BIOPSYCHOLOGY



Open Education Resource Initative
OERI

LibreTexts™

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

OERI

Biopsychology

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (https://LibreTexts.org) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden born from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by NICE CXone Expert and was supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org or visit our main website at https://LibreTexts.org.

This text was compiled on 04/14/2026

TABLE OF CONTENTS

About the Authors

Licensing

1: Biopsychology as a Course of Study

- 1.1: What is Biopsychology?
- 1.2: History of Biopsychology
- 1.3: Connecting Biology to Behavior
- 1.4: What are the Divisions of Biopsychology?

2: Research Methods and Ethical Considerations of Biological Psychology and Neuroscience

- 2.1: Introduction to Research in Biological Psychology
- 2.2: Non-Invasive Techniques: Structural Imaging
- 2.3: Non-Invasive Techniques: Direct Functional Imaging Techniques
- 2.4: Non-Invasive Techniques: Indirect Functional Imaging Techniques
- 2.5: Other Non-Invasive Techniques
- 2.6: Invasive Techniques
- 2.7: Ethical Considerations

3: Evolution, Genes, and Behavior

- 3.1: Evolution of Minds and Behavior
- 3.2: Evolution - Mechanisms and Evidence
- 3.3: Human Evolution
- 3.4: Evolutionary Theories in Psychology
- 3.5: DNA, Chromosomes, and Gene-Environment Correlations
- 3.6: The Nature-Nurture Question
- 3.7: Epigenetics in Psychology
- 3.8: Human Genetics
- 3.9: Genes and Behavior

4: Nervous System Anatomy

- 4.1: Nervous System Tissue
- 4.2: Nervous System Structure and Terminology
- 4.3: Nervous System Development
- 4.4: The Central Nervous System
- 4.5: The Peripheral Nervous System
- 4.6: Communication and the Endocrine System

5: Communication within the Nervous System

- 5.1: Neurons and their Basic Functions
- 5.2: Neurons Generate Voltage Changes to Code Information
- 5.3: Synaptic Transmission

6: The Effects of Psychoactive Drugs

- 6.1: Psychopharmacology and Psychoactive Drug Classification
- 6.2: Mechanisms of Psychoactive Drug Action
- 6.3: Substance Use and Addiction - Theories, Brain Mechanisms, and Treatment
- 6.4: Specific Drug Effects, Misuse, and Risks
- 6.5: Psychotherapeutics

7: Senses

- 7.1: Senses Overview
- 7.2: Vision
- 7.3: Hearing
- 7.4: Touch
- 7.5: Taste and Smell

8: Sensation and Perception

- 8.1: Sensation vs Perception
- 8.2: Psychophysical Methods
- 8.3: Perception and Action
- 8.4: Gestalt Principles of Perception

9: Movement

- 9.1: Classification of Muscle Types and Functions
- 9.2: Synaptic Control of Movement- Neuromuscular Junction and Motor Units
- 9.3: Spinal Reflexes
- 9.4: Motor Circuitry- Neural Structures and Pathways
- 9.5: Impact of Disease States and Neural Damage on Motor Control

10: Learning and Memory

- 10.1: Learning, Genes, and Adaptation
- 10.2: Types of Learning and Biological Adaptation
- 10.3: Introduction to Memory and the Brain
- 10.4: Synaptic Mechanisms of Learning and Memory
- 10.5: Brain Mechanisms of Memory Disorders
- 10.6: Localization of Types and Stages of Memory
- 10.7: Hormones, Learning, and Memory

11: Wakefulness and Sleep

- 11.1: Consciousness
- 11.2: What is Sleep?
- 11.3: Stages of Sleep
- 11.4: The Brain and Sleep
- 11.5: Theories of Sleep
- 11.6: Sleep Disorders

12: Ingestive Behaviors - Eating and Drinking

- 12.1: Maintaining Water Balance
- 12.2: Triggering Drinking Behavior - Osmometric and Volumetric Thirst

- 12.3: Ingestive Behavior Basics
- 12.4: Digestion and the Hormonal Control of Metabolism
- 12.5: Hunger - Theories, Detectors, and the Hypothalamus
- 12.6: Ingestive Behavior in Context

13: Sexuality and Sexual Development

- 13.1: Sex Hormones, Sexual Motivation, and Reproductive Behaviors
- 13.2: Sexual Development, Anatomy and Physiology
- 13.3: Differences of Sexual Development
- 13.4: Gender
- 13.5: Sexual Orientation
- 13.6: Sex Differences and Evolutionary Interpretations

14: Intelligence and Cognition

- 14.1: Overview of Brain Mechanisms in Intelligence, Language, and Cognition
- 14.2: Cognition, Intelligence, and Language as Psychological Adaptations
- 14.3: High Intelligence-Nature or Nurture; Learning Disabilities
- 14.4: Cognition- Categories, Concepts, Schemas and the Brain
- 14.5: Brain Mechanisms and Intelligence

15: Language and the Brain

- 15.1: Introduction to Language
- 15.2: Broca's Area, Wernicke's Area, and Additional Language-Processing Areas in the Brain
- 15.3: Other Brain Models of Spoken and Written Language
- 15.4: Handedness, Language, and Brain Lateralization
- 15.5: The Right Hemisphere's Contribution to Language

16: Emotion and Stress

- 16.1: The Neurological Bases of Emotions
- 16.2: Valence of Emotions
- 16.3: Fight or Flight
- 16.4: Social Neuroscience
- 16.5: Stress and health

17: Biological Bases of Psychological Disorders

- 17.1: Biological Factors in Psychological Disorders- An Introduction
- 17.2: Biological Basis of Mood Disorders- Major Depressive and Bipolar Disorders
- 17.3: Biological Basis of Anxiety Disorders
- 17.4: Biological Basis of Schizophrenia Spectrum Disorder

18: Supplemental Content

- 18.1: Chapter 3- Population Evolution (and the Modern Synthesis)
- 18.2: Chapter 3- Phenotypic Variation and Population Genetics
- 18.3: Chapter 3- Adaptive Evolution
- 18.4: Chapter 3- Modern Primates and Their Evolution
- 18.5: Chapter 3- Material Culture
- 18.6: Chapter 10- Computer Models and Learning in Artificial Neural Networks

- 18.7: Chapter 10- Acetylcholine, Enriched Experience, and Memory
- 18.8: Chapter 10- Memory Types and Stages
- 18.9: Chapter 10- The Psychology of Memory Encoding, Storage, and Retrieval
- 18.10: Chapter 10- Psychological Factors Affecting Learning, Memory, and Forgetting
- 18.11: Chapter 14- Traditional Models of Human Intelligence
- 18.12: Chapter 14- Problem Solving, Categories and Concepts
- 18.13: Chapter 14- The Brain and Evolution of Cognition
- 18.14: Chapter 14- Mirror Neurons, Theory of Mind, Social Cognition, and Neuroscience of Its Disorders
- 18.15: Chapter 15- A Gene Essential for Speech
- 18.16: Chapter 15- The Origins of Language

Index

Glossary

Detailed Licensing

About the Authors

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

The Authors

Naomi Bahm is a professor of Psychology at Cosumnes River College, part of the Los Rios Community College District in Sacramento, California. She earned her B.A. in Biology and Psychology from the College of Saint Benedict (Saint Joseph, Minnesota), followed by serving as a Peace Corps Volunteer in the aquaculture program in Cameroon, Africa. She then earned her M.A. and Ph.D. in Integrative Biology from the University of California at Berkeley, combining electroencephalograph (EEG) methods and the Adult Attachment Interview (AAI). Naomi now spends part of her “free” time as an AAI trainer. In addition to teaching, she enjoys reading, sewing, and spending time with family and friends (especially outdoors).

Textbook Authors

Author Images	Author Bios
 Naomi Bahm	Naomi Bahm is a professor of Psychology at Cosumnes River College, part of the Los Rios Community College District in Sacramento, California. She earned her B.A. in Biology and Psychology from the College of Saint Benedict (Saint Joseph, Minnesota), followed by serving as a Peace Corps Volunteer in the aquaculture program in Cameroon, Africa. She then earned her M.A. and Ph.D. in Integrative Biology from the University of California at Berkeley, combining electroencephalograph (EEG) methods and the Adult Attachment Interview (AAI). Naomi now spends part of her “free” time as an AAI trainer. In addition to teaching, she enjoys reading, sewing, and spending time with family and friends (especially outdoors).
 Bucky Bhadha	Bucky Bhadha has taught a range of psychology and development courses. She got her bachelor's degree in India and then graduate degrees from the School of Education, and Department of Psychology at the University of Southern California. She sees herself as an antiracism advocate working through open educational resources, open pedagogy, and the Prison Education Project. In her free time, Bucky enjoys playing board games with family and traveling.
 Alan Keys	Alan Keys is a Professor of Psychology at Sacramento City College where he has been teaching courses and labs in Biological Psychology, Psychopharmacology, and Research Methods since 2000. During his tenure at SCC, he also served for many years as the Faculty Research Coordinator where he worked with faculty on classroom-based research and played a key role in the initial development of Student Learning Outcome Assessment across the college. Alan completed his doctoral studies in Behavioral Neuroscience at UCLA and postdoctoral work at Oregon Health Sciences University where he did extensive work with animal models of drug abuse with a primary focus on the connection between drug-seeking behavior and alterations in neural circuitry associated with reward. He has remained passionate about the field of Biological Psychology and is committed to inspiring students to discover their own calling and to provide them with the tools and support to pursue them.
 Kenneth Koenigshofer	Kenneth Koenigshofer received his Ph.D. in physiological psychology from the University of California, Riverside, and is now retired after 38 years of teaching at Chaffey College, California State Polytechnic University at Pomona, and the University of Maryland. He is author of Mind Design: The Adaptive Organization of Human Nature, Minds, and Behavior (Pearson, 2010), revised as an e-book edition in 2016. His research focus on the evolution of cognition and intelligence led to publication of his theory of the evolution of general intelligence in the Spring 2017 volume (38:2) of the Journal of Mind and Behavior. Ken and his family live in a beach town north of San Diego.

Author Images	Author Bios
 Kevin Smith	Kevin R. Smith is a professor of psychology at Rio Hondo college. He earned his Ph.D. in psychology from the University of California, Irvine in 2008 doing functional MRI work on the auditory system. His hobbies include gardening and trying to hear as much live music as possible.

Licensing

*A detailed breakdown of this resource's licensing can be found in **Back Matter/Detailed Licensing**.*

CHAPTER OVERVIEW

1: Biopsychology as a Course of Study

> This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

1.1: What is Biopsychology?
1.2: History of Biopsychology
1.3: Connecting Biology to Behavior
1.4: What are the Divisions of Biopsychology?

This page titled 1: Biopsychology as a Course of Study is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

1.1: What is Biopsychology?

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe the field of biopsychology, its areas of focus, and other names for the field.
2. Discuss a common myth about the brain and why it is false.
3. Describe several brain imaging methods.
4. Discuss materialism (physicalism) and determinism as assumptions underlying biopsychology.
5. Describe the principle of brain organization referred to as "the localization of function."

1.1.1 Overview

Biopsychology is the study of biological mechanisms of behavior and mental processes. It examines the role of the nervous system, particularly the brain, in explaining behavior and the mind. This section defines biopsychology, critically examines a common myth about the brain, and briefly surveys some of the primary areas of research interest in biopsychology.

1.1.2 Biopsychology - The Interaction of Biology and Psychology

Psychology is the scientific study of behavior and mental processes in animals and humans. Modern psychology attempts to explain behavior and the mind from a wide range of perspectives. One branch of this discipline is biopsychology which is specifically interested in the biological causes of behavior and mental processes. Biopsychology is also referred to as biological psychology, behavioral neuroscience, physiological psychology, neuropsychology, and psychobiology. The focus of biopsychology is on the application of the principles of biology to the study of physiological, genetic, evolutionary, and developmental mechanisms of behavior in humans and other animals. It is a branch of psychology that concentrates on the role of biological factors, such as the central and peripheral nervous systems, neurotransmitters, hormones, genes, and evolution on behavior and mental processes.

Biological psychologists are interested in measuring biological, physiological, or genetic variables in an attempt to relate them to psychological or behavioral variables. Because all behavior is controlled by the central nervous system (brain and spinal cord), biopsychologists seek to understand how the brain functions in order to understand behavior and mental activities. Key areas of focus within the field include sensation and perception; motivated behavior (such as hunger, thirst, and sex); control of movement; learning and memory; sleep and biological rhythms; and emotion. With advances in research methods, more complex topics such as language, reasoning, decision making, intelligence, and consciousness are now being studied intensely by biological psychologists.

1.1.3 Do humans use only 10% of their brains?

The frequently repeated claim that humans use only 10% of their brains is false. The exact origin of this myth is unknown, but misinterpretations of brain research are likely to blame. In experiments with animal brains during the 1800's through the early 1900's, Marie-Jean-Pierre Flourens and Karl Lashley destroyed and/or removed as much as 90% of the brain tissue of their animal subjects. Nevertheless, these animals could still perform basic behavioral and physiological functions. Some who read these results made the incorrect assumption that this meant that animals were using only 10% of their brains. Subsequently, this interpretation was generalized to humans (Elias and Saucer, 2006). Furthermore, prominent psychologists and researchers, such as Albert Einstein, Margaret Mead, and William James, were also quoted as saying that humans are using only a small portion of their brain (Elias and Saucer, 2006), fueling the 10% myth.

Due to advances in biopsychology and other related fields, we now have a greater understanding of the complexity of the brain. We may not be using our brains as efficiently as possible at all times, but we are using the entirety of our brain as each part contributes to our daily functioning. Studies of humans with brain damage have revealed that the effects of brain damage are correlated with where the damage is and how extensive it is. In other words, where damage occurs determines what functions are impacted and more damage has more of an effect. This reflects a key organizational principle of the brain: the localization of function. This

principle means that specific psychological and behavioral processes are localized to specific regions and networks of the brain. For example, we now know that damage to an area of the brain known as the primary visual cortex, at the very back of your head in the occipital lobe, will result in blindness even though the rest of your visual system, including your eyes, is functioning normally. This syndrome is known as cortical blindness, to distinguish it from blindness that is caused by damage to the eyes. We now know that damage to a small area less than the size of a quarter at the very base of your brain results in disruption of feeding and regulation of body weight. Damage to another area of the brain located near your temples disrupts your ability to form new memories for facts and events, while leaving your ability to learn new motor tasks (such as skating or riding a bike) completely unaffected. Damage to another brain area causes face blindness, or prosopagnosia, a disorder in which the afflicted individual can still see normally except that they cannot recognize familiar faces, even the faces of close family members or even their own face in a photograph.

In the pages that follow in this textbook, you will learn many amazing things about the brain, and the nervous system in general. Get ready for many surprises as we explore the 3 pounds of brain tissue between our ears that make up the most complex piece of matter in the known universe. In this book, we examine some of what scientists now know about this astonishing organ, the brain, and how it functions to produce mind and behavior.

1.1.4 The Brain

The most important organ controlling our behavior and mental processes is the brain. Therefore, biopsychologists are especially interested in studying the brain, its neurochemical makeup, and how it produces behavior and mental processes (Wickens, 2021).

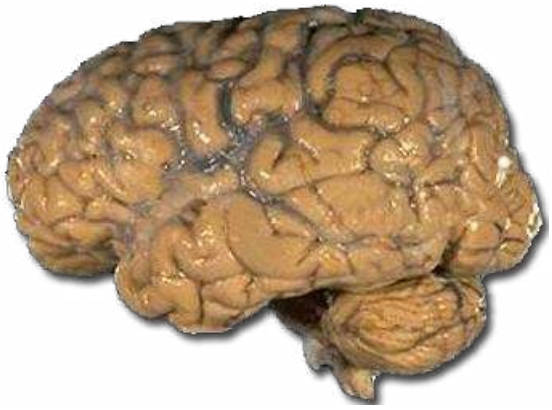


Figure 1.1.1: Photo of the left side of a human brain. Front is at the left. The massive cerebral cortex hides many subcortical brain structures beneath. Note the folds on the surface of the cerebral cortex. This folding increases in mammal species with increasing complexity of the brain of the species and is thought to originate from the "cramming" of more cortical tissue into the skull over evolutionary time. (Image from Wikimedia Commons; File:Human brain NIH.png; https://commons.wikimedia.org/wiki/F..._brain_NIH.png; image is from NIH and is in the public domain.).

Modern technology through neuroimaging techniques has given us the ability to look at living human brain structure and functioning in real time. Neuroimaging tools, such as functional magnetic resonance imaging (fMRI) scans, are often used to observe which areas of the brain are active during particular tasks in order to help psychologists understand the link between brain and behavior.

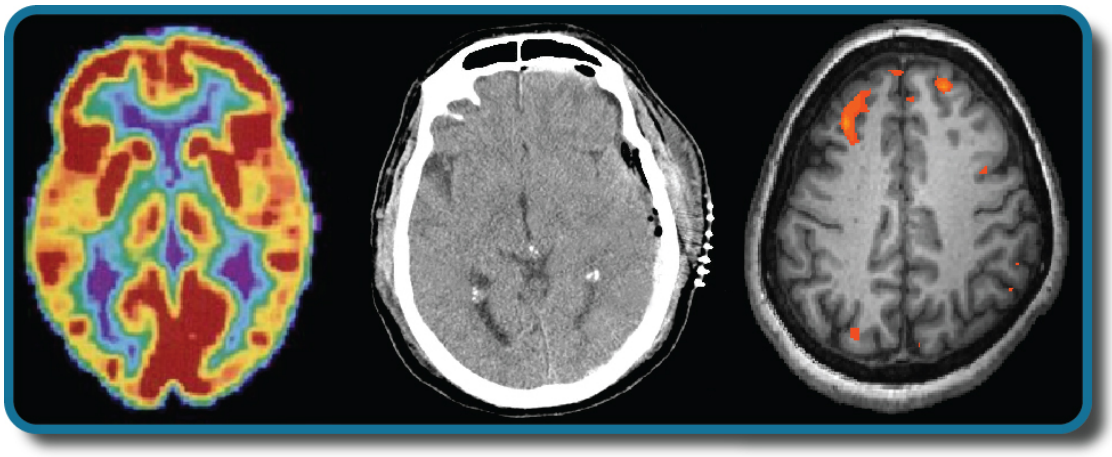


Figure **1.1.2**: Brain Imaging Techniques (Copyright; author via source)

Different brain-imaging techniques provide scientists with insight into different aspects of how the human brain functions. Three types of scans include (left to right) PET scan (positron emission tomography), CT scan (computed tomography), and fMRI (functional magnetic resonance imaging). (credit “left”: modification of work by Health and Human Services Department, National Institutes of Health; credit “center": modification of work by "Aceofhearts1968"/Wikimedia Commons; credit “right”: modification of work by Kim J, Matthews NL, Park S.)

Magnetic resonance imaging (MRI) scans of the head are often used to help psychologists understand the links between brain and behavior. As will be discussed later in more detail, different tools provide different types of information. A functional MRI, for example, provides information regarding brain activity while an MRI provides only information about structure.

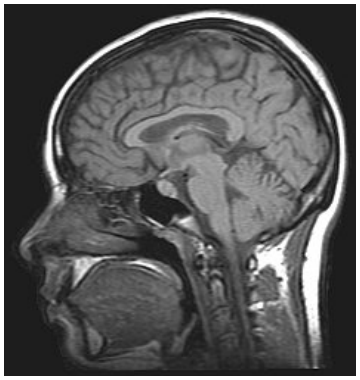


Figure **1.1.3**: MRI of Human Brain (public domain; via Wikimedia Commons)

1.1.5 A Scientific Model of the Universe: Two Basic Assumptions

The modern scientific view of the world assumes that there are at least two fundamental properties of the universe. The first assumption common to most scientists is that the entire universe is material or physical, composed exclusively of matter (including so-called dark matter) and energy (including electromagnetic energies such as light, heat, ultraviolet, and various other radiant energies, as well as other physical energies and forces such as electrical, gravitational, nuclear, and so forth). The entire universe is governed by physical laws. This view of the universe is called **materialism or physicalism**--the view that everything that exists in the universe consists of matter, energy, and other physical forces and processes.

Most important to biopsychology is the application of this principle to psychology and psychological processes. If everything in the universe is physical, then applied to psychology, including biopsychology, this means that the mind, our mental processes and subjective mental experiences, must also be entirely physical processes in an entirely material brain. Using this fundamental assumption of the modern scientific view of the universe, this means that the mind is entirely material, dependent upon the physical activities of an entirely material organ, the brain. On this view, the mind is what the brain does, and the brain and its processes are

completely physical, material, just matter and energy in highly specific and organized form. This does not mean that all matter is conscious, nor does it mean that the mind is just energy. The key idea is that the matter and energy must be organized in a particular way. That is, a mind, consciousness, can only emerge from matter, energy, and physical processes if they are organized in a very specific and complex form--that form that we know as a brain and its physical operations. Where did brains come from and how did they acquire the specific organization of matter and energy needed to make a conscious mind within them? The scientific answer is evolution.

Although this is the view among most biological psychologists, there are a few who believe, like many students do, that the brain, along with the rest of life, was created by a divine being and that therefore the mind has divine origins. Typically, this belief is accompanied by the assumption that the mind is not physical, but that it is akin to the soul and the soul is believed to be non-material. Belief that the mind is non-material and therefore independent of the physical brain and its physical processes is known as **mind-body dualism or mind-brain dualism,** which literally means that the mind and the functioning of the brain (assumed to be entirely physical) are two (dual) separate processes, completely independent of one another. The origin of dualism is often attributed to the 17th century French philosopher and mathematician, Rene Descartes. If this view were true, then we would expect that brain damage would have no effect on the mind. However, brain damage does affect the mind and the specific location of the damage produces more or less specific, fairly predictable, effects on the mind, modifying the mind and behavior in various ways. Examples of this are coma due to head injury; the effects of Parkinson's disease on movement after the disease damages areas of the brain known as the basal ganglia; changes in personality and emotion due to injury to the front of the brain, specifically the frontal lobes; memory loss in Alzheimer's Disease; and so on. Though you don't have to accept the assumption of physicalism when studying the brain if your religious beliefs are contrary to the idea, nevertheless it is important that you be aware of the assumption of physicalism/materialism that most biological psychologists accept, at least as a working hypothesis, if not a philosophical position, as they do their brain research.

The second major assumption among most scientists is **determinism**--the belief that all events in the universe have prior causes and that these causes are external to the human will. This implies that humans do not have free will. Instead human behavior is caused by events external to us such as our upbringing, our social and cultural environment, by our brain structure and functioning, and by our genes and our evolution as a species. In some versions of this viewpoint, since we do not have control over many of the factors in our environment, our genes, and our evolution as a species, our brain function and thus our behavior is actually controlled by causes outside of our control. On this view, free will is an illusion that arises from our awareness of our mental processes as we make choices based on our selection of various behavioral options that we see open to us, but what we often fail to realize is that those choices are determined by many factors beyond our awareness and control (Koenigshofer, 2010, 2016). Free will vs. determinism is an issue that is far from being resolved and remains controversial even among scientists, including biological psychologists. Investigation by biological psychologists of the brain processes involved in choice and decision-making is ongoing and may eventually shed light on this difficult issue. Again, it is not necessary for you to be a determinist to study the brain, but it is important for you to be aware of the doctrine of determinism as you consider the implications of brain research as you progress through this textbook and your course in biological psychology.

1.1.6 References

Elias, Lorin J. and Saucier, Deborah M. (2006). Neuropsychology: Clinical and Experimental Foundations. Pearson.

Fredrickson, B. L., Maynard, K. E., Helms, M. J., Haney, T. L., Siegler, I. C., & Barefoot, J. C. (2000). Hostility predicts magnitude and duration of blood pressure response to angand Referenceser. *Journal of behavioral medicine*, *23*(3), 229–243. https://doi.org/10.1023/a:1005596208324

Koenigshofer, K. (2010). *Mind Design: The Adaptive Organization of Human Nature, Minds, and Behavior.* Pearson Educational Solutions, Boston.

Koenigshofer, K. (2016). *Mind Design: The Adaptive Organization of Human Nature, Minds, and Behavior.* Revised Edition. Amazon e-book.

Wickens, Andrew P. (2021). Introduction to Biopsychology. Sage Publishing.

1.1.7 Attributions

Adapted from Introduction to Psychology by Jennifer Walinga is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

---

This page titled 1.1: What is Biopsychology? is shared under a CC BY-NC-SA 3.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

1.2: History of Biopsychology

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Discuss the history of biopsychology.
2. Discuss the important contributors to biopsychology and the contributions each made.
3. Explain the current direction of biopsychology.

1.2.1 Overview

Biopsychology as a scientific discipline is relatively new. However, early philosophers such as Hippocrates wrote about the importance of biological factors in human behavior. In this section, we will examine the emergence of biopsychology as a scientific field. We will further outline important events and contributions to biopsychology, beginning with early philosophers and their understanding of the biological aspects of behavior and mental processes. We then examine a number of more recent historical figures who made major contributions to the development of biological psychology and discuss their contributions to the discipline. Lastly, we will look at current directions in the field.

1.2.2 Biopsychology as a Discipline

Biopsychology as a scientific discipline emerged from a variety of scientific and philosophical traditions in the 18th and 19th centuries. Although the exact date of inception of biological psychology is unknown, there have been a number of milestones in its emergence. William James in his book, *The Principles of Psychology* (1890), argued that the scientific study of psychology should be grounded in an understanding of biology (Walinga, 2014). Like many early psychologists, James had extensive training in physiology.

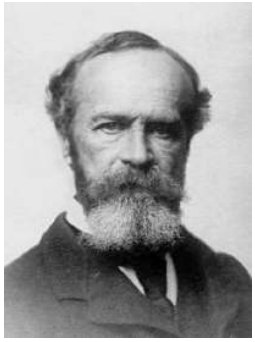


Figure 1.2.1: William James. (Public Domain; via Wikimedia Commons)

In *An Outline of Psychobiology,* Knight Dunlap used the term "psychobiology" to explain the role of biology in behavior. In 1949, Donald Hebb wrote his influential book, *The Organization of Behavior*, where he introduced the first comprehensive theory on how the brain might create and control complex psychological functioning, such as thought, memories, emotions and perceptions. In other words, even before biopsychology was formally recognized as a branch of psychology or as a discipline,

the connection between psychology and biology was recognized. With case studies such as Phineas Gage and invention of methods such as the electroencephalogram (EEG) and computed tomography (CT or CAT) scans, scientists began to link the brain to specific behavior and cognition as biopsychology as a discipline began to emerge.

1.2.3 The Origins of Biopsychology

It is important to examine the historical path of our understanding of the brain and its role in our behavior and mental processes. Examining the history of biopsychology allows us to understand its development over time, highlighting instances where researchers were wrong about the nature of brain-behavior relationships and revealing what we have yet to explain (Saucier and Elias, 2006). Studying the history of a scientific discipline gives us a roadmap of where we have traveled from and in what direction we need to go.

1.2.3.1 Early Views of Brain and Behavior

Early philosophers, such as Aristotle (384-322 B.C.E.), believed that one's mind resided in the heart. He believed that since our blood started from the heart, the soul also originated there. Plato (428-347 B.C.E.) argued that the executor of reason was the heart and our animalistic desires and emotions were controlled by the liver (Gross, 1987). Many ancient cultures, including the Chinese, Indian, and Egyptian also shared the same belief (Carlson, 2014). When the Egyptians embalmed a person, the heart was saved and buried with the individual but the brain was discarded (Klein and Thorne, 2006). However, there were early Greeks, such as Hippocrates (460-377 B.C.E.) , who believed that it was the brain and not the heart where the locus of the mind resided. He wrote: "It ought to be generally known that the source of our pleasure, merriment, laughter, and amusement, as of our grief, pain, anxiety, and tears is none other than the brain. It is specially the organ which enables us to think, see, and hear.....It is the brain too which is the seat of madness and delirium, of the fears and frights which assail us" (Gross, 1987, p. 843-844).

During the 3rd Century B.C., in Alexandria, a science museum was established. Since human dissection was practiced in that city for centuries, the anatomical study of the human body flourished (Gross, 1987). Two neuroanatomists, Herophilos and Erasistratos, contributed to our knowledge of the human brain. Herophilos distinguished the cerebellum (at the very base of the back of the brain) and the cerebrum (the two cerebral hemispheres). He hypothesized that since the cerebellum was denser than the other parts of the brain, it must control the muscles (a guess of impressive accuracy). And he provided the first clear description of the cavities within the brain known as ventricles (Figure 1.2.1). Erasistratos continued the work of Herphilos and proposed that human intelligence was related to the number of convolutions (ridges) in the brain; the more convolutions an individual's brain had, the more intelligent that person would be.

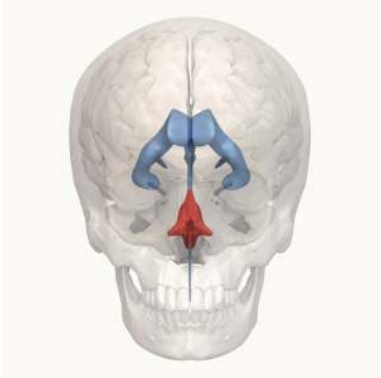


Figure **1.2.2**: Human Ventricles (CC BY-SA 2.1; BodyParts3D by DBCLS via Wikimedia Commons)

During the Roman Empire, the Greek anatomist Galen (A.D. 130-200) dissected the brains of sheep, monkeys, dogs, pigs, among other non-human mammals (Carlson, 2014). He believed the brain to be the site of sensation and thought, and the controller of movement (Gross, 1987). He stated that spinal cord was an extension of the brain and chronicled the relationship between the

spinal nerves and specific muscles each controlled. For the next advance in understanding spinal function, we must await Bell and Magendie in the 19th Century.

1.2.3.2 Middle Ages



Figure **1.2.3**: Drawing of Ibn-Sina. (Public Domain; via Wikimedia Commons)

Avicenna (980-1037 C.E.) was a Persian physician who some regard as the father of modern medicine. He recognized biological psychology in his treatment of illnesses that involved emotions (Syed, 2002). In his book, *The Canon of Medicine (Qanun: Law of Medicine)*, Avicenna dedicated three chapters to neuropsychiatric disorders where he defined madness (Junun) as originating in the middle part of the brain. He discussed a condition similar to schizophrenia, which he called Junun Mufrit, characterized by agitation, behavioral and sleep disturbances, giving inappropriate answers to questions, and an occasional inability to speak. Avicenna also discovered the cerebellar vermis, which he simply called the vermis, and the caudate nucleus. Both terms are still used in neuroanatomy today (Mohamed, 2008). He further explained that "humidity" inside the head contributed to mood changes and when this moisture inside the brain goes beyond limits, the brain loses control over its rationality and mental disorders appear (Haque, 2004). Lastly he presented detailed knowledge about skull fractures and their surgical treatments (Aciduman, Arda, Ozakturk, and Telatar, 2009). He also discussed phenomena such as insomnia, mania, hallucinations, nightmares, dementia, epilepsy, stroke, paralysis, vertigo, melancholia, and tremors (Mohamed, 2008). He was also the first person to associate mental deficits with deficits in the brain's middle ventricle or frontal lobe (Mohamed, 2008).

1.2.3.3 Renaissance

By the beginning of the 16th century, Andreas Vesalius, founder of modern anatomy, rekindled interest in neuroscience (Gross, 1987). Through his dissection of human cadavers, Vesalius found problems with the Galenic view of anatomy (which emphasized three major systems--the heart, brain, and blood--and the importance of a balance between four bodily fluids--blood, phlegm, and black and yellow bile). Vesalius published seven books, with the fourth and seventh books devoted to the nervous system (Van Laere, 1993). Vesalius outlined seven pairs of 'brain nerves', each with a specialized function. In his seventh book which was solely devoted to the brain, he successfully describes the cerebral membranes, the ventricular system, and the cerebrum (Van Laere, 1993). Other scholars furthered Vesalius' work by adding their own detailed sketches of the human brain.

René Descartes (1596-1650) proposed the reflexive theory in terms of our behavior (Elias and Saucier, 2006). This theory explained reflexive behavior as an "external stimuli [that] would move the skin, in turn moving the filaments, releasing the animal spirits and innervating the muscles" (Elias and Saucier, 2006, p 6). His mechanical explanation of movement without conscious intervention allowed for scientific study of behavior. Furthermore, his theory encouraged using animal models to study biological bases of behavior (Cobb, 2002). He also studied the physiology of the brain, proposing the theory of dualism (the view that mind and brain are separate) to tackle the issue of the brain's relation to the mind. He suggested that the pineal gland was where the mind interacted with the body after recording the brain mechanisms responsible for circulating cerebrospinal fluid (Elias and Saucier, 2006).



Figure 1.2.4: Descartes Drawing - Reflexive role of pain in contracting muscle. (Public Domain; via Wikimedia Commons)

1.2.4 Discoveries of the 19th Century

1.2.4.1 Localization of Function

Starting in early 1800's, a series of researchers began to catalog different brain parts and their presumed role in behavior. In 1811, Jean-Cesar Legallois (1770-1840) discovered that when he destroyed tissue in the medulla, a procedure called lesioning, there was an immediate cessation of respiration (Elias and Saucier, 2006). Charles Bell (1774-1842) and Francois Magendie (1783-1840) discovered the different function of spinal cord nerves where the ventral roots transmitted motor impulses and posterior roots received sensory input (Bell-Magendie Law). Given the specialization found in spinal cord nerves, Bell recommended that further investigation of the the entire nervous system should examine functional and anatomical segregation (Elias and Saucier, 2006).

Taking on the challenge, Franz Joseph Gall (1758-1828) and Johann Spurzheim (1776-1832) formulated phrenology where measurement of the skull was used to determine personality characteristics of an individual. Gall made the claim that here were twenty-seven distinct cognitive abilities that could be localized on the cortex of the human brain (the cortex composing the outermost part of the brain). He believed that the design of the brain changed as each of us developed certain characteristics and this resulted in corresponding changes in the skull. Where the skull rises, one can see excellence in the human qualities associated with the underlying brain and where there were depressions in the skull, human defects were represented (Lau, Loi and Abdullah, 2021). Measuring the skull by using a technique called cranioscopy would allow the scientist to detect deformation and bumps on the skull which would outline the person's personality (Elias and Saucier, 2006).

In 1848, John Martyn Harlow treated Phineas Gage and documented his case. Gage, who was a railroad worker, had his frontal lobe pierced by an iron tamping rod in a blasting accident. He survived the trauma but suffered extensive damage to his left prefrontal cortex (Macmillan, 2001). Through Gage's case study, Harlow showed the connection between the prefrontal cortex damage, executive functioning, and personality changes.

In 1861, Paul Broca (1824-1880) presented a case where a young man had sustained damage to the left frontal lobe and as a result was unable to produce spoken speech. However, the patient was able to comprehend speech and had average intelligence. After his death, Broca performed an autopsy and determined that the patient had a lesion in the frontal lobe in the left cerebral hemisphere. Broca collected similar cases of damage to the left frontal lobe with the patient losing their ability to speak (Lau, Loi and Abdullah, 2021). One such case was Phineas Gage who lost his speech after his accident. Broca published his findings from the autopsies of twelve patients in 1865. Today, we recognize Broca's Area in the frontal lobe as a critical region in speech production. His work inspired others to perform careful autopsies with the aim of linking more brain regions to sensory and motor functions.

Carl Wernicke (1848-1904) further developed the theory of the specialization of specific brain structures in language comprehension and production. He wrote a paper in 1874, suggesting that housed in the left temporal lobe was an auditory comprehension center. He stated that when this area, now known as Wernicke's area, was damaged, the individual could not

comprehend other's speech or use words correctly (Elias and Saucier, 2006). The concept of "localization of function", specific functions residing in specific parts of the brain, was established the end of the 19th century. Respiration and the medulla and the roles of Broca's and Wernicke's areas in language are examples of this concept.

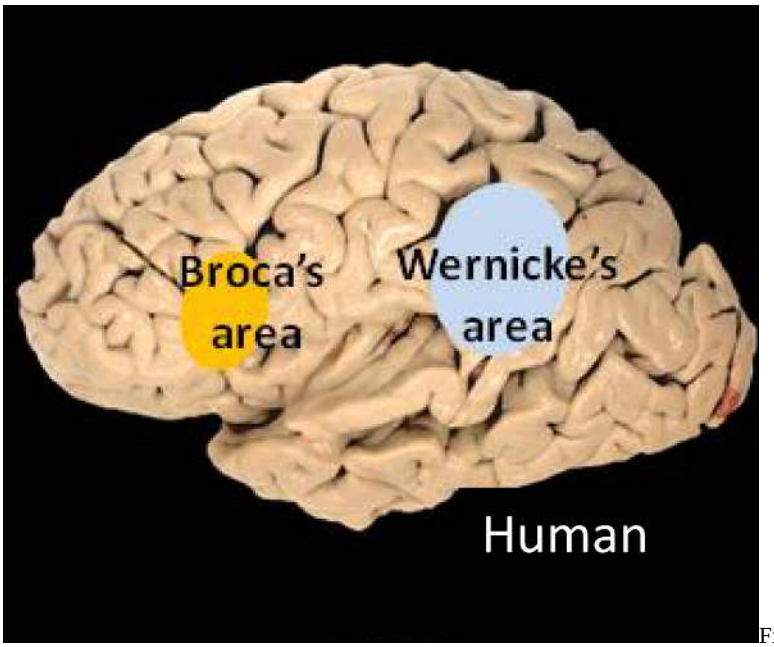


Figure 1.2.5: The two language areas of the brain. (Public Domain; via Wiki media Commons)

1.2.4.2 The Neuron

The last monumental neurological discovery of the 19th century was the identification of the specific structural unit responsible for these brain functions, the neuron. Studies of the brain became more sophisticated after the invention of the microscope and the development of a staining procedure by Camillo Golgi during late 1873 (Elias and Saucier, 2006). Golgi (1843-1926) was a histologist (one who studies the cellular structure of tissue) and a physician (Klein and Thorne, 2006). He used a silver chromate salt to reveal the intricate structures of single neurons (Elias and Saucier, 2006). Through this technique, he was able to reveal that the neuron has three distinct structures: dendrite, cell body, and axon (Elias and Saucier, 2006). Furthermore, he revealed that the axon can be short or very long, enabling the cell body to send messages to neurons far away. Therefore the axon must be the output carrier for the neuron. This technique is called the Golgi stain, named in his honor.

The Golgi stain technique was used by Santiago Ramón y Cajal to outline the neural connections in the brain (Elias and Saucier, 2006). While Golgi believed that neurons were fused together to form a continuous circuit, Cajal demonstrated that neurons are not physically connected but do have a mechanism for communicating (Elias and Saucier, 2006). This led to the formation of the **neuron doctrine**, the hypothesis that the functional unit of the brain is the neuron, also known as a nerve cell. Today, the field accepts neuronal theory which states that the nervous system is made up of individual nerve cells called neurons (Klein and Thorne, 2006).

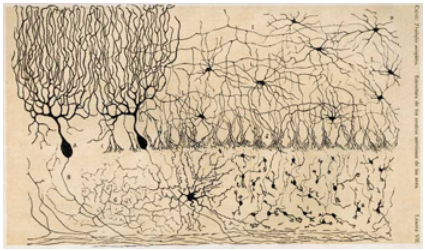


Figure 1.2.6: Ramon y Cajal's Drawing of A Neuron. (Public Domain; via Wikimedia Commons)

Golgi and Ramón y Cajal shared the Nobel Prize in Physiology or Medicine in 1906 for their extensive observations, descriptions, and categorizations of neurons throughout the brain. Golgi was praised for his staining technique and Cajal for developing the

neuron doctrine.

One of the most important questions given Cajal's neuron doctrine was the mechanism for communication between neurons. The role of electricity in nerves and electrical communication between neurons was first observed in dissected frogs by Luigi Galvani. In 1780, he showed that an electrical stimulus applied to the motor nerve of a frog's nerve cell produced a twitch. This experiment suggested that electricity was an important functional element of the nervous system (Klein and Thorne, 2006).

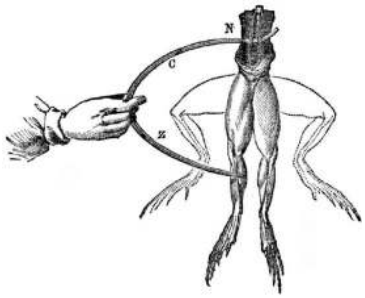


Figure 1.2.7: Drawing showing Galvani's Frog Leg Electricity Experiment. (Public Domain; via Wikimedia Commons)

1.2.5 Discoveries of the 20th Century

1.2.5.1 The Synapse

Charles Scott Sherrington (1857-1952) accepted Cajal's neuron doctrine and set out to discover the point of functional contact between neurons. He called this connection point the **synapse**, which means "binding together" (Klein and Thorne, 2006). His discovery was made by studying spinal reflexes in dogs. Confirming Galvani's electrical theory, he demonstrated the electrical transmission between neurons (Elias and Saucier, 2006). He received the Nobel Prize in 1932 for his research. A German pharmacologist and physician, Otto Loewi (1873-1961), investigated chemical communication between neurons. He removed the heart of two frogs and placed them in separate containers, each one filled with a chemical solution that allowed the hearts to continue to beat. He stimulated the vagus nerve of the first heart which slowed down its beating. He shared the solution from the first heart's container with the second heart and the beating of the second heart also slowed. He concluded that when he stimulated the vagus nerve of the first heart, a chemical was released to slow down its beating and when the solution from the first heart's container was shared with the second heart, the chemical directed the second heart to also decrease its beating (Klein and Thorne, 2006). He originally called this chemical vagusstoff, referring to its secretion after the stimulation of vagus nerve. This chemical was later identified as a **neurotransmitter**, acetylcholine (ACh). ACh was originally identified in 1915 by Henry Hallett Dale. However, Loewi is credited with establishing its role in chemical communication between nerve cells (Klein and Thorne, 2006).

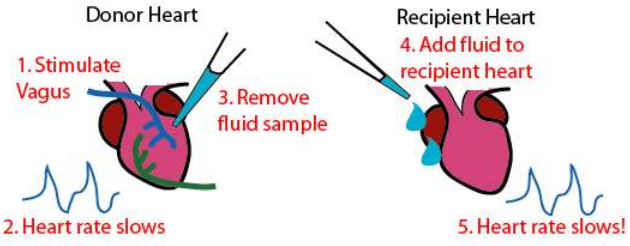


Figure 1.2.8: Loewi's experiment, describing the existence of Vagusstoff. (CC BY-SA 3.0; via Wikimedia Common)

Today, dozens of neurotransmitters have been identified in the brain. After acetylcholine, noradrenaline (now known as norepinephrine) was discovered in the 1940's, followed by dopamine and serotonin in the next decade and gamma-aminobutyric acid (GABA), glutamate, and glycine in the 1960's (Wickens, 2021). Later discoveries of neuropeptides, neuromodulators, and certain gases have added to the complexity of chemical communication in the brain. Our understanding of the impact and role of these chemical messengers and their effect on behavior and mental processes has given rise to the field of **psychopharmacology**.

1.2.5.2 Neural Conduction

Previous research had clearly demonstrated that electrical impulses played a role in neural communication. But the finer details of how this was occurring had not been discovered (Wickens, 2021). A series of discoveries led to our present understanding of neural conduction.

In 1902 and again in 1912, Julius Bernstein (1836-1917) was the first researcher to discover the **action potential** - the unique electrical signal traveling the length of an axon (Elias and Saucier, 2006). Additionally, he demonstrated that this electric conduction was due to change in **ionic concentration** of the cell's intracellular fluid. However, given the microscopic size of the cell, it was difficult to measure and study mammalian cells. In 1936, John Z. Young, an Oxford biologist, located a neuron in the body of a squid. The axon was nearly 1 mm in diameter which was 100 to 1,000 times larger than a mammalian axon and could be studied for about six hours after its dissection (Elias and Saucier, 2006). In 1952, two Cambridge University physiologists, Alan Hodgkin and Andrew Huxley published a set of papers describing their research on the **giant squid axon**. Much of our understanding of neural functioning came from the giant squid. Hodgkin and Huxley created a technique to measure electrical changes of the neuron by inserting an electrode inside and one positioned outside of the neuron without damaging it. This enabled them to measure **changes in electrical charge** when the neuron was at rest and when it produced an action potential (Wickens, 2021). A detailed discussion of both concepts, resting and action potential will follow in future chapters.

1.2.5.3 Studying Brain Activity and Brain Imaging

With the invention of machines such as the electroencephalograph (EEG), CT scan, and magnetic resonance imaging (MRI), scientists were able to peak into the activity and images of patients' brains while they were alive and engaged in mental activity. In 1924, Hans Berger recorded the first human brain EEG. This machine allowed researchers to study the overall electrical activity of the brain. In 1953, Kleitman and Aserinsky discovered Rapid Eye Movement (REM) using EEG on sleeping subjects. Today EEGs are helpful in diagnosing epilepsy and sleep disorders. Many credit their work as the beginning of modern study of sleep. The first image of the brain by CT scan was in 1971 by Godfrey Hounsfield. He used the CT scan to locate a brain tumor (Wells, 2005). CT scans can be used to locate brain tumors, injury, or atrophy. However, the CT scan uses Xrays to provide an image of the brain which could pose risk to brain tissue. Magnetic resonance imaging (MRI) was also used in 1970's to give pictures of the brain. An MRI does not use Xrays and is therefore found to be a safer method for producing human brain scans.

In 1950, positron emission tomography (PET) scans were used to look at brain activity or communication in a living patient. Researchers were able to look at different parts of the brain in different situations. In other words, they could look at real time activity of the brain as the person laughed or was scared to see what ares of the brain lit up. This gave an indication of what areas of the brain maybe responsible for different emotions or behaviors. The problem with PET scans was that an individual receiving a PET scan had to drink or be injected with a mildly radioactive substance, called a tracer. Once in the bloodstream, the amount of tracer in any given region of the brain can be monitored. As brain areas become more active, more blood flows to that area. A computer monitors the movement of the tracer and creates a rough map of active and inactive areas of the brain during a given behavior. PET scans show little detail, are unable to pinpoint events precisely in time, and require that the brain be exposed to radiation. In 1990s, a group in Bell Laboratories invented functional Magnetic Resonance Imaging (fMRI). This machine shows changes in brain activity over time by tracking blood flow and oxygen levels. The fMRI provides more detailed images of the brain’s structure, as well as better accuracy in time, than is possible in PET scans.

Today, these instruments have greatly contributed to our understanding of the brain and its function in behavior and cognition. With their high level of detail, MRI and fMRI are often used to compare the brains of healthy individuals to the brains of individuals diagnosed with psychological disorders. This comparison helps determine what structural and functional differences exist between these populations.

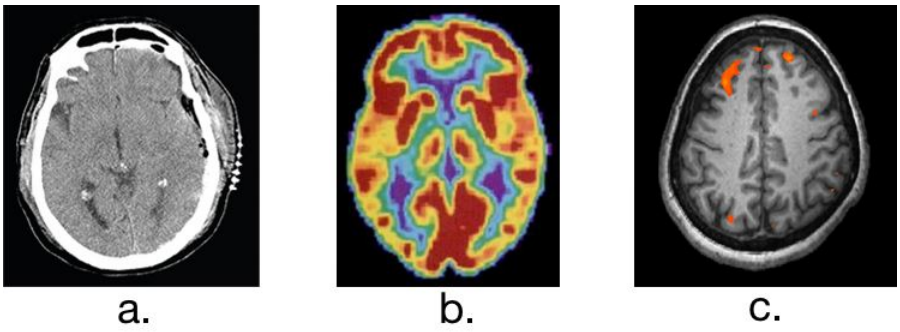


Figure 1.2.9: Images of brain using different instruments. a. CT scan; b. PET; c. fMRI (CC BY 3.0; via Wikimedia Commons)

1.2.6 Today

In April 2013, the BRAIN (Brain Research through Advancing Innovative Neurotechnologies) Initiative was launched. The focus of this initiative was to advance our understanding of the human brain. In the fiscal year 2020, 10.1 billion U.S. dollars has been allocated to neuroscience studies (Mikulic, 2021). Different institutions, agencies and foundations such as Food and Drug Administration, Defense Advanced Research Projects Agency, have joined in this effort.

1.2.7 References

Aciduman, Ahmet, Arda, Berna, Ozakturk, Fatma G., Telatar, Umit F. (2009). What does Al-Qanun Fi Al-Tibb (the Canon of Medicine) say on head injuries?. *Neurosurgical Review, 32* (3), 255-263.

Carlson, Neil R. (2014). Foundations of Behavioral Neuroscience, 9th Edition. Pearson Publishing.

Cobb, Matthew (2002). Timeline: Exorcizing the animal spirits: Jan Swammerdam on nerve function. *Nature Reviews Neuroscience, 3* (5), 395-400.

Gross, Charles G. (1987). Neuroscience, Early History of. In Adelman, George (ed.), Encyclopedia of Neuroscience, Kirkhauswer Verlag AG, pp. 843-847.

Haque, Amber (2004). Psychology from Islamic perspective: Contributions of early Muslim scholars and challenges to contemporary Muslim psychologists. *Journal of Religion and Health, 43* (4), 357-377.

Klein, Stephen B. and Thorne, B. Michael (2006). Biological Psychology. Worth Publishers.

Lau, Yoke Lian, Loi, Chek Kim and bin Abdullah, Mohd Nor Azan (2021). The Historical Development of the study of Broca's Aphasia. *Malang Neurology Journal, 7* (2), 125-128.

Macmillan, Malcolm (2001). John Martyn Harlow: Obscure country physician? *Journal of the History of Neurosciences: Basic and Clinical Perspectives, 10* (2), 149-162.

Mikulic, Matej (2021). Total neuroscience funding by the National Institutes for Health (NIH). Statica.com. https://www.statista.com/statistics/...isticContainer

Mohamed, Wael MY (2008). History of Neuroscience: Arab and Muslim contributions to modern neuroscience. *IBRO History of Neuroscience*: 255.

Syed, Ibrahim B. (2002). Islamic medicine: 1000 years ahead of its times. *Journal of the International Society for the History of Islamic Medicine 2*, 2-9.

Van Laere J. Vesalius en het zenuwstelsel [Vesalius and the nervous system]. Verh K Acad Geneeskd Belg. 1993;55(6):533-76. Dutch. PMID: 8209578.

Wells, P. N. T. (2005). “Sir Godfrey Newbold Hounsfield KT CBE. 28 August 1919 – 12 August 2004: Elected F.R.S. 1975”. Biographical Memoirs of Fellows of the Royal Society. **51**: 221–235. doi:10.1098/rsbm.2005.0014.

Wickens, A. P. (2021). Introduction to Biopsychology. United Kingdom: SAGE Publications.

1.2.8 Attributions

Introduction to Psychology - 1st Canadian Edition by Jennifer Walinga is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, except where otherwise noted.

Spielman, Rose M.; Jenkins, William; and Lovett, Marilyn, "Psychology 2e" (2020). *Open Access Textbooks*. 1.
https://commons.erau.edu/oer-textbook/1

This page titled 1.2: History of Biopsychology is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

1.3: Connecting Biology to Behavior

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Discuss the Case of Phineas Gage and its contribution to biological psychology.

1.3.1 Overview

While experiments are necessary to establish cause and effect relationships, in-depth studies of unique individuals or groups of people who share an experience can be used to inform our understanding of things that we can not study experimentally. Surgical errors, extreme mistreatment, and tragic accidents are impactful events that can alter individuals significantly, providing unique opportunities to study the effects of experiences which can not be ethically studied experimentally. There have been a number of these case studies which have revealed the role of different parts of the brain on our thinking and behavior. One such case is Phineas Gage. Gage lived 12 years after a rod pierced his skull, damaging his left frontal lobe. Researchers were able to gather information about his functioning before and observe his cognitive ability and personality after the accident. His case enabled the field to understand the role of frontal lobe in personality and mental processes.

1.3.2 The Tale of Phineas Gage



Figure **1.3.1**: Phineas Gage Portrait After His Accident. (Public Domain; via Wikipedia Common)

The case of Phineas Gage is worthy of expanded coverage as his tragic accident establishes a clear connection between the brain and who we are. Gage, a 25-year-old man, was employed in railroad construction at the time of the accident. As the company's most capable employee, with a well-balanced mind and a sense of leadership, he was directing a rock-splitting workgroup while preparing the bed of the Rutland and Burlington Railroad south of Cavendish, Vermont, USA. At 4:30 PM on September 13, 1848, he and his group were blasting a rock, and Gage was assigned to put gunpowder in a deep hole inside it.

The moment he pressed the gunpowder into the hole with a bar, the friction caused sparks, and the powder exploded. The resulting blast projected the meter-long bar, which was 1.25 inches in diameter and weighed about 13.2 pounds, through his skull at high speed. The bar entered his left cheek, destroyed his eye, passed through the left front of his brain, and left his head at the top of the skull on the right side. Gage was thrown on his back and had some brief convulsions, but he woke up and spoke in a few minutes, walked with a little help, and sat in an ox cart for the 0.7-mile trip to where he was living.



Figure **1.3.2**: The injury of Phineas Gage With Frontal Lobe shown in Red. (Public Domain; via Wikimedia Common)

About 30 minutes after the accident, a doctor arrived to provide medical care. Gage had lost a lot of blood, and the next days that followed were quite difficult. The wound became infected, and Phineas was anemic and remained semi-comatose for more than two weeks. He also developed a fungal infection in the exposed brain that needed to be surgically removed. His condition slowly improved after doses of calomel and beaver oil. By mid-November he was already walking around the city.

1.3.2.1 The Consequences

For three weeks after the accident, the wound was treated by doctors. During this time, he was assisted by Dr. John Harlow, who covered the head wound and then reported the case in the Boston Medical Surgery Journal. In November 1849, invited by the professor of surgery at Harvard Medical School, Henry Jacob Bigelow, Harlow took Gage to Boston and introduced him to a meeting of the Boston Society for Medical Improvement .

In his reports, Harlow described that the physical injury profoundly altered Gage's personality. Although his memory, cognition, and strength had not been altered, his once gentle personality slowly degraded. He became a man of bad and rude ways, disrespectful to colleagues, and unable to accept advice. His plans for the future were abandoned, and he acted without thinking about the consequences. And here was the main point of this curious story: Gage became irritable, irreverent, rude and profane, aspects that were not part of his way of being. His mind had changed radically. His transformation was so great that everyone said that “Gage is no longer himself.”

As a result of this personality change, he was fired and could no longer hold a steady job. He became a circus attraction and even tried life in Chile, later returning to the United States. However, there is something still little known about Gage: his personality changes lasted for about four years, slowly reverting later. As a proof of this, he worked as a long-haul driver in Chile, a job that required considerable planning and focus skills. He died on May 21, 1861, 12 years after the accident, from an epileptic seizure that was almost certainly related to his brain injury.



Figure **1.3.3**: Frontispiece, showing multiple views of the exhumed skull, and tamping iron, of brain injury survivor Phineas Gage (Public Domain; via Wikimedia Commons

After his body was removed from its grave, Gage's mother donated his skull to Dr. Harlow who in turn donated it to Harvard University.

1.3.2.2 The Legacy

Gage's case is considered to be one of the first examples of scientific evidence indicating that damage to the frontal lobes may alter personality, emotions, and social interaction. Prior to this case, the frontal lobes were considered silent structures, without function and unrelated to human behavior. Scottish neurologist, David Ferrier, was motivated by this fact to investigate the role of frontal lobes in brain function. Ferrier removed the frontal lobes in monkeys and noted that there were no major physiological changes, but the character and behavior of the animals were altered. In other words, he confirmed the role of the frontal lobes that was suggested by Gage's accident in an experiment with a non-human animal.

Knowledge that the frontal lobe was involved with emotions continued to be studied. The surgeon Burkhardt in 1894 performed a series of surgeries in which he selectively destroyed the frontal lobes of several patients in whom he sought to control psychotic symptoms, being the modern prototype of what was later known through Antonio Egas Moniz as psychosurgery. Today, it is well understood that the prefrontal cortex of the brain controls the organization of behavior, including emotions and inhibitions.

Folkloric as it may be, but nonetheless remarkable, the contribution of Phineas Gage's case should not be overlooked, as it provided scientists the baseline for the promotion of studies in neuropsychiatry, and a source of inspiration for world medicine. In 2012, a team of neuroscientists used computer tomography of Gage's skull with typical brain MRI scans to simulate how extensive Gage's brain damage was. They confirmed that most of the damaged area was the left frontal lobe. However, surrounding areas and their neural network were also extensively severed. And it is not just the researchers who keep coming back to Gage. Medical and psychology students still learn about Gage from their history lessons. Neurosurgeons and neurologists still sometimes use Gage as a reference when evaluating certain cases. The final chapter of his life also offers us a thought-provoking discovery about cases of massive brain damage, indicating that rehabilitation may be possible.

1.3.3 Summary

Phineas Gage made a huge contribution to our understanding of the frontal lobe damage and its subsequent change in personality. Furthermore, his case expanded knowledge in neurology in several areas, including the study of brain topography in behavioral disorders, the development of psychosurgery, and finally the study of brain rehabilitation. Also, Gage's case had a tremendous influence on early neuropsychiatry. The specific changes observed in his behavior pointed to theories about the localization of brain function and correlated with cognitive and behavioral sequelae, thereby acquainting us with the role of the frontal cortex in higher-order actions such as reasoning, behavior and social cognition. In those years, while neuropsychiatry was in its infancy, Gage's extraordinary story served as one of the first pillars of evidence that the frontal lobe is involved in personality, which helped solidify his remarkable legacy in world medical history.

1.3.3.1 Attributions

Adapted from Phineas Gage’s Great Legacy by Vieira Teles Filho, Ricardo. Licensed CC BY 4.0.

This page titled 1.3: Connecting Biology to Behavior is shared under a CC BY license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

1.4: What are the Divisions of Biopsychology?

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Discuss the different divisions of biopsychology.

1.4.1 Overview

Neuroscientists study how the nervous system develops, its structure, and what it does. Biopsychology is an integrative discipline that draws knowledge from the other neuroscience disciplines. It then applies that knowledge to the study of behavior and mental processes. In this section, we will examine a few of the disciplines of neuroscience that are of particular interest to biopsychology.

1.4.2 Biopsychology - An Integrative Discipline

Neuroscience is the scientific study of the nervous system. The more neuroscientific research progresses, the more clearly it is established that human behavior and mental processes—the key interests for psychological study—are intimately intertwined with activity in the brain. It encompasses the branch of biology that deals with the anatomy, biochemistry, molecular biology, and physiology of neurons and neural circuits. It also encompasses cognition (thinking) and human behavior.

Our focus, biopsychology is a subset of neuroscience that focuses on behavior and mental processes. Consequently, one of the many names for biopsychology is "behavioral neuroscience". Biopsychology aims to understand the biological basis of human behavior and mental processes. It is an integrative discipline that draws knowledge from different divisions of neuroscience. These divisions include but are not limited to:

- **Neurophysiology** - the study of the function of the nervous system, focusing on neurons (nerve cells) and their communication.
- **Neuroanatomy** - the study of the anatomy or structure of nervous tissue and neural structures of the nervous system.
- **Neuropharmacology** - the study of how drugs affect cellular function in the nervous system.
- **Developmental neuroscience** - aims to describe the cellular basis of brain development and to address the underlying mechanisms. The field draws on both neuroscience and developmental biology to provide insight into the cellular and molecular mechanisms by which complex nervous systems develop.
- **Cognitive neuroscience** - concerned with the scientific study of biological structures and processes underlying cognition, with a focus on the neural elements of mental processes.
- **Computational neuroscience** - includes both the study of the information processing functions of the nervous system, and the use of digital computers to study the nervous system. It is an interdisciplinary science that links the diverse fields of neuroscience, cognitive science and psychology, electrical engineering, computer science, physics and mathematics.
- **Neurology** - the medical specialty dealing with disorders of the nervous system. It deals with the diagnosis and treatment of all categories of disease involving the central, peripheral, and autonomic nervous systems.
- **Social neuroscience** uses the brain and body to understand how we think and act, with a focus on how we think about and act toward other people. It is an interdisciplinary field that uses a range of neuroscience measures to understand how other people influence our thoughts, feelings, and behavior. (We will examine this branch more in depth in Cognition and Intelligence Chapter)
- **Educational neuroscience or neuroeducation**, seeks to link an understanding of genetics and brain processes with how people learn, for the purpose of informing classroom instruction and experiences.

1.4.3 Summary

In these sections, you have learned about the disciplines who have contributed to biopsychology in its aim to understand the biological aspect of behavior and mental processes. Biopsychology does not stand on its own. Disciplines such as neuroanatomy contribute to our understanding of the different structures of the nervous system, including the brain. Neuroanatomy has helped with our understanding of how some of the major parts of the brain control and contribute to different behaviors. As each of these

disciplines expand, biopsychology is able to draw information from each and develop.

Attributions

- Outline of Neuroscience; Wikipedia; licensed CC BY-SA 3.0
- Ito, T. A. & Kubota, J. T. (2021). Social neuroscience. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology.* Champaign, IL: DEF publishers. (CC BY-NC-SA)
- Boundless Neuroscience; licensed CC BY-SA 4.0

This page titled 1.4: What are the Divisions of Biopsychology? is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

CHAPTER OVERVIEW

2: Research Methods and Ethical Considerations of Biological Psychology and Neuroscience

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

2.1: Introduction to Research in Biological Psychology
2.2: Non-Invasive Techniques: Structural Imaging
2.3: Non-Invasive Techniques: Direct Functional Imaging Techniques
2.4: Non-Invasive Techniques: Indirect Functional Imaging Techniques
2.5: Other Non-Invasive Techniques
2.6: Invasive Techniques
2.7: Ethical Considerations

This page titled 2: Research Methods and Ethical Considerations of Biological Psychology and Neuroscience is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

2.1: Introduction to Research in Biological Psychology

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Differentiate between the purpose of structural and functional imaging.
2. Describe the differences between spatial resolution and temporal resolution.

2.1.1 Overview

This chapter will describe the various ways that biological psychologists study the brain. There are many ways to categorize the techniques that are used when studying the brain. We will start by covering the non-invasive techniques, where we are able to study the brain without getting direct physical access to the brain (think of fixing a broken pipe in a wall without having to open the wall up). Then we will move into the invasive techniques, where we study the brain by having direct access (an example would be fixing a broken pipe in a wall by tearing a hole in the wall). Then we will discuss various neuropsychological techniques, where we learn about the brain using people with some sort of brain "issue." For example, people with epilepsy have been extensively studied and we can learn a lot about how the brain works from them. Finally, the last section will address ethical considerations of biological psychology research.

In this section, we will start with a discussion of many modern day neuroscience terms that are important for understanding the different techniques in future sections.

2.1.2 The Terminology of Modern Research Techniques

We have come a long way since Phineas Gage with how we study the brain. Many techniques now allow us to understand how the brain works without waiting for a horrific accident to occur or conducting some sort of surgery (although, as you will see, we still use surgical techniques to study the brain). Techniques have been developed that allow us to see what the brain looks like, as a still image (structurally) or in action (functionally).

2.1.2.1 Structural Imaging

Structural imaging techniques are useful in many situations such as locating tumors, sites of physical brain damage, or finding size differences between the structures of the brain between various groups. Magnetic resonance imaging (MRI), for example, is one such technique that is commonly used to study the brain and to diagnosis knee and shoulder injuries. Structural imaging techniques allow us to look inside the brain (or body) without having to go inside.

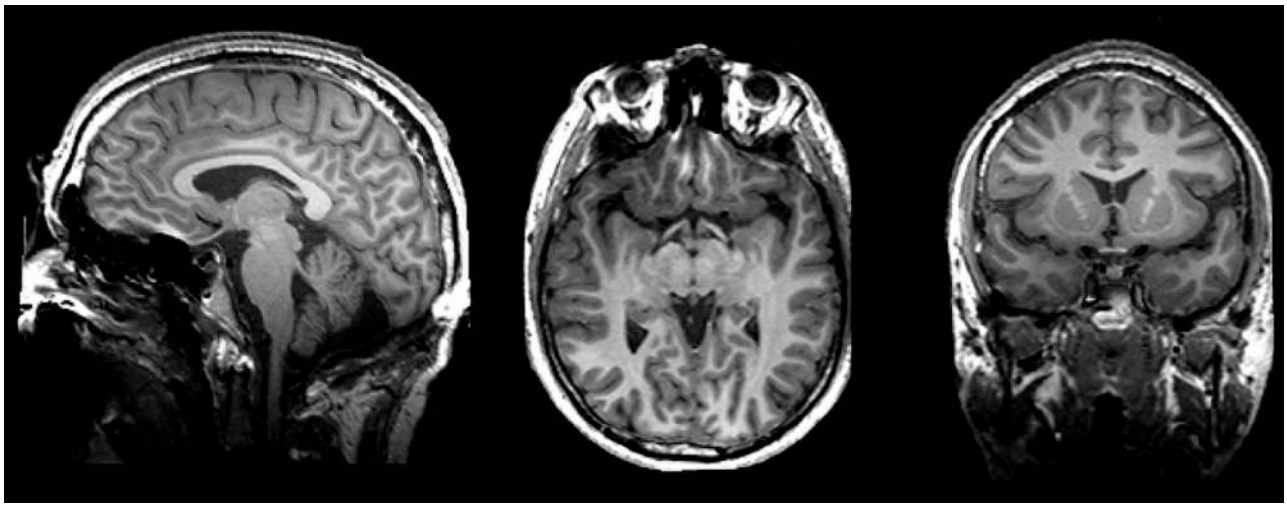


Figure **2.1.1**: An MRI (A structural image) of the brain. Note the ability to see specific gyri (bumps) and sulci (grooves).

A series of MRI images can be used to create a picture of the brain.

2.1.2.2 Functional Imaging

Many researchers are also interested in how the brain works. Some studies begin with the scientific question of “what does this part do?” Or more commonly, “Where in the brain does this happen?” **Functional imaging** techniques allow researchers to learn about the brain activity during various tasks by creating images based on the electrical activity or the absorption of various substances that occurs while a subject is engaging in a task. Such techniques can be used, for example, to visualize the parts of the brain that respond when we're exposed to stimuli that upset us or make us happy.

2.1.2.3 Temporal Versus Spatial Resolution

Within functional imaging techniques, researchers are frequently focused on one of two questions. They may ask “When does this activity occur?” Or “Where does this activity occur?” Some techniques are better for answering one of these questions, whereas other techniques are better for answering the other question. We describe how well a technique can determine *when* the activity has occurred as **temporal resolution.** For example, was the brain region activity occurring sometime in the last hour, the last minute, the last second, or within milliseconds? While some techniques are excellent at determining precisely when the activity occurred and other techniques are quite terrible at it. Additionally, we can describe how well a technique can determine *where* the activity has occurred as **spatial resolution.** For example, did the activity occur in the temporal lobe somewhere or can we narrow that down to a specific gyrus (ridge) or sulcus (groove) of the cerebral cortex? If it occurred on a particular gyrus can we narrow it down to a particular portion of that gyrus? As with temporal resolution, some techniques are excellent at determining precisely where the activity occurred whereas other techniques are less accurate.

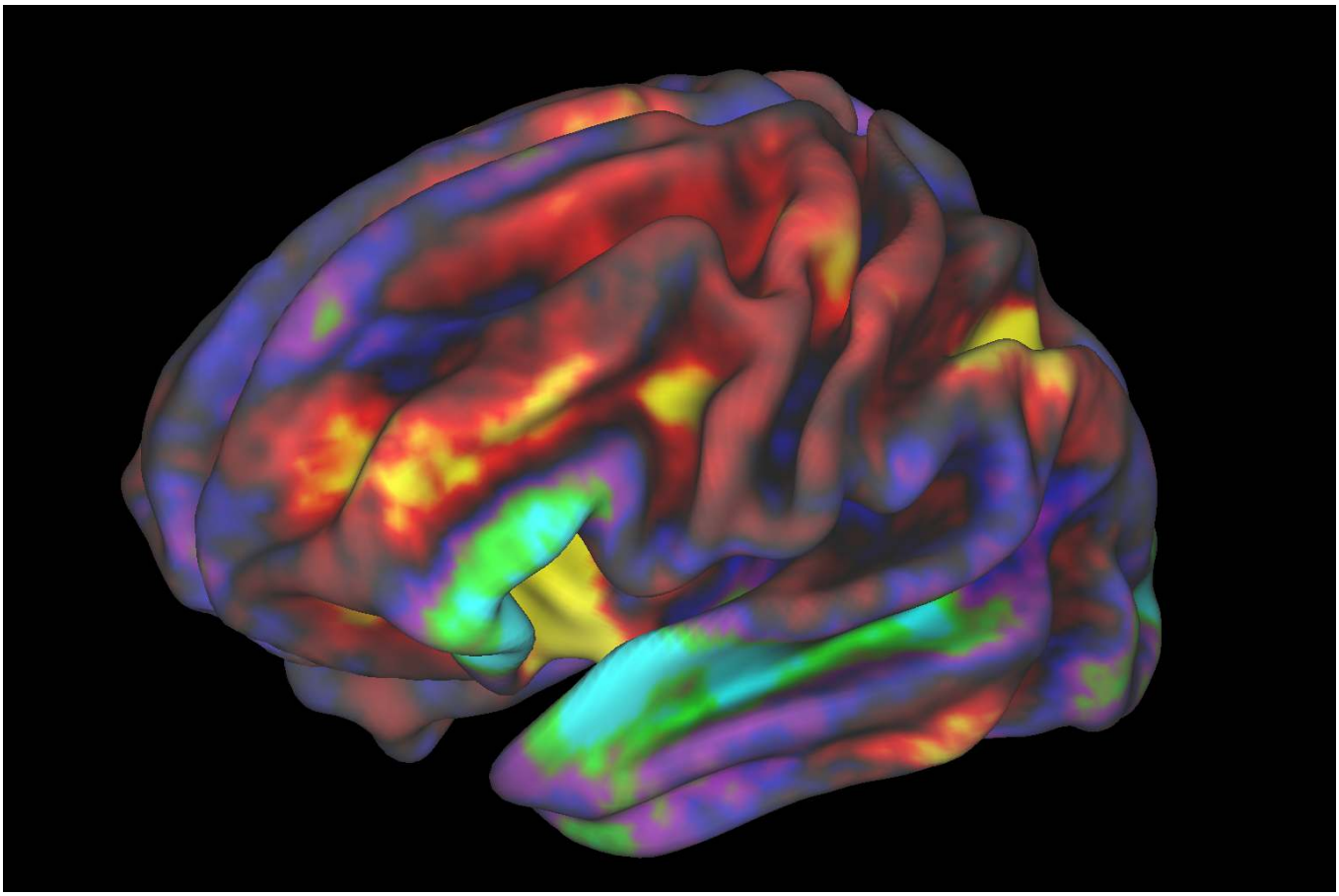


Figure **2.1.2**: Functional MRI of a teenage brain conducting a working memory test. The regions in red and yellow, primarily in the frontal lobe and occipital lobe, are the most active.

2.1.3 Summary

How we study the brain has come a long way since the days of Phineas Gage. Although, as we will discuss, we still learn about the brain from accidents and other traumatic brain events, we have a variety of other techniques that we can use now to study the brain in healthy individuals. These techniques allow us to answer questions about what the brain looks like, what specific parts do, and when they do it.

2.1.4 Attributions

- Psychophysiological Methods in Neuroscience by Zachary Infantolino and Gregory A. Miller is licensed under a CC BY-NC-SA 4.0 International License.
- Figure **2.1.1**: "Deadstar's brain" by Daniele Oberti is licensed under CC BY-NC-ND 2.0
- Figure **2.1.2**: "fMRI Image of Preteen Brain" by NIH Image Gallery is licensed under CC BY-NC 2.0

This page titled 2.1: Introduction to Research in Biological Psychology is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

2.2: Non-Invasive Techniques: Structural Imaging

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Explain how X-rays, CT scans, and MRI scans differ.
2. Describe the pros and cons of the three main structural imaging techniques.

2.2.1 Overview

Structural imaging techniques typically come in three different options: X-rays, computed tomography (CT) or computed axial tomography (CAT) scans, and magnetic resonance imaging (MRI) scans. Each uses different types of technology to provide a representation of a structure without having to remove the skin or bone that protects that structure. Each of these non-invasive techniques has advantages and disadvantages.

2.2.2 **X-Rays**

German physicist Wilhelm Röntgen (1845–1923) was experimenting with electrical current when he discovered that a mysterious and invisible “ray” would pass through his flesh but leave an outline of his bones on a screen coated with a metal compound. In 1895, Röntgen made the first durable record of the internal parts of a living human: an “X-ray” image (as it came to be called) of his wife’s hand. Scientists around the world quickly began their own experiments with X-rays, and by 1900, X-rays were widely used to detect a variety of injuries and diseases. In 1901, Röntgen was awarded the first Nobel Prize for physics for his work in this field.

The **X-ray** is a form of high energy electromagnetic radiation with a short wavelength capable of penetrating solids and ionizing gases. As they are used in medicine, X-rays are emitted from an X-ray machine and directed toward a specially treated metallic plate placed behind the patient’s body. The beam of radiation results in darkening of the X-ray plate. X-rays are slightly impeded by soft tissues, which show up as gray on the X-ray plate, whereas hard tissues, such as bone, largely block the rays, producing a light-toned “shadow.” **Thus, X-rays are best used to visualize hard body structures such as teeth and bones**. Figure 2.2.1 depicts an X-ray of a knee. Like many forms of high energy radiation, however, X-rays are capable of damaging cells and initiating changes that can lead to cancer. This danger of excessive exposure to X-rays was not fully appreciated for many years after their widespread use.



Figure 2.2.1: X-Ray of a Knee: High energy electromagnetic radiation allows the internal structures of the body, such as bones, to be seen in X-rays like these.

Due to the development of other techniques that are considerably better at looking at soft tissue, X-rays are largely not used for studying the brain anymore.

2.2.3 Modern Medical Imaging

X-rays can depict a two-dimensional image of a body region, and only from a single angle. In contrast, more modern medical imaging technologies produce data that are integrated and analyzed by computers to produce three-dimensional (3D) images or images that reveal aspects of the body functioning.

Computed Tomography

Tomography refers to imaging by sections. **Computed (or computerized) tomography (CT)** is a noninvasive imaging technique that uses computers to analyze several cross-sectional X-rays in order to reveal small details about structures in the body. The technique was invented in the 1970s and is based on the principle that, as X-rays pass through the body, they are absorbed or reflected at different levels. In the technique, a patient lies on a motorized platform while a computerized axial tomography (CAT) scanner rotates 360 degrees around the patient, taking X-ray images. Figure **2.2.2** shows a CT scanner with a platform for the subject to lie on. A computer combines these images into a two-dimensional view of the scanned area, or "slice." Figure **2.2.3** shows a series of slices of the brain for one subject.



Figure **2.2.2**: A CT scanner at the University of Pittsburg Medical Center East.

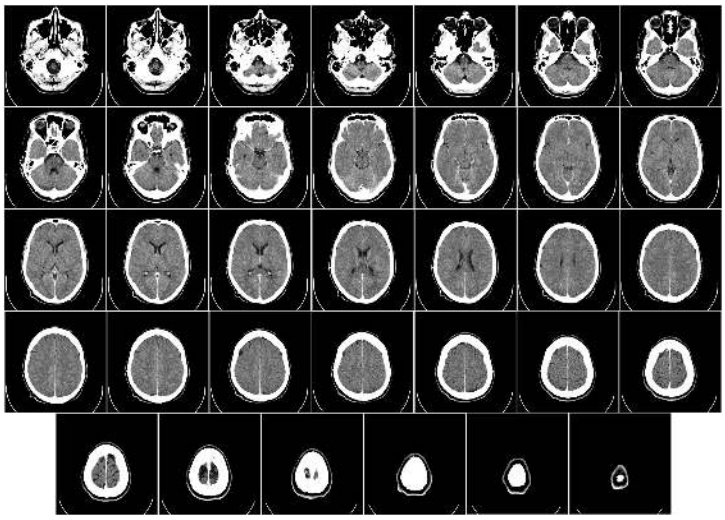


Figure **2.2.3**: A series of axial CT scans of the brain of one subject. Each image is a slice of the brain starting with a bottom slice and incrementally moving to higher and higher slices.

Since 1970, the development of more powerful computers and more sophisticated software has made CT scanning routine for many types of diagnostic evaluations. It is especially useful for soft tissue scanning, such as of the brain and the thoracic and abdominal viscera. Its level of detail is so precise that it can allow physicians to measure the size of a mass down to a millimeter. The main disadvantage of CT scanning is that it exposes patients to a dose of radiation many times higher than that of X-rays. Whether this is particularly dangerous is still being debated (McCollough et al., 2015).

2.2.3.1 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is a noninvasive medical imaging technique based on a phenomenon of nuclear physics discovered in the 1930s, in which matter exposed to magnetic fields and radio waves was found to emit radio signals. In 1970, a physician and researcher named Raymond Damadian noticed that malignant (cancerous) tissue gave off different signals than normal body tissue. He applied for a patent for the first MRI scanning device, which was in use clinically by the early 1980s. The early MRI scanners were crude, but advances in digital computing and electronics led to their advancement over any other technique for precise imaging, especially to discover tumors. MRI also has the major advantage of not exposing patients to radiation.

Drawbacks of MRI scans include their much higher cost, and patient discomfort with the procedure. The MRI scanner subjects the patient to such powerful electromagnets that the scan room must be shielded. The patient must be enclosed in a metal tube-like device for the duration of the scan, sometimes as long as thirty minutes, which can be uncomfortable and impractical for ill patients. The device is also so noisy that, even with earplugs, patients can become anxious or even fearful. These problems have been overcome somewhat with the development of “open” MRI scanning, which does not require the patient to be entirely enclosed in the metal tube. Figure **2.2.4** shows an MRI machine with a platform for the patient to lie on. Patients with iron-containing metallic implants (internal sutures, some prosthetic devices, and so on) cannot undergo MRI scanning because it can dislodge these implants.



Figure **2.2.4**: An MRI machine.

2.2.3.2 Using Structural Imaging Techniques to Study a Disorder: Autism Spectrum Disorder

One example that we will use throughout this chapter is that of how we use these research techniques to study Autism Spectrum Disorder (ASD). ASD is a developmental disorder frequently characterized by issues including various combinations of interaction issues, communication difficulties, and even repetitive behaviors. Throughout each section, we will discuss some of the ways the main tools of brain research have been used to examine this disorder.

Structural imaging techniques with ASD have focused on which brain structures have physical differences. MRIs have found a thicker frontal cortex (Carper & Courchesne, 2005) and a thinner temporal cortex (Hardan et al., 2006) in patients with ASD. These areas are notable because the frontal cortex is linked to communication and language abilities and the temporal cortex is linked to auditory processing (ie. language input), both of which are issues that many with ASD struggle with.

2.2.4 Summary

Structural imaging techniques, including X-rays, CT scans, and MRIs, allow researchers to get a look at what the brain looks like without having to do anything invasive to the patient, like surgery. These techniques, particularly CT scans and MRIs are extremely useful in constructing an image of what the brain looks like and allow doctors to detect any structural abnormalities in their patients. They also allow researchers to learn about the sizes of different structures in the brain and possibly correlate those differences to various functions.

2.2.5 References

McCollough, C. H., Bushberg, J. T., Fletcher, J. G., & Eckel, L. J. (2015). Answers to common questions about the use and safety of CT scans. *Mayo Clinic Proceedings*, *90*(10), 1380–1392.

Carper, R. A., & Courchesne, E. (2005). Localized enlargement of the frontal cortex in early autism. *Biological Psychiatry*, *57*(2), 126–133.

Hardan, A. Y., Muddasani, S., Vemulapalli, M., Keshavan, M. S., & Minshew, N. J. (2006). An MRI study of increased cortical thickness in autism. *American Journal of Psychiatry*, *163*(7), 1290–1292.

2.2.6 Attributions

Anatomy & Physiology by Lindsay M. Biga, Sierra Dawson, Amy Harwell, Robin Hopkins, Joel Kaufmann, Mike LeMaster, Philip Matern, Katie Morrison-Graham, Devon Quick & Jon Runyeon is licensed under a CC BY-SA 4.0 International License.

Figure **2.2.1**: "Plain radiograph of the right knee" by Ptrump16 is licensed under CC BY-SA 4.0

Figure **2.2.2**: "New UPMC East" by Davey Nin is licensed under CC BY 2.0.

Figure **2.2.3**: "Computed tomography of human brain - large - CT scan" is in the Public Domain, CC0

Figure **2.2.4**: "MRI" by Liz West is licensed under CC BY 2.0

This page titled 2.2: Non-Invasive Techniques: Structural Imaging is shared under a CC BY-SA license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

2.3: Non-Invasive Techniques: Direct Functional Imaging Techniques

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Apply the terms spatial and temporal resolution to EEG and MEG.
2. In basic terms, describe EEG and MEG.
3. Describe the key characteristic of direct functional imaging techniques.

2.3.1 Overview

In this section, we will discuss the two main direct functional imaging techniques, electroencephalography (EEG) and magnetoencephalography (MEG). We will also generally discuss what makes a technique a direct brain imaging technique.

2.3.2 EEG

Electroencephalography (EEG) is one technique for studying brain activity. This technique uses at least two and up to 256 electrodes to measure the difference in electrical charge (the **voltage**) between pairs of points on the head. These electrodes are typically fastened to a flexible cap (similar to a swimming cap) that is placed on the participant's head. Figure **2.3.1** shows a patient wearing such a cap. From the scalp, the electrodes measure the electrical activity that is naturally occurring within the brain. They do not introduce any new electrical activity.



Figure **2.3.1**: A subject with a group of EEG electrodes on their head.

Given that this electrical activity must travel through the skull and scalp before reaching the electrodes, localization of activity is less precise when measuring from the scalp, but it can still be within several millimeters when localizing activity that is near the scalp. While EEG is lacking with respect to spatial resolution, one major advantage of EEG is its temporal resolution. Data can be recorded thousands of times per second, allowing researchers to document events that happen in less than a millisecond. EEG analyses typically investigate the change in amplitude (wave height) or frequency (number of waves per unit of time) components of the recorded EEG on an ongoing basis or averaged over dozens of trials (see Figure **2.3.2**). The EEG has been used extensively in the study of sleep. When you hear references to "brain waves", those are references to information obtained using EEG.

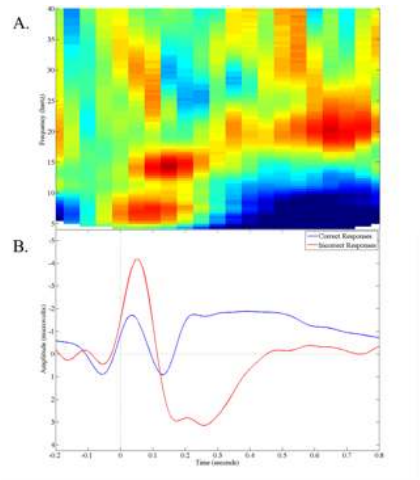


Figure **2.3.2**: Example of EEG analysis output. Panel A represents changes in the relative strength of different frequencies in the EEG data over time. Panel B represents changes in the amplitude in the instantaneous EEG voltage over time.

2.3.3 MEG

Magnetoencephalography (MEG) is another technique for noninvasively measuring neural activity. The flow of electrical charge (the current) associated with neural activity produces very weak magnetic fields that can be detected by sensors placed near the participant's scalp. Figure **2.3.3** depicts a subject in an MEG machine. The number of sensors used varies from a few to several hundred. Due to the fact that the magnetic fields of interest are so small, special rooms that are shielded from magnetic fields in the environment are needed in order to avoid contamination of the signal being measured. MEG has the same excellent temporal resolution as EEG. Additionally, MEG is not as susceptible to distortions from the skull and scalp. Magnetic fields are able to pass through the hard and soft tissue relatively unchanged, thus providing better spatial resolution than EEG. MEG analytic strategies are nearly identical to those used in EEG. However, the MEG recording apparatus is much more expensive than EEG, so MEG is much less widely available.



Figure **2.3.3**: A subject/patient getting an MEG.

2.3.4 General Information About Direct Imaging Techniques

EEG and MEG both have excellent temporal resolution and are useful when someone is particularly interested in studying the timing of brain activity. For example, if someone is reading a sentence that ends with an unexpected word, how long after reading the unexpected word does the brain react to it? In addition to these types of questions, EEG and MEG methods allow researchers to investigate the degree to which different parts of the brain "talk" to each other. This allows for a better understanding of brain networks, such as their role in different tasks and how they may function abnormally in psychopathology.

Direct imaging techniques are those that allow for a direct measure of brain activity. EEG and MEG are both considered direct brain imaging techniques since EEG measures the electrical activity from groups of neurons and MEG measures the magnetic fields that the electrical activity gives off. Neither of these techniques relies on measuring something else with an assumption that they are linked. This is not true in the next set of techniques we will discuss.

2.3.4.1 Using Direct Functional Imaging Techniques to Study a Disorder: Autism Spectrum Disorder

EEG and MEG have been used to examine ASD. One of the findings included a delay in the brain wave associated with auditory stimuli. In short, there are differences in the time for processing auditory sounds in children with ASD compared to those without ASD. Furthermore, this delay appears more pronounced in children with ASD who have language developmental delays as opposed to children with ASD without linguistic delays (Roberts et al., 2019). This delay has even been proposed to help clinicians diagnose autism in young children.

2.3.5 Summary

Direct imaging techniques are extremely useful ways to measure electrical brain activity in a non-invasive way. Both EEG and MEG are most useful at identifying differences in timing patterns of electrical activity.

2.3.6 References

Roberts, T. P., Matsuzaki, J., Blaskey, L., Bloy, L., Edgar, J. C., Kim, M., Ku, M., Kuschner, E. S., & Embick, D. (2019). Delayed M50/M100 evoked response component latency in minimally verbal/nonverbal children who have autism spectrum disorder. *Molecular Autism*, *10*(1).

2.3.7 Attributions

- Psychophysiological Methods in Neuroscience by Zachary Infantolino and Gregory A. Miller is licensed under a CC BY-NC-SA 4.0 International License.
- Figure **2.3.1**: "EEG Brain Scan" by Tim Sheerman-Chase is licensed under CC BY 2.0
- Figure **2.3.3**: "170321-F-LW859-017" by Airman Magazine is licensed under CC BY-NC 2.0

This page titled 2.3: Non-Invasive Techniques: Direct Functional Imaging Techniques is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

2.4: Non-Invasive Techniques: Indirect Functional Imaging Techniques

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe the key characteristics of indirect functional brain imaging techniques.
2. In basic terms, describe functional MRI (fMRI) and positron emission tomography (PET).
3. Discuss the pros and cons of FMRI and PET.

2.4.1 Overview

EEG and MEG are direct functional imaging techniques as they measure the actual activity in the brain. In this section, we will discuss what makes a technique an *indirect* brain imaging technique and the two main indirect imaging techniques, functional MRI (fMRI) and positron emission tomography (PET), will be introduced.

2.4.2 FMRI and PET

Indirect brain imaging techniques rely on an assumption that activity in the brain correlates to something else that we are able to measure. In these cases, these techniques measure blood flow in the brain. The assumption is that blood flow in the brain is related to the activity level in that area of the brain. Of course, with any assumption, there is always the risk that it could be wrong. Thankfully there is extensive research examining this assumption and the scientific consensus currently is that blood flow is an appropriate indication of brain activity. The two main indirect brain imaging techniques that we will cover are functional MRI (fMRI) and positron emission tomography (PET).

Functional magnetic resonance imaging (fMRI) is a method that is used to assess changes in the activity of tissue, such as measuring changes in neural activity in different areas of the brain during thoughts or experiences. This technique builds on the principles of structural MRI techniques and also uses the property that, when neurons fire, they use energy, which must be replenished. Glucose and oxygen, two key components for energy production, are supplied to the brain from the blood stream as needed. Oxygen is transported through the blood using hemoglobin, which contains binding sites for oxygen. When these sites are saturated with oxygen, it is referred to as oxygenated hemoglobin. When the oxygen molecules have all been released from a hemoglobin molecule, it is known as deoxygenated hemoglobin. As a set of neurons begin firing, oxygen in the blood surrounding those neurons is consumed, leading to a reduction in oxygenated hemoglobin. The body then compensates and provides an abundance of oxygenated hemoglobin in the blood surrounding that activated neural tissue. When activity in that neural tissue declines, the level of oxygenated hemoglobin slowly returns to its original level, which typically takes several seconds. Figure **2.4.1** shows a subject about to go into a functional MRI machine.



Figure **2.4.1**: A Functional MRI machine. While it is pictured here with the patient outside of the magnet, to scan the patient, they would be moved into the circular magnet.

fMRI measures the change in the concentration of oxygenated hemoglobin, which is known as the blood-oxygen-level-dependent (BOLD) signal. This leads to two important facts about fMRI. First, fMRI measures blood volume and blood flow, and from this we infer neural activity; as stated previously, fMRI does not measure neural activity directly. Second, fMRI data typically have poor temporal resolution; however, when combined with structural MRI, fMRI provides excellent spatial resolution. Temporal resolution for fMRI is typically on the order of seconds, whereas its spatial resolution is on the order of millimeters. Generally speaking, under most conditions there is an inverse relationship between temporal and spatial resolution—one can increase temporal resolution at the expense of spatial resolution and vice versa. In other words, as one increases the other decreases.

This method is valuable for identifying specific areas of the brain that are associated with different physical or psychological tasks. Clinically, fMRI may be used prior to neurosurgery in order to identify the brain areas that are associated with language so that the surgeon can avoid those areas during the operation. fMRI allows researchers to identify differential or convergent patterns of activation associated with tasks. For example, if participants are shown words on a screen and are expected to indicate the color of the letters, are the same brain areas recruited for this task if the words have emotional content or not? Does this relationship change in psychological disorders such as anxiety or depression? Is there a different pattern of activation even in the absence of obvious performance differences? fMRI is an excellent tool for comparing brain activation in different tasks and/or populations. Figure **2.4.2** provides an example of results from fMRI analyses overlaid on a structural MRI image. The blue and orange shapes represent areas with significant changes in the BOLD signal, thus changes in neural activation.

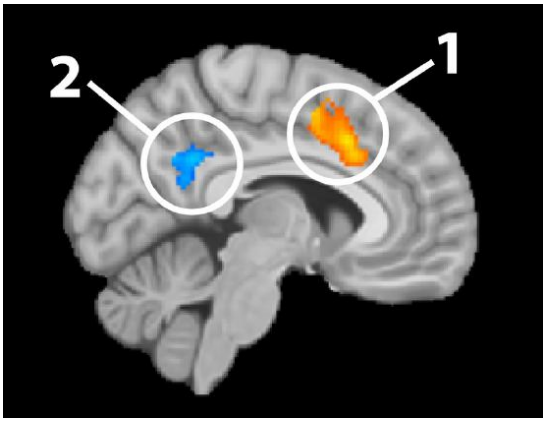


Figure **2.4.2**: Example of fMRI analyses overlaid on an sMRI image. Area 1 (orange) indicates an increase in the BOLD signal, and Area 2 (blue) indicates a decrease in the BOLD signal. We infer that neural activity increased in the Area 1 and decreased in Area 2.

Positron emission tomography (PET) is a medical imaging technique that is used to measure processes in the body, including the brain (see Figure **2.4.3** for a PET scanner). This method relies on a positron-emitting tracer atom that is introduced into the blood

stream in a biologically active molecule, such as glucose, water, or ammonia. A positron is a particle much like an electron but with a positive charge. One example of a biologically active molecule is fludeoxyglucose, which acts similarly to glucose in the body. Fludeoxyglucose will concentrate in areas where glucose is needed—commonly areas with higher metabolic (energy) needs. Over time, this tracer molecule emits positrons, which are detected by a sensor. The spatial location of the tracer molecule in the brain can be determined based on the emitted positrons. This allows researchers to construct a three-dimensional image of the areas of the brain that have the highest metabolic needs, typically those that are most active. Images resulting from PET usually represent neural activity that has occurred over tens of minutes, which is very poor temporal resolution for some purposes. PET images are often combined with computed tomography (CT) images to improve spatial resolution, as fine as several millimeters. Tracers can also be incorporated into molecules that bind to neurotransmitter receptors, which allow researchers to answer some unique questions about the action of neurotransmitters. Unfortunately, very few research centers have the equipment required to obtain the images or the special equipment needed to create the positron-emitting tracer molecules, which typically need to be produced on site.



Figure **2.4.3**: A PET scanner.

2.4.2.1 Using Indirect Functional Imaging Techniques to Study a Disorder: Autism Spectrum Disorder

PET and fMRI studies of ASD have found different levels of neuronal activity in the amygdala and the hippocampus compared to subjects without ASD. These areas are notable because they are a part of the “social brain.” These studies have largely focused on patients with ASD when they are viewing faces. As the viewing of faces is a large part of socializing (for example, reading expressions and making eye contact) and socializing is one area where many autistic patients have issues, these studies help provide further information for doctors and researchers to use. (See Philip et al. (2012) for a review of the fMRI studies of ASD.)

2.4.3 Summary

The use of indirect functional imaging techniques has allowed researchers and doctors to see which parts of the brain are active during various tasks. Both fMRI and PET allow researchers to measure blood flow in order to make conclusions about changes in brain activity. These techniques have excellent spatial resolution, but poor temporal resolution.

2.4.4 References

Philip, R. C. M., Dauvermann, M. R., Whalley, H. C., Baynham, K., Lawrie, S. M., & Stanfield, A. C. (2012). A systematic review and meta-analysis of the fmri investigation of Autism Spectrum Disorders. *Neuroscience & Biobehavioral Reviews*, *36*(2), 901–942.

2.4.5 Attributions

- Psychophysiological Methods in Neuroscience by Zachary Infantolino and Gregory A. Miller is licensed under a CC BY-NC-SA 4.0 International License.
- Figure **2.4.1**: "Neurodegenerative Diseases and Cognition" by Berkeley Lab is licensed under CC BY-NC-ND 2.0
- Figure **2.4.3**: "NMCSD Nuclear Medicine Department Conducts PET Scan" is in the Public Domain

This page titled 2.4: Non-Invasive Techniques: Indirect Functional Imaging Techniques is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

2.5: Other Non-Invasive Techniques

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe how transcranial magnetic stimulation (TMS) is different from other functional imaging techniques.
2. Differentiate between depolarization and hyperpolarization with respect to TMS.

2.5.1 Overview

This section is frequently combined with one of the previous sections in many textbooks because many of the researchers who do functional brain imaging work also conduct TMS work (the subject of this section). However, this technique is different from the previous techniques in many important ways. In this section, we will discuss what TMS is, when it is used, and how it is different from the previous imaging techniques.

2.5.2 Transcranial Magnetic Stimulation

Another technique that is worth mentioning is transcranial magnetic stimulation (TMS). TMS is a noninvasive method that causes **depolarization** or **hyperpolarization** in neurons near the scalp. In TMS, a coil of wire is placed just above the participant's scalp (as shown in Figure **2.5.1**). When electricity flows through the coil, it produces a magnetic field. This magnetic field travels through the skull and scalp and affects neurons near the surface of the brain. When the magnetic field is rapidly turned on and off, a current is induced in the neurons, leading to depolarization or hyperpolarization, depending on the number of magnetic field pulses. Single- or paired-pulse TMS depolarizes site-specific neurons in the cortex, causing them to fire. If this method is used over primary motor cortex, it can produce or block muscle activity, such as inducing a finger twitch or preventing someone from pressing a button. If used over primary visual cortex, it can produce sensations of flashes of light or impair visual processes. This has proved to be a valuable tool in studying the function and timing of specific processes such as the recognition of visual stimuli. Repetitive TMS produces effects that last longer than the initial stimulation. Depending on the intensity, coil orientation, and frequency, neural activity in the stimulated area may be either attenuated or amplified. Used in this manner, TMS is able to explore **neural plasticity**, which is the ability of connections between neurons to change. This has implications for treating psychological disorders as well as understanding long-term changes in neuronal excitability.

Note that TMS is different from the previous techniques in that we are not taking images of what the brain is doing. TMS disrupts or stimulates the brain and actively changes what the brain is doing.



Figure **2.5.1**: Woman with TMS wand pressed against the back right side of her head.

Figure 2.5.1: Woman with TMS wand pressed against the back right side of her head. Credit: "TMS" by Baburov is licensed under CC BY-SA 4.0

2.5.2.1 Using Transcranial Magnetic Imaging to Study a Disorder: Autism Spectrum Disorder

TMS studies, as with most research techniques, can come in the form of basic research (research intended to inform our understanding) and applied research (research intended to solve a problem). Basic research in neuroscience is typically driven by research questions aimed at a general understanding of how the brain and nervous system work. Some TMS studies have used TMS to reduce brain activity in the right amygdala during the processing of faces with negative emotions (Baeken et al., 2010). Although this research wasn't specific to autism, it is not hard to see the connection between understanding how the amygdala works and ASD. Furthermore, studies have tried to use TMS to treat ASD. Studies thus far have focused on using TMS to change activity levels and possibly stimulate neural plasticity. There was even a transcranial magnetic stimulation therapy for autism conference held in 2014 to discuss the use of the tool in the treatment of ASD. Indeed, there are myriad of possibilities for how this tool can be used in the future. (See Oberman et al. (2015) for a review of TMS treatments for ASD.)

2.5.3 Summary

TMS is a technique where researchers and doctors are able examine the changes in function that are produced when a brain area is excited or inhibited. For research purposes, it has been useful to help us map the brain. For clinical purposes, TMS is being studied more and more to understand whether it can cause alterations in the neural networks that can lead to benefits for people with various disorders.

2.5.4 References

Baeken, C., De Raedt, R., Van Schuerbeek, P., Vanderhasselt, M. A., De Mey, J., Bossuyt, A., & Luypaert, R. (2010). Right prefrontal HF-rtms attenuates right amygdala processing of negatively valenced emotional stimuli in healthy females. *Behavioural Brain Research*, *214*(2), 450–455.

Oberman, L. M., Enticott, P. G., Casanova, M. F., Rotenberg, A., Pascual-Leone, A., & McCracken, J. T. (2015). Transcranial magnetic stimulation in autism spectrum disorder: Challenges, promise, and roadmap for future research. *Autism Research*, *9*(2), 184–203.

2.5.5 Attributions

- Figure 2.5.1: "TMS" by Baburov is licensed under CC BY-SA 4.0

This page titled 2.5: Other Non-Invasive Techniques is shared under a CC BY-SA license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

2.6: Invasive Techniques

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe what single cell recordings, lesion studies, direct cortical stimulation, split-brain studies, and the Wada procedure each refer to.
2. Explain the pros and cons of single cell recordings.
3. Describe when lesion studies, direct cortical stimulation, split-brain studies, and the Wada procedure may be used.
4. Discuss the limitations of lesion/surgery studies.

2.6.1 Overview

In this section, we will discuss some of the other ways that researchers and doctors study the brain. These techniques are different from what was previously discussed in that they are more invasive, techniques that require entering the brain as opposed to taking measurements from the skull. Each of the following techniques are not used on everyday "healthy" human volunteers. Rather, these techniques are typically used when there is something going on in the brain of interest and doctors/researchers need to investigate.

2.6.1.1 Single Cell Recordings

One technique that is used to study animals in neuroscience, known as **single cell recordings** allows for us to record the activity of a cell, at least in theory. The idea of single cell recordings is that we can place a very tiny recording device, known as a microelectrode, into a single neuron and then we can try and figure out what will “activate” that particular neuron. For example, in the visual system, you may find a neuron that activates when a line moves in a certain direction in a certain location. We would then conclude that this neuron processes moving lines from a particular location.

Furthermore, single cell recordings have excellent spatial and temporal resolution. The researcher can tell exactly where the activity is coming from and exactly when the activity is occurring.

However, single cell recordings are usually extracellular (outside of the cell). That is, they don’t record from inside a single cell but, rather, they record from outside a few cells. Also, consider that the neuron that responds to a line in a particular location that is moving in a particular direction likely does not respond to much else. So, it is extremely difficult to determine what exactly each cell does through single cell recordings. Recording from one area ignores what is happening everywhere else in the brain.

2.6.1.2 Lesion Studies

A **lesion** is a site of damage in the brain. In neuroscience, we conduct lesion studies on both animals and human subjects. In animals, lesions can be made in a specific area by the researcher. Researchers are able to correlate the deficits in function with the area of damage. For example, if a researcher damages area X, and now the animal is unable to enter into REM (rapid eye movement) sleep, one can reasonably conclude that area X serves some function related to REM sleep. Although the same can be said for lesion studies of humans, accidents, or medical necessities are generally the source of human lesion subjects. You'll recall that we began this chapter by mentioning the tragic - but educational - case of Phineas Gage.

Lesion studies can allow for very specific conclusions to be made about very specific brain areas. However, in human subjects, many of the lesion patients have damage to multiple areas. In general, this makes it more difficult to make conclusions about the function of the brain areas. If the person has damage to areas X, Y, and Z, and is unable to enter into REM sleep, we are uncertain whether the area that is related to REM sleep is area X, Y, or Z or some combination of them.

2.6.1.3 Neurosurgical Techniques

Another way the brain has been studied by neuroscientists is through various techniques that are employed before or during brain surgery. One such technique, **direct cortical stimulation**, occurs when a researcher applies a small electrical current directly to the brain itself. This stimulation can cause excitation or inhibition depending on how much stimulation is given. In order to do direct

cortical stimulation, the subject must have their brain exposed during surgery. One may reasonably ask the question, “Why would we ever do this?” Well, when someone is having brain surgery, there is likely a reason. For example, if a patient has a tumor in a medial portion of the brain, doctors may have to go through healthy brain tissue in order to reach the tumor so that they can remove it. Doctors must choose carefully which part of the healthy brain tissue they will damage in order to get to the tumor. One way of figuring out which area would do the least damage is to do a technique known as **cortical mapping**. During cortical mapping, direct cortical stimulation is applied to various parts of the healthy brain tissue to map out their functions. This allows doctors to choose the path of least damage.

Alternatively, cortical mapping can now occur through surgically implanted subdural strip and grid electrodes that will allow the researchers/doctors to stimulate the brain areas in between surgeries, as opposed to during surgery. Additionally, in recent years, researchers have been examining whether TMS is an appropriate (and non-surgical) substitution for direct cortical stimulation.

2.6.1.3.1 Split Brain Studies

Sometimes when surgeons perform surgery to improve the lives of their patients, they can unintentionally create other issues. One famous example of this involves patients who were subjected to a procedure that effectively disrupts the communication between the two sides of the brain. Split-brain research refers to the study of those who received this treatment and the knowledge resulting from this work (Rosen, 2018). Under what circumstances would such a seemingly radical procedure be used - and what are its effects?

In order to treat patients with severe epilepsy, doctors cut the corpus callosum in the brain, which is the main structure that connects the two hemispheres. Doing this kept the electrical activity that was causing the epileptic seizures confined to one hemisphere and helped get the epilepsy under control. However, this also disconnected the two hemispheres from each other, which led to some interesting studies, where researchers were able to study the functions of each hemisphere independently. These studies will be discussed later when we cover lateralization of functions.

2.6.1.3.2 Wada Procedure

One additional way to study the contributions of each hemisphere separately is through a procedure known as a Wada. In a Wada procedure, a barbiturate (a depressant drug used for various purposes including sedation) is used to put one half of the brain “to sleep” and then the contributions of the other hemisphere can be studied. Wada procedures are typically used for similar purposes as are cortical mapping techniques such as direct cortical stimulation. But, instead of mapping specific functions to specific areas (as with direct cortical stimulation), the Wada procedure maps functions to hemispheres. Usually, the Wada is used to identify which hemisphere is responsible for language processing and memory tasks. Although scientists know that language functions are usually in the left hemisphere, it is not always the case (particularly in left-handed individuals), so the Wada will help determine which hemisphere is dominant for language functions. For memory functions, both hemispheres play a significant role, but during the Wada, doctors are able to determine which hemisphere has stronger memory function.

2.6.1.4 One Major Concern With Lesion/Surgery Studies

One thing to remember about all studies of lesion or surgical patients is that the ability to generalize to the population during these studies may be questionable. It is important to keep in mind that that the reason these patients are studied is because they had some sort of issue with their brain. It is reasonable to wonder whether their brains are representative of “normal subjects,” that is, subjects who do not have lesions or other issues.

For example, perhaps someone with epilepsy, after having years of seizures, has a different brain organization than someone without epilepsy. In that circumstance, what we learn from them in a split brain study may not be applicable to a non-epileptic population.

2.6.2 Summary

There are many ways for researchers and doctors to learn about the brain and how it functions. As discussed in this section, we can use animals to study the brain and nervous system and try to draw parallels between what we learn about them and how our own brains work. We also discussed how people with brain lesions or other brain disorders (such as epilepsy) have led scientists to develop techniques to help them function, and how these techniques (split-brain surgeries or Wada procedures) have provided us with valuable information about the brain as well.

2.6.3 References

Gazzaniga, M. C. (1973). The split brain in man. In R. E. Ornstein (Ed.), *The nature of human consciousness: A book of readings* (pp. 87-100). San Francisco: W. H. Freeman

Rosen V. (2018). One Brain. Two Minds? Many Questions. *Journal of undergraduate neuroscience education : JUNE : a publication of FUN, Faculty for Undergraduate Neuroscience*, *16*(2), R48–R50.

This page titled 2.6: Invasive Techniques is shared under a CC BY license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

2.7: Ethical Considerations

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Explain the "Three R's".
2. Explain why research is still conducted on nonhuman animal subjects.
3. Describe the Tuskegee Syphilis Study and explain the ramifications of studies like these on research.

2.7.1 Overview

In this section, we will discuss the ethics of conducting research, both on nonhuman animal subjects and on human subjects. We will discuss a handful of examples of ethical violations and how these violations have led to negative outcomes in some situations.

2.7.2 Ethics in Neuroscience Research

Research has a very complicated history with respect to ethics. This is true when discussing our treatment of nonhuman animal subjects and our treatment of human subjects as well. Let's start by discussing the ethical considerations for nonhuman animal subject research.

2.7.2.1 Nonhuman Animal Subject Research

One area of controversy regarding research techniques is the use of nonhuman animal subjects. One of the keys to behaving in an ethical manner is to ensure that one has given informed consent to be a subject in a study. Obviously, animals are unable to give consent. For this reason, there are some who believe that researchers should not use nonhuman animal subjects in any case.

There are others that advocate for using nonhuman animal subjects because nonhuman animal subjects many times will have distinct advantages over human subjects. Their nervous systems are frequently less complex than human systems, which facilitates the research. It is much easier to learn from a system with thousands of neurons compared to one with billions of neurons like humans. Also, nonhuman animals may have other desirable characteristics such as shorter life cycles, larger neurons, and translucent embryos. However, it is widely recognized that this research must proceed with explicit guidelines ensuring the safe treatment of the animals. For example, any research institution that will be conducting research using nonhuman animal subjects must have an Institutional Animal Care and Use Committee (IACUC). IACUCs review the proposed experiments to ensure an appropriate rationale for using nonhuman animals as subjects and ensure ethical treatment of those subjects.

Furthermore, many researchers who work with nonhuman animal subjects adhere to the Three R's: Replacement, Reduction, and Refinement (Russell & Burch, 1959).

Replacement suggests that researchers should seek to use inanimate systems as a replacement for nonhuman animal subjects whenever possible. Furthermore, replacement is also suggested to replace higher level organisms with lower level organisms whenever possible. The idea is that instead of choosing a primate to conduct the study, researchers are encouraged to use a lower level animal such as an invertebrate (a sea slug, for example) to conduct the study.

Reduction refers to reducing the number of nonhuman animal subjects that will be used in the particular study. The idea here is that if a study can learn sufficient information from one nonhuman animal, then they should only use one.

Finally, refinement is about how the nonhuman animals are cared for. The goal is to minimize discomfort that the subject experiences and to enhance the conditions that the subject experiences throughout their life. For a full discussion of the Three R's, see Tannenbaum and Bennett (2015).

In conclusion, many researchers argue that what we have learned from nonhuman animal subjects has been invaluable. These studies have led to drug therapies for treating pain and other disorders; for instance, most drugs are studied using animals first, to ensure they are safe for humans. Animal nervous systems are used as models for the human nervous systems in many areas. Sea

slugs (Aplysia californica) have been used to learn about neural networks involved in learning and memory. Cats have been studied to learn about how our brain's visual system is organized. Owls have been used to learn about sound localization in the auditory system. Indeed, research using nonhuman animal subjects has led to many important discoveries.

2.7.2.2 Human Subject Research

What about research on human subjects? We do not have to go very far back in history to find situations where researchers behaved in unethical ways towards their human subjects. One of the most famous ethical violations in history is that many experiments were conducted using concentration camp prisoners as subjects during the holocaust.

Throughout the years, psychologists have engaged in various studies that have pushed the envelope of ethical research, such as Milgram's study of obedience or Zimbardo's Stanford prison study. Studies such as these have led to the development of strict ethical guidelines for human research. As with research on nonhuman animal subjects, there is a committee known as an Institutional Review Board (IRB) whose role is to approve research proposals. These committees ensure that there is an appropriate reason for completing the research with human subjects and that the safety of the human subjects are appropriately considered.

To further complicate matters, here in the United States, we have our own history of when ethical violations intersected with racial/ethnic divides.

Indeed, members of some groups have historically faced more than their fair share of the risks of scientific research, including people who are institutionalized, are disabled, or belong to racial or ethnic minorities. A particularly tragic example is the Tuskegee syphilis study conducted by the US Public Health Service from 1932 to 1972 (Reverby, 2009). The participants in this study were poor African American men in the vicinity of Tuskegee, Alabama, who were told that they were being treated for “bad blood.” Although they were given some free medical care, they were not treated for their syphilis. Instead, they were observed to see how the disease developed in untreated patients. Even after the use of penicillin became the standard treatment for syphilis in the 1940s, these men continued to be denied treatment without being given an opportunity to leave the study. The study was eventually discontinued only after details were made known to the general public by journalists and activists. It is now widely recognized that researchers need to consider issues of justice and fairness at the societal level.

“They Were Betrayed”

In 1997— 65 years after the Tuskegee Syphilis Study began and 25 years after it ended— President Bill Clinton formally apologized on behalf of the US government to those who were affected. Here is an excerpt from the apology:

So today America does remember the hundreds of men used in research without their knowledge and consent. We remember them and their family members. Men who were poor and African American, without resources and with few alternatives, they believed they had found hope when they were offered free medical care by the United States Public Health Service. They were betrayed.

Read the full text of the apology at http://www.cdc.gov/tuskegee/clintonp.htm.



Figure **2.7.1**: President Bill Clinton greets a survivor of the Tuskegee Syphilis Study on African-American men.

The racism behind the unethical behavior in the Tuskegee study (and other studies) has led to a general distrust of research from some minorities. This distrust in research has led to a lack of volunteers from minority communities in research. Distrust in research also has other consequences, and as recently as 2021 has been cited as a contributing factor for the disproportionate impact of the COVID-19 pandemic on minority communities (Carson, et all., 2021). Unfortunately, when a large portion of the research conducted is on Caucasian samples, it is unclear whether or not the results generalize to non-Caucasian groups. Given the complexities of the human brain, researchers continue to push for a sample that is truly representative of the population at large to ensure that their research results are generalizable.

2.7.3 Summary

In this section, we discussed the ethical considerations of conducting research on nonhuman animal subjects and human subjects. We discussed what rules and guidelines have been put in place to ensure ethical conduct from researchers with their subjects. We also discussed some examples of violations of ethical research practices and the consequences of those violations.

2.7.4 References

Carson, S.L, Casillas, A., Castellon-Lopez, Y., Mansfield, L. N., Morris, D., Barron, J., Ntejume, E., Landovitz, R., Vassar, S., Norris, K., Dubinett, S. M., Garrison, N. A., Brown, A. F. (2021). COVID-19 Vaccine Decision-making factors in racial and ethnic minority communities in Los Angeles, California. *JAMA Network Open 4*(9)

Reverby, S.M. (2009). *Examining Tuskegee: The infamous syphilis study and its legacy.* Chapel Hill, NC: University of North Carolina Press. (As cited by Jhangiani et al., 2019.)

Russell W.M.S. and Burch R.L. (1959) *The Principles of Humane Experimental Technique*. London (UK): Methuen & Co. Ltd (special edition published by Universities Federation for Animal Welfare, 1992).

Tannenbaum J. and Bennett, B.T. (2015). Russell and Burch's 3Rs Then and Now: The Need for Clarity in Definition and Purpose. *Journal of the American Association for Laboratory Animal Science*, *54*(2), 120-132.

2.7.5 Attributions

Figures:

- President Bill Clinton greets a survivor of the Tuskegee Syphilis Study on African-American men by Sharon Farmer White House Photograph Office, Public domain, via Wikimedia Commons

Text:

- Moral Foundations of Ethical Research by Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler, & Dana C. Leighton is licensed under a CC-BY-SA 4.0 International License.

This page titled 2.7: Ethical Considerations is shared under a CC BY-SA license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

CHAPTER OVERVIEW

3: Evolution, Genes, and Behavior

> This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

3.1: Evolution of Minds and Behavior
3.2: Evolution - Mechanisms and Evidence
3.3: Human Evolution
3.4: Evolutionary Theories in Psychology
3.5: DNA, Chromosomes, and Gene-Environment Correlations
3.6: The Nature-Nurture Question
3.7: Epigenetics in Psychology
3.8: Human Genetics
3.9: Genes and Behavior

This page titled 3: Evolution, Genes, and Behavior is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

3.1: Evolution of Minds and Behavior

Learning Objectives

1. Discuss the claim that nothing in psychology makes sense except in the light of evolution.
2. Explain the process of natural selection and provide an example.
3. Describe what is meant by adaptation and explain what is meant by a psychological adaptation.
4. Explain how evolution can be nonrandom, yet still without purpose.
5. Describe what William James meant by the term "instincts" applied to humans, according to Cosmides and Tooby.
6. Discuss how evolutionary psychologists conceptualize human nature.
7. Discuss evolution as the unifying principle for psychology, given that evolution is the unifying principle for all of biology.
8. Discuss how two types of selection (hint: one of these is kin selection) may explain the duality of human moral nature and feelings of moral conflict.
9. Describe the primary features of evolutionary psychology as an approach to the study of the mind and brain.

3.1.1 Overview

Why do humans and other animals move, why do they have particular mental capabilities and not others, and what laws or principles govern the organization of behavior and mental processes? In this section, the concept of evolution will be introduced and the insights it can provide regarding mind and behavior considered. When we recognize that mental activities and behavior in humans and other animals are the product of a biological organ, the brain, we can then apply principles of biological science to an understanding of psychology. This approach focuses attention on the biological functions served by mental abilities and behavior and how they contribute to adaptation to environmental change.

3.1.2 Evolution, Natural Selection, and Psychological Adaptation

> "In the distant future I see open fields for far more important researches. Psychology will be based on a new foundation, that of the necessary acquirement of each mental power and capacity by gradation." Charles Darwin, *On the Origin of Species* (1859/1996, p. 394)

> "Nothing in biology makes sense except in the light of evolution." Theodosius Dobzhansky (1973, p. 125)

When looking at nature, it is important to ask why things in nature have the properties that they do. This is especially important when thinking about organisms. Why do organisms possess the traits that they do? Why do birds have feathers? Why do we have emotions and why are our emotions so similar to the emotions found in a wide range of other animal species? Why do we have the capacity to think, why does our thinking take the forms that it does, and why are many of our thinking processes fundamentally similar to the thinking processes found in many other animal species? Why do we see color, while many other species don't? Why do we have the capacity to form mental images of the past and of imagined futures?

Psychology is the scientific study of mental processes and behavior. These processes are, of course, all functions of a biological organ, the brain. Like other organs, the brain and its operations have evolutionary origins. Understanding evolution can give us fresh insights about our psychology and the psychology of other species. The characteristics of organisms, including mental and behavioral traits, perform biological functions that contribute to survival and reproduction.

The minds and behavior of organisms have been shaped by evolutionary forces over millions of years, just like their anatomical and physiological features. To the trained eye, the mark of evolution is still evident in the psychology of present-day humans and other animals. And, as Darwin's principle of the continuity of species predicts, human and animal species, although diverse, show fundamental similarities in mind and behavior just as they do in their anatomy and physiology. Although there are many factors that affect evolution, the key driving force in evolution is the process of natural selection (Dawkins, 1976, 1982).

3.1.3 Natural Selection

Darwin was not the first to propose that species evolve, that they change over time. Darwin's brilliant insight was the discovery of the primary process which guides evolutionary change--natural selection, the survival and reproduction of the "fittest". "Fit" in this context does not mean the healthiest or the strongest, rather it is a reference to the ability to pass one's genes on to the next

generation. Natural selection is similar to artificial selection, a process used by animal breeders to enhance characteristics which they deem desirable. But in natural selection it is the environment that does the selection.

Natural selection acts on natural variation within populations. One important feature of life is that individual organisms within any sexually-reproducing species vary from one another in countless ways. Natural selection works because some individuals are better suited to the environment than others; those individuals better fit to the environment have a competitive advantage and naturally survive and reproduce in greater numbers and therefore pass on their heritable (genetic) traits to succeeding generations with greater frequency. This is the essence of evolution.

Evolution by natural selection is inevitable given three factors:

- genetic variation in heritable traits in a population of reproducing organisms due to mutation and the mixing of genes during sexual reproduction;
- competition among members of the population for limited environmental resources;
- those individuals who are better suited, better "fit," to the demands and opportunities of the environment have a competitive advantage and therefore survive and reproduce in greater numbers. The result is the proliferation of the genes and traits which provide superior fitness to the environment. Fitness is measured not simply in terms of survival, but more importantly by reproductive success, the number of living offspring that are produced.

In other words, some individuals, by chance, happen to have genetic characteristics that make them better prepared to survive and reproduce, leading their genes to be selected and to persist in future generations. By this process of selection, traits which give a competitive edge in solving problems of survival and reproduction and in exploiting environmental opportunities tend to be preserved over generations, while traits less well-suited are removed. For example, those better at predator avoidance, better at obtaining energy from food, better at successful mating, or better at inventing ways to use the environment for adaptive advantage survive and reproduce more offspring. By this process of natural selection, over countless generations, species of living things evolve characteristics (including mental and behavioral traits) that are well-adapted to solve problems and to exploit opportunities in their particular environments. Gradually, over long periods of time, organisms acquire adaptive "design" by completely natural and unintentional processes. Note that as the environment changes, new traits may evolve. In a sense, evolutionary change tracks changes in the environment, creating an improving match between organism and environment over evolutionary time.

Nevertheless, relatively rapid changes in the environment may outpace the rate of evolutionary change. This means that adaptations that evolved for an earlier environment may not necessarily be well suited to the current environment. This is important in understanding some of the psychological traits of humans that were once advantageous, but which have now become non-adaptive. For example, some experts have argued that humans have an innate disposition to be territorial and to be suspicious and wary of strangers, of outsiders, and perhaps even hostile to those outside of their own group. Although this may have been an adaptive psychological trait in our ancient Pleistocene past when humans lived in small groups of hunter-gatherers, in today's world these tendencies may dispose us toward prejudice and even dangerous and wasteful wars.

3.1.4 Environmental Selection and Sexual Selection

The type of natural selection discussed above is environmental selection--selection by the environment of which genetic variants in a population will survive long enough to have a chance to reproduce. Note that the environment "selects" who will survive and who won't by virtue of the fact that the environment presents challenges to survival such as disease, predators, and insufficient supplies of resources such as energy (in the form of sunlight for plants, or food for animals), territory, and water. The "selection" by the environment occurs when individuals that just happen to be better suited to meet the challenges presented by the environment naturally have a competitive advantage and survive and reproduce in greater numbers than those unlucky individuals who are less well fit to the environment.

In addition to environmental selection, in sexually reproducing species, there is a second layer of selection, sexual selection--selection based on the "attractiveness" of potential sexual partners. All other factors being equal, individuals with genes that make them more attractive to the opposite sex tend to have more reproductive opportunities and, at least in times past, tend to have more offspring than those perceived to be less attractive. This effect is in part due to the fact that bodily features associated with health and likely reproductive success tend to be perceived as sexually attractive (presumably as a result of brain evolution).

3.1.5 Evolution is neither random nor purposeful

It is important to note, as described above, that although evolution depends on random variations (in traits and genes) among individuals within a population, evolution is not a random process--natural selection acts on the random variation, constraining it toward successful adaptation. As Buss and Hawley (2010, p. ix) state, “Individual differences are indispensable for natural selection. Without heritable variants, natural selection—the only known process capable of creating and maintaining functional adaptations—could not occur.” In this way, evolution is not random as some who don't understand the process are apt to claim--natural selection gives it direction.

Furthermore, it is important to note that evolution is not purposeful, even though it is not random. Evolution happens automatically, without purpose, simply as a result of differential rates of reproduction in a population occurring as a consequence of the fact that individuals in a population with traits (and genes) that are more successful in a particular environment end up surviving and producing more offspring than their competitors--that's natural selection. This results in changes in gene frequencies in populations of organisms, and that is evolution. As biologist, Theodosius Dobzhansky (1964, p.449), stated: "My genes are different sequences of the same four "letters" of the "genetic alphabet" which also compose the genes of a fish or of a corn plant. Genes reproduce themselves generally with an astonishing accuracy; the sequences of the four 'letters,' the nucleotide bases, are usually identical in hundreds of billions of cells of the bodies of the parents and of their progeny. Occasionally, there occur, however, changes, "misprints," mutations. Self-reproduction plus mutation make possible natural selection. Natural selection makes possible evolution." We might even say that given these conditions, natural selection and evolution are not only possible, but inevitable.

3.1.6 Adaptations

The heritable features which organisms use to solve problems of survival and reproduction are called **adaptations**. These evolved adaptations can be anatomical, such as having wings or fur; physiological, such as digestive processes or having an immune system; and behavioral and mental (i.e. psychological). The evolved behavioral and mental traits of an organism are its **psychological adaptations** (e.g. having a fear response to danger; feelings of sexual attraction which draw you toward desirable potential mates; having a "sweet tooth" that drives you to seek out and ingest high caloric foods; having tender feelings toward your offspring motivating care-giving; understanding cause-effect relations and making causal inferences; having the mental ability to imagine future actions and to mentally anticipate the probable outcomes of those actions; and so on). According to psychologists who favor an evolutionary perspective, the mind can be seen as a large collection of evolved psychological adaptations (not all psychologists agree with this view; see Panksepp and Panksepp, 2000). These psychological adaptations are built by evolution into the structure of the brain and its physical operations. These psychological adaptations can be quite specific and concrete, such as the innate emotions related to pair-bonding and mating (e.g. "falling in love") or inborn human taste preferences for sweets and fats (see Cosmides and Tooby, 1997), or they can be quite general and abstract *instincts of thought* such as the innate disposition to understand the world in terms of cause-effect, to be sensitive to predictability between events, and to form categories and inferences based on similarities among things. These and other instincts of thought make up much of what we call intelligence (Koenigshofer, 2017).



Figure **3.1.2**: Cultural overlays on a set of psychological adaptations. Human mating involves a number innate "instinctive" emotions and psychological needs which have evolved to foster reproduction of genes. Genetic evolution of these psychological adaptations has been accompanied by cultural evolution (selection and cultural transmission of learned behaviors via tradition, modeling, and other mechanisms of cultural transmission) of traditional practices associated with the genetically evolved psychological adaptations, the "reproductive instincts." (Image from Wikimedia Commons; File:Long Wedding Dress for Couple with Flowers.jpg; https://commons.wikimedia.org/wiki/F...th_Flowers.jpg; by epSos.de; licensed under the Creative Commons Attribution 2.0 Generic license).

In humans, psychological adaptations such as the emotions involved in mate selection, pair bonding, and mating are often associated with learned ritual behaviors that are culturally transmitted over generations and become part of the culture of particular human groups (see Figure 3.1.2). Such cultural practices are often based on commonly held "myths" that unify and help identify large groups of humans as culturally distinct from one another while facilitating cooperative effort among the members of such large groups toward common goals. According to one author (Harari, 2014), this ability to form cultural myths that imbue large numbers of strangers with a common identity, permitting them to work toward common goals, is unique to humans and accounts, in large part, for the unprecedented success of our species compared to all others. However, it is important to realize that this ability to form unifying myths depends upon features of the human brain that don't exist in sufficient degree in other animal species. Myths associated with cultural practices ranging from marriage or child birth to religious ritual or imperialism require the ability to form abstract concepts. Although research shows that non-human animal species can form concepts (Smith et al., 2010; Zentall et al., 2008), human concepts are highly abstract, apparently uniquely so, and ability for high levels of abstraction may depend upon the unique complexity of the human cerebral cortex (Koenigshofer, 2017; see Chapter 14 on Intelligence and Cognition). However, in addition, to ability to form abstract concepts, ability to learn and to transmit learned knowledge and behaviors from one generation to the next (cultural transmission) was essential to human uniqueness, and this ability to transmit learned information from generation to generation, by tradition and other non-genetic means, also depended upon particular features of the human brain (see Chapters 14 and 15 and sections 18.5 and 18.13). Thus, the great accomplishments of human civilization, which set us so clearly apart from other species, ultimately depended upon human brain evolution setting the stage, with human cultural evolution superimposed and critically dependent on brain evolution. One specific example may help drive the point home. Many experts believe that human tool making and tool use were extremely important in human evolution and early human adaptive success (see sections 3.3 and 18.5). The manufacture of stone tools by early humans was a complex and sophisticated cognitive and manual task beyond the capability of all other primates. Orban and colleagues (2006) identified a set of regions in the dorsal intraparietal sulcus (IPS) of the human cerebral cortex which they believe perform the complex visual analysis needed for the precision with

which humans manipulate tools; these brain areas are not found in monkeys. Thus, a key product of human cultures was only possible because of unique features evolved in the human brain. This is only one example of a general claim: human cultural achievements that make our species so unique depend upon unique features of human brain evolution.

3.1.7 Individual Selection, Kin Selection, and Human Moral Conflict

From an evolutionary perspective, the minds and behavior of humans and animals, just like their anatomy and physiology, have evolved in service of survival and reproduction, either through the reproduction of one's own genes, or indirectly through the survival and reproduction of close relatives and their genes (Dawkins, 1976). The former (individual selection) favors the self-interested side of human nature; the latter (kin selection) favors our altruistic side including prosocial behaviors such as caring, giving, sharing, and cooperation (some of the adaptive benefits of living in groups in social species such as humans, wolves, lions, chimpanzees, elephants, etc). In humans at least, self-interest and the interests of others often conflict, generating psychological tension experienced as moral dilemma. For humans, it is likely that there is an optimal balance between these two opposing behavioral dispositions; an extreme, maladaptive imbalance can lead to psychopathologies such as narcissistic and anti-social personality disorders. The duality of human moral nature may have roots in the duality of these two evolutionary processes, selection for survival and reproduction of one's own genes and kin selection, selection for survival and reproduction of the genes of close genetic relatives (Koenigshofer, 2010, 2016).

3.1.8 Evolutionary Psychology

Because psychological adaptations are located in the brain and its operations, these adaptations can be understood as information processing features of neural systems. On this view, the brain is a computational machine comprised of an enormous number of computational systems or modules, essentially mini-computers, made of circuits of neurons whose circuit configurations and operations have been shaped by evolutionary forces, developmental processes, and environmental experiences.

As noted in the opening quote above, Darwin recognized the relevance of evolution to psychology. Others have followed his lead. As Cosmides and Tooby (1997, p. 1) state, "In the final pages of the Origin of Species, after he had presented the theory of evolution by natural selection, Darwin made a bold prediction: 'In the distant future I see open fields for far more important researches. Psychology will be based on a new foundation, that of the necessary acquirement of each mental power and capacity by gradation.' Thirty years later, William James tried to do just that in his seminal book, Principles of Psychology, one of the founding works of experimental psychology (James, 1890). In Principles, James talked a lot of "instincts". This term was used to refer (roughly) to specialized neural circuits that are common to every member of a species and are the product of that species' evolutionary history. Taken together, such circuits constitute (in our own species) what one can think of as 'human nature.'"

Today, the influence of these early thinkers is expressed in evolutionary approaches to psychology. Cosmides and Tooby (1997, p. 1) state: "The goal of research in evolutionary psychology is to discover and understand the design of the human mind. Evolutionary psychology is an approach to psychology, in which knowledge and principles from evolutionary biology are put to use in research on the structure of the human mind. It is not an area of study, like vision, reasoning, or social behavior. It is a way of thinking about psychology that can be applied to any topic within it. In this view, the mind is a set of information-processing machines that were designed by natural selection to solve adaptive problems faced by our hunter-gatherer ancestors." Furthermore, they explain, "Psychology is that branch of biology that studies (1) brains, (2) how brains process information, and (3) how the brain's information-processing programs generate behavior. Once one realizes that psychology is a branch of biology, inferential tools developed in biology -- its theories, principles, and observations -- can be used to understand psychology" (Cosmides and Tooby, 1997 p. 3).

Evolutionary psychologists have applied this approach to emotion (Johnston, 1999; Ketelaar, 2015), intelligence and cognition (Bouchard, 2014; Koenigshofer, 2017; Pika, et al., 2020), personality (Buss and Hawley, 2010; Figueredo, et al., 2009), language (Corballis, 2010; Fitch, 2010; Pinker, 2003), social cognition (Fiddick, 2015; Vonk, et al., 2015), and a wide range of other psychological processes. Evolution by natural selection is gaining increasing acceptance among psychologists as "a strong candidate for central inclusion in a unifying meta-theory of psychology" (Marsh and Boag, 2013, p. 655; also see Goetz & Shackelford, 2006), just as evolution has become the unifying principle for all of biology. As you read about evolution and its mechanisms in the following sections, keep in mind that not only physical traits evolve, but mental and behavioral traits and capabilities (psychological traits) do as well.

In the remainder of this chapter we examine the processes of evolution in greater detail including primate and human evolution, and evolutionary theories in psychology. To emphasize the importance of evolution to an understanding of mind and behavior, recall the claim stated above: *Nothing in psychology "makes sense except in the light of evolution."*

3.1.9 Summary

Behavior and the information processing mechanisms that underlie behavior are biological processes (in their origin and functions), dependent upon a biological organ, the brain. Just as evolution by natural selection is the organizing principle for all of biology, it must be a key organizing principle for psychology as well. Natural selection is the primary driving force of evolution, although other processes (examined in other sections of this chapter) are also involved. Natural selection occurs automatically without purpose or goal simply as a mechanistic consequence of two facts of life: 1) that individuals of any species vary from one another in their heritable traits, and 2) that some individuals by chance, because of the traits they possess (and the genes that underlie those traits), have a competitive advantage that favors their rates of survival and reproduction in a particular environment compared to those variants that are less fit. Therefore, the individuals (and the underlying genes) that are best fit to the environment end up leaving more surviving offspring and therefore have a disproportionately greater impact on the genes and traits of future generations. As a consequence, evolution occurs over generations, causing species to accumulate traits that are a better fit to survival and reproduction in the particular environments they occupy. For biological psychology, it is important to recognize that in addition to anatomy and physiology, psychological traits are also shaped into adaptive form by these same processes. Thus, to understand human and animal behavior, an evolutionary approach provides a biological context within which we can examine origins, functions, and organizing principles of minds and behavior. This approach does not negate or deny the influence of learning and culture but views these processes as biological phenomenon with evolutionary and neurophysiological roots in the brain.

3.1.10 References

Bouchard, T. J. (2014). Genes, evolution and intelligence. *Behavior genetics*, *44*(6), 549-577.

Buss, D. M., & Hawley, P. H. (Eds.). (2010). *The evolution of personality and individual differences*. Oxford University Press.

Corballis, M. C. (2010). Mirror neurons and the evolution of language. *Brain and language*, *112*(1), 25-35.

Cosmides, L., & Tooby, J. (1997, January 13). Evolutionary psychology: A Primer. Retrieved February 20, 2021, from https://www.cep.ucsb.edu/primer.html (Links to an external site.)

Darwin, C. (1859/1996). The origin of species. Oxford, UK: Oxford University Press

Dawkins, R. (1976). The selfish gene. Oxford University Press.Dawkins, R. (1976). The Selfish Gene. Oxford: Oxford University Press.

Dawkins, R. (1982). The Extended Phenotype: The Gene as the Unit of Natural Selection. San Francisco: Freeman.

Dobzhansky, T. (1964). Biology, Molecular and Organismic. *American Zoologist* 4 (4), 443-452.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. The American Biology Teacher, 35,

125-129.

Fiddick, L. (2015). Social by Design: How Social Psychology Can Be More Cognitive Without Being Less Social. In *Evolutionary Perspectives on Social Psychology* (pp. 15-26). Springer, Cham.

Figueredo, A. J., Gladden, P., Vasquez, G., Wolf, P. S. A., & Jones, D. N. (2009). 16 Evolutionary theories of personality. *The Cambridge handbook of personality psychology*, 265.

Fitch, W. T. (2010). *The evolution of language*. Cambridge University Press.

Goetz, A. T., & Shackelford, T. K. (2006). Modern application of evolutionary theory to psychology: Key concepts and clarifications. *The American journal of psychology*, 567-584.

Grant, B. R., & Grant, P. R. (1989). Natural selection in a population of Darwin's finches. *The American Naturalist*, *133* (3), 377-393.

Grant, R. B., & Grant, P. R. (2003). What Darwin's finches can teach us about the evolutionary origin and regulation of biodiversity. *BioScience*, *53* (10), 965-975.

Harari, Y. N. (2014). *Sapiens: A brief history of humankind*. Random House.

Johnston, V. S. (1999). *Why we feel: The science of human emotions*. Perseus Publishing.

Kennair, L. E. O. (2002). Evolutionary psychology: An emerging integrative perspective within the science and practice of psychology. *Human Nature Review*, *2*, 17-61.

Ketelaar, T. (2015). Evolutionary psychology and emotion: A brief history. In *Evolutionary perspectives on social psychology* (pp. 51-67). Springer, Cham.

Koenigshofer, K. (2010). *Mind Design: The Adaptive Organization of Human Nature, Minds, and Behavior.* Pearson Educational Solutions, Boston.

Koenigshofer, K. (2016). *Mind Design: The Adaptive Organization of Human Nature, Minds, and Behavior.* Revised Edition. Amazon e-book.

Koenigshofer, K. A. (2017). General Intelligence: Adaptation to Evolutionarily Familiar Abstract Relational Invariants, Not to Environmental or Evolutionary Novelty. *The Journal of Mind and Behavior*, *38* (2), 119-153.

Marks, J. Chapter 2, Evolution, in *Explorations: An Open Invitation To Biological Anthropology*; https://pressbooks-dev.oer.hawaii.ed...r/__unknown__/; licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, except where otherwise noted.

Marsh, T., & Boag, S. (2013). Evolutionary and differential psychology: conceptual conflicts and the path to integration. *Frontiers in psychology*, *4*, 655.

Orban, G. A., Claeys, K., Nelissen, K., Smans, R., Sunaert, S., Todd, J. T., Wardak, C., Durand, J. & Vanduffel, W. (2006). Mapping the parietal cortex of human and non-human primates. *Neuropsychologia* 44 (13):2647 –67. [GAO, arKV, GV]

Panksepp, J., & Panksepp, J. B. (2000). The seven sins of evolutionary psychology. *Evolution and cognition*, *6*(2), 108-131.

Pika, S., Sima, M. J., Blum, C. R., Herrmann, E., & Mundry, R. (2020). Ravens parallel great apes in physical and social cognitive skills. *Scientific reports*, *10*(1), 1-19.

Pinker, S. (2003). *The language instinct: How the mind creates language*. Penguin UK.

Smith, J. D., Beran, M. J., Crossley, M. J., Boomer, J., & Ashby, F. G. (2010). Implicit and explicit category learning by macaques (Macaca mulatta) and humans (Homo sapiens). *Journal of Experimental Psychology: Animal Behavior Processes*, *36* (1), 54.

Vonk, J., McGuire, M., & Johnson-Ulrich, Z. (2015). The Evolution of Social Cognition. In *Evolutionary Perspectives on Social Psychology* (pp. 81-94). Springer, Cham.

Zentall, T. R., Wasserman, E. A., Lazareva, O. F., Thompson, R. K., & Rattermann, M. J. (2008). Concept learning in animals. *Comparative Cognition & Behavior Reviews*. *3*, 13–45. https://doi.org/10.3819/ccbr.2008.3000

3.1: Evolution of Minds and Behavior is shared under a CC BY license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer.

3.2: Evolution - Mechanisms and Evidence

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

3.2.1 Learning Objectives

1. Describe how the present-day theory of evolution was developed.
2. Explain natural selection and its role in evolution.
3. Explain convergent and divergent evolution.
4. Describe homologous and vestigial structures.
5. Discuss misconceptions about the theory of evolution.
6. Explain the categories of evidence for evolution.
7. Describe species and speciation.
8. Explain kin selection, inclusive fitness, and the evolution of altruistic behaviors such as helping, cooperation, and giving.

3.2.2 Overview

The theory of evolution is the unifying theory of biology, meaning it is the framework within which biologists ask questions about the living world. Its power is that it provides direction for predictions about living things that are borne out in experiment after experiment. Recall the claim by geneticist, Theodosius Dobzhansky, quoted in Section 3.1, that “nothing makes sense in biology except in the light of evolution." He meant that the tenet that all life has evolved and diversified from a common ancestor is the foundation from which we approach all questions in biology. It is important to keep this claim in mind when we think about brain evolution and psychology. The evolutionary links among animal species help explain the similarities in their brains and behavior.

3.2.2.1 **Summary of Key Points:**

1. Evolution: All species of living organisms, from bacteria to baboons to blueberries, evolved at some point from a different species. Although it may seem that living things today stay much the same, that is not the case—evolution is an ongoing process.
2. Understanding Evolution: Evolution by **natural selection** describes a mechanism for how species change over time. That species change had been suggested and debated well before Darwin began to explore this idea. The view that species were static and unchanging was grounded in the writings of Plato, yet there were also ancient Greeks who expressed evolutionary ideas. Darwin was the first to describe the primary mechanism of evolutionary change--natural selection.
3. When Darwin proposed his theory of evolution, the mechanisms of inheritance were unknown. Now, we know a great deal about those mechanisms. Here are some key definitions. **Chromosomes** are long strings of genes made of the molecule, deoxyribonucleic acid (DNA). A **gene** is a segment of a chromosome coding for synthesis of a specific protein. **Alleles** are alternative forms of a gene at the same location on the chromosome (for example a blue gene for eye color and a brown gene for eye color are alleles). **Genotype** is the genetic makeup of an individual. **Phenotype** is the actual anatomical, physiological, and behavioral/psychological characteristics of an individual.



Figure **3.2.1**: All organisms are products of evolution adapted to their environment. (a) Saguaro (Carnegiea gigantea) can soak up 750 liters of water in a single rain storm, enabling these cacti to survive the dry conditions of the Sonora desert in Mexico and the Southwestern United States. (b) The Andean semiaquatic lizard (Potamites montanicola) discovered in Peru in 2010 lives between 1,570 to 2,100 meters in elevation, and, unlike most lizards, is nocturnal and swims. Scientists still do not know how these cold-blood animals are able to move in the cold (10 to 15°C) temperatures of the Andean night. (credit a: modification of work by Gentry George, U.S. Fish and Wildlife Service; credit b: modification of work by Germán Chávez and Diego Vásquez, ZooKeys)

3.2.2.2 Understanding Evolution

Darwin's evolution by natural selection describes a mechanism for how species change over time. Darwin described evolution as "descent with modification."

In the eighteenth century, James Hutton, a Scottish naturalist, proposed that geological change occurred gradually by the accumulation of small changes from processes operating like they are today over long periods of time. This contrasted with the predominant view that the geology of the planet was a consequence of catastrophic events occurring during a relatively brief past. Hutton's view was popularized in the nineteenth century by the geologist Charles Lyell who became a friend to Darwin. Lyell's ideas were influential on Darwin's thinking: Lyell's notion of the greater age of Earth gave more time for gradual change in species, and the process of change provided an analogy for gradual change in species. In the early nineteenth century, Jean-Baptiste Lamarck published a book that detailed a mechanism for evolutionary change. This mechanism is now referred to as an inheritance of acquired characteristics by which modifications in an individual that are caused by its environment, or the use or disuse of a structure during its lifetime, could be inherited by its offspring and thus bring about change in a species. A simple example or two will illustrate why this is wrong--the idea suggests that if you go to the gym and get buff, then you will have buff kids, or if you lose a finger in an accident, that you will have children missing a finger, and that such changes will eventually become characteristics of the species. Today we know that if the change in characteristics does not change genes, the change cannot be passed on to future generations. While this mechanism for evolutionary change was discredited, Lamarck's ideas were an important influence on evolutionary thought.

3.2.2.3 Charles Darwin and Natural Selection

In the mid-nineteenth century, the actual mechanism for evolution was independently conceived of and described by two naturalists: Charles Darwin and Alfred Russel Wallace. Importantly, each naturalist spent time exploring the natural world on expeditions to the tropics. From 1831 to 1836, Darwin traveled around the world on *H.M.S. Beagle*, including stops in South America, Australia, and the southern tip of Africa. Wallace traveled to Brazil to collect insects in the Amazon rainforest from 1848 to 1852 and to the Malay Archipelago from 1854 to 1862. Darwin's journey, like Wallace's later journeys to the Malay Archipelago, included stops at several island chains, the last being the Galápagos Islands, west of Ecuador. On these islands, Darwin observed species of organisms on different islands that were clearly similar, yet had distinct differences. For example, the ground finches inhabiting the Galápagos Islands comprised several species with a unique beak shape. The species on the islands had a graded series of beak sizes and shapes with very small differences between the most similar. He observed that these finches closely resembled another finch species on the mainland of South America. Darwin imagined that the island species might be species modified from one of the original mainland species. Upon further study, he realized that the varied beaks of each finch

helped the birds acquire a specific type of food. For example, seed-eating finches had stronger, thicker beaks for breaking seeds, and insect-eating finches had spear-like beaks for stabbing their prey.

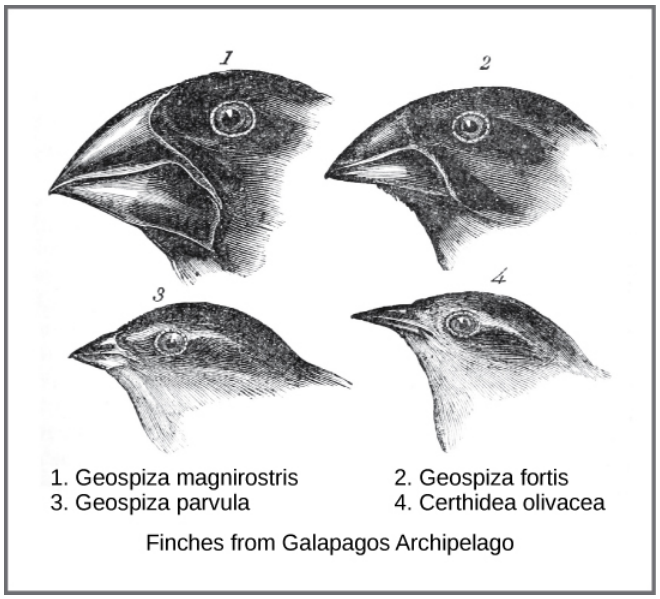


Figure **3.2.2**: Darwin observed that beak shape varies among finch species. He postulated that the beak of an ancestral species had adapted over time to equip the finches to acquire different food sources.

Wallace and Darwin both observed similar patterns in other organisms and they independently developed the same explanation for how and why such changes could take place. Darwin called this mechanism natural selection. Natural selection, also known as "survival of the fittest," is more accurately described as survival and *reproduction* of the fittest. Survival alone is insufficient. Evolution involves differential rates of reproduction. Survival alone without reproduction has no effect on the genetic evolution of a species. Natural selection results in the more prolific reproduction of individuals with traits that contribute to survival and reproduction in a changing environment; this increased rate of reproduction in individuals with traits better fit to the environment compared to individuals with traits less fit leads to evolutionary change.

For example, a population of giant tortoises found in the Galapagos Archipelago was observed by Darwin to have longer necks than those that lived on other islands with dry lowlands. These tortoises were "selected" because they could reach more leaves and access more food than those with short necks. In times of drought when fewer leaves would be available, those that could reach more leaves had a better chance to eat and survive than those that couldn't reach the food source. Consequently, long-necked tortoises would be more likely to be reproductively successful and pass the long-necked trait to their offspring. Over time, only long-necked tortoises would be present in the population, as short-necked animals failed to survive and reproduce.

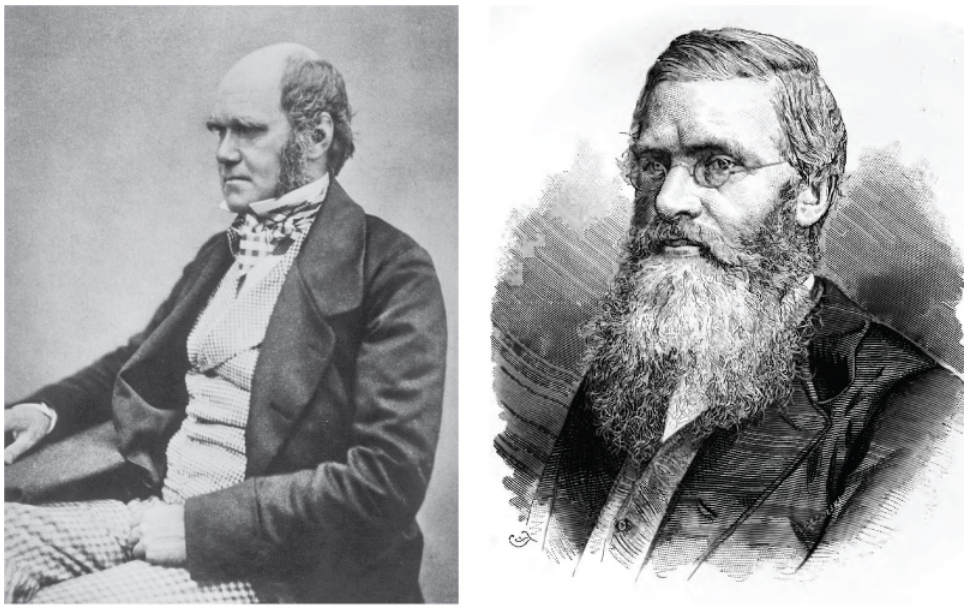
Natural selection, Darwin argued, was an **inevitable outcome of three principles** that operated in nature.

First, **most characteristics of organisms are inherited**, or passed from parent to offspring. Although no one, including Darwin and Wallace, knew how this happened at the time, it was a common understanding.

Second, more offspring are produced than are able to survive, so resources for survival and reproduction are limited. The capacity for reproduction in all organisms outstrips the availability of resources to support their numbers. Thus, there is **competition for** those **resources in each generation**. Both Darwin and Wallace's understanding of this principle came from reading an essay by the economist Thomas Malthus who discussed this principle in relation to human populations.

Third, **offspring vary** among each other in regard to their characteristics **and those variations are inherited**. Darwin and Wallace reasoned that **offspring with inherited characteristics which allow them to best compete for limited resources will survive and have more offspring than those individuals with variations that are less able to compete**. Because characteristics are inherited, these traits will be better represented in the next generation. This will lead to change in populations over generations in a process that Darwin called **descent with modification**. Ultimately, natural selection leads to greater adaptation of the population to its local environment; it is the only mechanism known for adaptive evolution.

Papers by Darwin and Wallace presenting the idea of natural selection were read together in 1858 before the Linnean Society in London. The following year Darwin's book, *On the Origin of Species*, was published. His book outlined in considerable detail his arguments for evolution by natural selection.



(a) (b)

Figure **3.2.3**: Both (a) Charles Darwin and (b) Alfred Wallace wrote scientific papers on natural selection that were presented together before the Linnean Society in 1858.

Demonstrations of evolution by natural selection are time consuming and difficult to obtain. As briefly discussed in Module 3.1, one of the best examples has been demonstrated in the very birds that helped to inspire Darwin's theory: the Galápagos finches. Peter and Rosemary Grant and their colleagues have studied Galápagos finch populations every year since 1976 and have provided important demonstrations of natural selection. The Grants found changes from one generation to the next in the distribution of beak shapes in the medium ground finch on the Galápagos island of Daphne Major. The birds have inherited variation in the bill shape with some birds having wide deep bills and others having thinner bills. During a period in which rainfall was higher than normal because of an El Niño (an unusual warming of the Pacific Ocean near the West coast of the Americas), the large hard seeds that large-billed birds ate were reduced in number; however, there was an abundance of the small soft seeds which the small-billed birds ate. Therefore, survival and reproduction were much better in the following years for the small-billed birds. In the years following this El Niño, the Grants measured beak sizes in the population and found that the average bill size was smaller. Since bill size is an inherited trait, parents with smaller bills had more offspring and the size of bills had evolved to be smaller. As conditions changed and a long period of drought ensued, larger seeds became more available. The trend toward smaller average bill size ceased and selection for larger beaks and body size resulted in a lasting increase in beak size.

As discussed in Section 3.1, another well known example of observable evolution is the peppered moth. Because of its importance as another example of natural selection in action, we review the details here. Prior to industrialization in England, the peppered moth had light colored wings that closely matched the color of the bark of the trees. This inherited trait provided camouflage for the moths helping to protect them from predator birds. Although most of the moths had light colored wings, a few each generation had darker wings, but without sufficient camouflage these darker colored moths had been easy prey and survived and reproduced with low frequency. As industrialization came to England, soot from factories began to slowly turn the bark of trees darker and darker. As this occurred, darker wings were now an advantage, helping to hide the darker moths from predators when they landed on trees, increasing their survival and reproductive rates over the lighter colored moths. As natural selection continued over decades, the light colored moths, once the predominant form, were increasingly replaced over generations by darker winged variants, until today, after no more than a hundred years of evolution, the peppered moth has dark wings (although as mentioned in Module 3.1 with cleaner air, numbers of the lighter variant are increasing as the tree bark continues to lighten). This transformation of the species as a consequence of natural selection, as the environment changed, is a good example of evolution rapid enough that it can be observed.

3.2.3 Process and Pattern of Evolution

Natural selection can only take place if there is variation, or differences, among individuals in a population. Importantly, these differences must have some genetic basis; otherwise, the selection will not lead to change in the next generation. This is critical because variation among individuals can be caused by non-genetic reasons, such as an individual being taller because of better nutrition rather than different genes.

Genetic diversity in a population comes from two main mechanisms: mutation and sexual reproduction. **Mutation**, a change in DNA, is the ultimate source of new **alleles** (alternative forms of a gene located at the same place on a chromosome), or new genetic variation in any population. The genetic changes caused by mutation can have one of three outcomes on the phenotype. A mutation affects the phenotype of the organism in a way that gives it reduced fitness—lower likelihood of survival or fewer offspring. A mutation may produce a phenotype with a beneficial effect on fitness. And, many mutations will also have no effect on the fitness of the phenotype; these are called neutral mutations. Mutations may also have a whole range of effect sizes on the fitness of the organism that expresses them in their phenotype, from a small effect to a great effect. Sexual reproduction also leads to genetic diversity: when two parents reproduce, unique combinations of alleles assemble to produce the unique genotypes and thus phenotypes in each of the offspring.

A heritable trait that helps the survival and reproduction of an organism in its present environment is called an **adaptation**. Scientists describe groups of organisms becoming adapted to their environment when a change in the range of genetic variation occurs over time that increases or maintains the “fit” of the population to its environment. The webbed feet of platypuses are an adaptation for swimming. The snow leopards’ thick fur is an adaptation for living in the cold. The cheetahs’ fast speed is an adaptation for catching prey.

Whether or not a trait is favorable depends on the environmental conditions at the time. The same traits are not always selected because environmental conditions can change. For example, consider a species of plant that grew in a moist climate and did not need to conserve water. Large leaves were selected because they allowed the plant to obtain more energy from the sun. Large leaves require more water to maintain than small leaves, and the moist environment provided favorable conditions to support large leaves. After thousands of years, the climate changed, and the area no longer had excess water. The direction of natural selection shifted so that plants with small leaves were selected because those populations were able to conserve water to survive the new environmental conditions.

Formation of New Species: Although all life on earth shares various genetic similarities, only certain organisms combine genetic information by sexual reproduction and have offspring that can then successfully reproduce. Scientists call such organisms members of the same biological species. For example, a mule is a cross between a donkey and a horse. However, since the mule cannot reproduce (donkey and horse do not have the same number of chromosomes), donkeys and horses are considered separate species. However, a cross between a Dachshund and a Norwegian Elkhound can result in offspring that can reproduce with another domestic dog, and thus all domestic dogs are considered the same species (all have the same number of chromosomes). **Speciation** (the formation of new species) occurs over a span of evolutionary time, so when a new species arises, there is a transition period during which the closely related species continue to interact.

The evolution of species has resulted in enormous variation in form and function. Sometimes, evolution gives rise to groups of organisms that become tremendously different from each other. When two species evolve in diverse directions from a common point, it is called **divergent evolution**. Such divergent evolution can be seen in the forms of the reproductive organs of flowering plants which share the same basic anatomies; however, they can look very different as a result of selection in different physical environments and adaptation to different kinds of pollinators. As you will see in the section on human evolution, human and apes lines diverged from one another millions of years ago, yet evidence of a common ancestry comes from many sources, including similarities in DNA.



(a) (b)

Figure 3.2.4: Flowering plants evolved from a common ancestor. Notice that the (a) dense blazing star (Liatrus spicata) and the (b) purple coneflower (Echinacea purpurea) vary in appearance, yet both share a similar basic morphology. (credit a: modification of work by Drew Avery; credit b: modification of work by Cory Zanker)

In other cases, similar phenotypes evolve independently in distantly related species. For example, flight has evolved in both bats and insects, and they both have structures we refer to as wings, which are adaptations to flight. However, the wings of bats and insects have evolved from very different original structures. This phenomenon is called **convergent evolution**, where similar traits evolve independently in species that do not share a recent common ancestry. The two species came to the same function, flying, but did so separately from each other.

3.2.3.1 Natural Selection and Adaptive Evolution

Natural selection only acts on the population's heritable traits: selecting for beneficial alleles and thus increasing their frequency in the population, while selecting against deleterious alleles and thereby decreasing their frequency—a process known as **adaptive evolution**. Natural selection does not act on individual alleles, however, but on entire organisms. An individual may carry a very beneficial genotype with a resulting phenotype that, for example, increases the ability to reproduce (fecundity), but if that same individual also carries an allele that results in a fatal childhood disease, that fecundity phenotype will not be passed on to the next generation because the individual will not live to reach reproductive age. Natural selection acts at the level of the individual; it selects for individuals with greater contributions to the gene pool of the next generation, known as an organism's evolutionary (Darwinian) fitness.

Fitness is often quantifiable and is measured by scientists in the field. However, it is not the absolute fitness of an individual that counts, but rather how it compares to the other organisms in the population. This concept, called relative fitness, allows researchers to determine which individuals are contributing additional offspring to the next generation, and thus, how the population might evolve.

3.2.4 Evidence of Evolution

The evidence for evolution is compelling and extensive. Looking at every level of organization in living systems, biologists see the signature of past and present evolution. Darwin dedicated a large portion of his book, On the Origin of Species, to identifying patterns in nature that were consistent with evolution, and since Darwin, our understanding has become clearer and broader.

3.2.4.1 Fossils

Fossils provide solid evidence that organisms from the past are not the same as those found today, and fossils show a progression of evolution. Scientists determine the age of fossils and categorize them from all over the world to determine when the organisms lived relative to each other. The resulting fossil record tells the story of the past and shows the evolution of form over millions of years. For example, scientists have recovered highly detailed records showing the evolution of humans and horses.



(a) (b)

Figure **3.2.5**: In this (a) display, fossil hominids are arranged from oldest (bottom) to newest (top). As hominids evolved, the shape of the skull changed. An artist's rendition of (b) extinct species of the genus Equus reveals that these ancient species resembled the modern horse (Equus ferus) but varied in size.

3.2.4.2 Anatomy and Embryology

Another type of evidence for evolution is the presence of structures in organisms that share the same basic form. For example, the bones in the appendages of a human, dog, bird, and whale all share the same overall construction (Figure 18.1.618.1.6) resulting from their origin in the appendages of a common ancestor. Over time, evolution led to changes in the shapes and sizes of these bones in different species, but they have maintained the same overall layout. Scientists call these synonymous parts **homologous structures.**

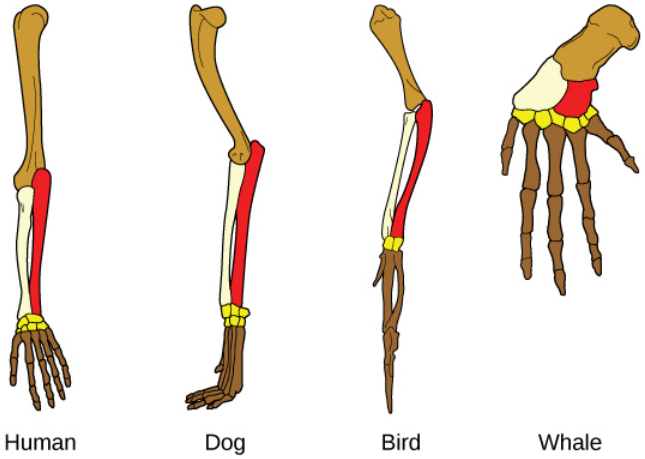


Figure **3.2.6**: The similar construction of these appendages indicates that these organisms share a common ancestor.

Some structures exist in organisms that have no apparent function at all, and appear to be residual parts from a past common ancestor. These unused structures without function are called **vestigial structures**. Examples of vestigial structures are wings on flightless birds, leaves on some cacti, and hind leg bones in whales.

Link to Learning

Visit this interactive site to guess which bones structures are homologous and which are analogous, and see examples of evolutionary adaptations to illustrate these concepts.

Another evidence of evolution is the **convergence of form** in organisms that share similar environments. For example, species of unrelated animals, such as the arctic fox and ptarmigan, living in the arctic region have been selected for seasonal white phenotypes

during winter to blend with the snow and ice. These similarities occur not because of common ancestry, but because of similar selection pressures—the benefits of not being seen, by prey or by predators, respectively.



(a) (b)

Figure **3.2.7**: The white winter coat of the (a) arctic fox and the (b) ptarmigan's plumage are adaptations to their environments. (credit a: modification of work by Keith Morehouse)

Embryology, the study of the development of the anatomy of an organism to its adult form, also provides evidence of relatedness between now widely divergent groups of organisms. Mutational tweaking in the embryo can have such magnified consequences in the adult that embryo formation tends to be conserved. As a result, structures that are absent in some groups often appear in their embryonic forms and disappear by the time the adult or juvenile form is reached. For example, all vertebrate embryos, including humans, exhibit gill slits and tails at some point in their early development. These disappear in the adults of terrestrial groups but are maintained in adult forms of aquatic groups such as fish and some amphibians. Great ape embryos, including humans, have a tail structure during their development that is lost by the time of (their) birth.

3.2.4.3 Biogeography

The geographic distribution of organisms on the planet follows patterns that are best explained by evolution in conjunction with the movement of tectonic plates over geological time. Broad groups that evolved before the breakup of the supercontinent Pangaea (about 200 million years ago) are distributed worldwide. Groups that evolved since the breakup appear uniquely in regions of the planet, such as the unique flora and fauna of northern continents that formed from the supercontinent Laurasia and of the southern continents that formed from the supercontinent Gondwana. The presence of members of the plant family Proteaceae in Australia, southern Africa, and South America is best explained by their presence prior to the southern supercontinent Gondwana breaking up.

The great diversification of marsupials in Australia and the absence of other mammals reflect Australia's long isolation. Australia has an abundance of endemic species—species found nowhere else—which is typical of islands whose isolation by expanses of water prevents species to migrate. Over time, these species diverge evolutionarily into new species that look very different from their ancestors that may exist on the mainland. The marsupials of Australia, the finches on the Galápagos, and many species on the Hawaiian Islands are all unique to their one point of origin, yet they display distant relationships to ancestral species on mainlands.

3.2.4.4 Molecular Biology

Like anatomical structures, the structures of the molecules of life reflect Darwin's "descent with modification" (i.e. evolution). Evidence of a common ancestor for all of life is reflected in the universality of DNA as the genetic material and in the near universality of the genetic code and the machinery of DNA replication and expression. Fundamental divisions in life between the three domains (archaea--single celled organisms without a cell nucleus; bacteria; and eukaryote--including plants, animals, and fungi) are reflected in major structural differences in otherwise conservative structures such as the components of ribosomes and

the structures of membranes. In general, the relatedness of groups of organisms is reflected in the similarity of their DNA sequences—exactly the pattern that would be expected from descent and diversification from a common ancestor.

DNA sequences have also shed light on some of the mechanisms of evolution. For example, it is clear that the evolution of new functions for proteins commonly occurs after gene duplication events that allow the free modification of one copy by mutation, selection, or drift (changes in a population's gene pool resulting from chance), while the second copy continues to produce a functional protein.

3.2.5 Misconceptions About Evolution

Although the theory of evolution generated some controversy when it was first proposed, it was almost universally accepted by biologists, particularly younger biologists, within 20 years after publication of *On the Origin of Species*. Nevertheless, the theory of evolution is a difficult concept and misconceptions about how it works abound.

Link to Learning

This site addresses some of the main misconceptions associated with the theory of evolution.

3.2.5.1 Evolution Is Just a Theory

Critics of the theory of evolution dismiss its importance by purposefully confounding the everyday usage of the word "theory" with the way scientists use the word. In science, a "theory" is understood to be a body of thoroughly tested and verified explanations for a set of observations of the natural world. Scientists have a theory of the atom, a theory of gravity, and the theory of relativity, each of which describes understood facts about the world. In the same way, the theory of evolution describes facts about the living world. As such, a theory in science has survived significant efforts to discredit it by scientists. In contrast, a "theory" in common vernacular is a word meaning a guess or suggested explanation; this meaning is more akin to the scientific concept of "hypothesis." When critics of evolution say evolution is "just a theory," they are implying that there is little evidence supporting it and that it is still in the process of being rigorously tested. This is a mischaracterization.

3.2.5.2 Individuals Evolve

Evolution is the change in genetic composition of a population over time, specifically over generations, resulting from differential reproduction of individuals with certain alleles. Individuals do change over their lifetime, obviously, but this is called development and involves changes programmed by the set of genes the individual acquired at birth in coordination with the individual's environment. When thinking about the evolution of a characteristic, it is probably best to think about the change of the average value of the characteristic in the population over time. For example, when natural selection leads to bill-size change in medium-ground finches in the Galápagos, this does not mean that individual bills on the finches are changing. If one measures the average bill size among all individuals in the population at one time and then measures the average bill size in the population several years later, this average value will be different as a result of evolution. Although some individuals may survive from the first time to the second, they will still have the same bill size; however, there will be many new individuals that contribute to the shift in average bill size.

3.2.5.3 Evolution Explains the Origin of Life

It is a common misunderstanding that evolution includes an explanation of life's origins. Conversely, some of the theory's critics believe that it cannot explain the origin of life. The theory does not try to explain the origin of life. The theory of evolution explains how populations change over time and how life diversifies the origin of species. It does not shed light on the beginnings of life including the origins of the first cells, which is how life is defined. The mechanisms of the origin of life on Earth are a particularly difficult problem because it occurred a very long time ago, and presumably it just occurred once. Importantly, biologists believe that the presence of life on Earth precludes the possibility that the events that led to life on Earth can be repeated because the intermediate stages would immediately become food for existing living things.

However, once a mechanism of inheritance was in place in the form of a molecule like DNA, either within a cell or pre-cell, these entities would be subject to the principle of natural selection. More effective reproducers would increase in frequency at the expense of inefficient reproducers. So while evolution does not explain the origin of life, it may have something to say about some of the processes operating once pre-living entities acquired certain properties.

3.2.5.4 Organisms Evolve on Purpose

Statements such as “organisms evolve in response to a change in an environment” are quite common, but such statements can lead to two types of misunderstandings. First, the statement must not be understood to mean that individual organisms evolve. The statement is shorthand for “a population evolves in response to a changing environment.” However, a second misunderstanding may arise by interpreting the statement to mean that the evolution is somehow intentional. A changed environment results in some individuals in the population, those with particular phenotypes, benefiting and therefore producing proportionately more offspring than other phenotypes. This results in change in the population, if the characteristics are genetically determined.

It is also important to understand that the variation that natural selection works on is already in a population and does not arise in response to an environmental change. For example, applying antibiotics to a population of bacteria will, over time, select a population of bacteria that are resistant to antibiotics. The resistance, which is caused by a gene, did not arise by mutation because of the application of the antibiotic. The gene for resistance was already present in the gene pool of the bacteria, likely at a low frequency. The antibiotic, which kills the bacterial cells without the resistance gene, strongly selects individuals that are resistant, since these would be the only ones that survived and divided. Experiments have demonstrated that mutations for antibiotic resistance do not arise as a result of antibiotic.

In a larger sense, evolution is not goal directed. Species do not become “better” over time; they simply track their changing environment with adaptations that maximize their reproduction in a particular environment at a particular time. Evolution has no goal of making faster, bigger, more complex, or even smarter species, despite the commonness of this kind of language in popular discourse. What characteristics evolve in a species are a function of the variation present and the environment, both of which are constantly changing in a non-directional way. What trait is fit in one environment at one time may well be fatal at some point in the future. This holds equally well for a species of insect as it does the human species.

3.2.6 **Population Evolution and the Modern Synthesis**

The mechanisms of inheritance, or genetics, were not understood at the time Charles Darwin and Alfred Russel Wallace were developing their idea of natural selection. This lack of understanding was a stumbling block to understanding many aspects of evolution. In fact, the predominant (and incorrect) genetic theory of the time, blending inheritance, made it difficult to understand how natural selection might operate. Darwin and Wallace were unaware of the genetics work by Austrian monk Gregor Mendel, which was published in 1866, not long after publication of Darwin's book, *On the Origin of Species* (1859). Mendel’s work was rediscovered in the early twentieth century at which time geneticists were rapidly coming to an understanding of the basics of inheritance. Initially, the newly discovered particulate nature of genes made it difficult for biologists to understand how gradual evolution could occur. But over the next few decades, genetics and evolution were integrated in what became known as **the modern synthesis**—the coherent understanding of the relationship between natural selection and genetics that took shape by the 1940s and is generally accepted today. In sum, the modern synthesis describes how evolutionary processes, such as natural selection, can affect a population’s genetic makeup, and, in turn, how this can result in the gradual evolution of populations and species. The theory also connects this change of a population over time, called microevolution, with the processes of macroevolution that gave rise to new species and higher taxonomic groups with widely divergent characters.

3.2.7 Kin Selection

In the discussion of natural selection, the emphasis was on how natural selection works on individuals to favor the more fit and disfavor the less fit in a population. The emphasis was on the survival (mortality selection), mating success (sexual selection), or family size (fecundity selection) of **individuals**. But what of the worker honeybee who gives up her life when danger threatens her hive? Or the mother bird who, feigning injury, flutters away from her nestful of young, thus risking death at the hands of a predator? How can evolution produce genes for such instinctive patterns of behavior when the owner of these genes risk failing the first test of fitness: survival?

A possible solution to this dilemma lies in the effect of such seemingly altruistic behavior on the overall ("inclusive") fitness of the family of the altruistic individual. Linked together by a similar genetic endowment, the altruistic member of the family enhances the chance that many of its own genes will be passed on to future generations by sacrificing itself for the welfare of its relatives. It is interesting to note that most altruistic behavior is observed where the individuals are linked by fairly close family ties. Natural selection working at the level of the family rather than the individual is called **kin selection**.

How good is the evidence for kin selection? Does the behavior of the mother bird really increase her chances of being killed? After all, it may be advantageous to take the initiative in an encounter with a predator that wanders near. But even if she does increase her risk, is this anything more than another example of maternal behavior? Her children are, indeed, her kin. But isn't natural selection simply operating in one of the ways Darwin described: to produce larger mature families?

Perhaps clearer examples of altruism and kin selection are to be found in those species of birds that employ "helpers". One example: Florida scrub jays (*Aphelocoma coerulescens coerulescens*). These birds occupy well-defined territories. When they reach maturity, many of the young birds remain for a time (one to four years) in the territory and help their parents with the raising of additional broods. How self-sacrificing! Should not natural selection have produced a genotype that leads its owners to seek mates and start raising their own families (to receive those genes)?

But the idea of kin selection suggests that the genes guiding their seemingly altruistic behavior have been selected for because they are more likely to be passed on to subsequent generations in the bodies of an increased number of younger brothers and sisters than in the bodies of their own children. To demonstrate that this is so, it is necessary to show that:

1. the "helping" behavior of these unmated birds is really a help and that
2. they have truly sacrificed opportunities to be successful parents themselves.

Thanks to the patient observations of Glen Woolfenden, the first point is established. He has shown that parents with helpers raise larger broods than those without. But the second point remains unresolved. Perhaps by waiting until they have gained experience with guarding nests and feeding young and until a suitable territory becomes available, these seemingly altruistic helpers are actually improving their chances of *eventually* raising larger families than they would have if they started right at it. If so, then once again we are simply seeing natural selection working through one of Darwin's criteria of individual fitness: ability to produce larger mature families.

The evolutionary advantage of helping ceases if the young are not actually siblings of the helper. It is well-established (e.g., by DNA analysis) that the females of many species of birds have "extramarital" affairs; that is, produce broods where the young have been sired by more than one male. Interestingly, it turns out that the more promiscuous the females of a given species, the less likely it is that they are assisted by helpers. Conversely, those species that employ helpers tend to be monogamous (however, there are a few exceptions.)

3.2.8 Kin Selection in Social Insects

The honeybee and other social insects provide the clearest example of kin selection. They are also particularly interesting examples because of the peculiar genetic relationships among the family members.

Male honeybees (drones) develop from the queen's unfertilized eggs and are haploid. Thus, all their sperm will contain exactly the same set of genes. This means that all their daughters will share exactly the same set of **paternal** genes, although they will share — on average — only one-half of their mother's genes. (Human sisters, in contrast, share one-half of their father's as well as one-half of their mother's genes.) So any behavior that favors honeybee sisters (75% of genes shared) will be more favorable to their genotype than behavior that favors their children (50% of genes shared).

Since that is the case, why bother with children at all? Why not have most of the sisters be sterile workers, caring for their mother as she produces more and more younger sisters, a few of whom will someday be queens? As for their brothers, worker bees share only 25% of their genes with them. Is it surprising, then, that as autumn draws near, the workers lose patience with the lazy demanding ways of their brothers and finally drive them from the hive?

3.2.9 No Perfect Organism

Natural selection is a driving force in evolution and can generate populations that are better adapted to survive and successfully reproduce in their environments. But natural selection cannot produce the perfect organism. Natural selection can only select on existing variation in the population; it does not create anything from scratch. Thus, it is limited by a population’s existing genetic variance and whatever new alleles (genetic variants) arise through mutation and gene flow (when some organisms from one population migrate into another population).

Natural selection is also limited because it works at the level of individuals, not alleles, and some alleles are linked together due to their physical proximity in the genome, making them more likely to be passed on together (linkage disequilibrium). Any given

individual may carry some beneficial alleles and some unfavorable alleles. It is the net effect of these alleles, or the organism's fitness, upon which natural selection can act. As a result, good alleles can be lost if they are carried by individuals that also have several overwhelmingly bad alleles; likewise, bad alleles can be kept if they are carried by individuals that have enough good alleles to result in an overall fitness benefit.

Finally, it is important to understand that not all evolution is adaptive. While natural selection selects the fittest individuals and often results in a more fit population overall, other forces of evolution, including genetic drift and gene flow, often do the opposite: introducing deleterious alleles to the population's gene pool. Evolution has no purpose—it is not changing a population into a preconceived ideal. It is simply the sum of the various forces described in this chapter and how they influence the genetic and phenotypic variance of a population.

3.2.10 Summary

Evolution is the process of adaptation through mutation and natural selection which allows better biologically fit (i.e. better adapted) characteristics to be passed to succeeding generations while less well adapted characteristics tend to be weeded out. Over time, organisms evolve characteristics that are beneficial to their survival and reproduction. For living organisms to adapt and change to environmental pressures, genetic variation must be present. With genetic variation, individuals have differences in form and function that allow some to survive environmental conditions better than others. These organisms pass their favorable traits to their offspring. Eventually, environments change, and what was once a desirable, advantageous trait may become an undesirable trait and organisms may further evolve. Evolution may be convergent with similar traits evolving in multiple species or divergent with diverse traits evolving in multiple species that came from a common ancestor. Evidence of evolution can be observed by means of DNA code and the fossil record, and also by the existence of homologous and vestigial structures.

The modern synthesis of evolutionary theory grew out of the combination of Darwin's and Wallace's formulations of evolution with Mendel's analysis of heredity, along with the more modern study of population genetics. The modern synthesis describes the evolution of populations and species, from small-scale changes among individuals to large-scale changes over paleontological time periods. To understand how organisms evolve, scientists can track populations' allele frequencies over time.

Kin selection and inclusive fitness involve selection for altruistic behaviors which benefit close genetic relatives.

3.2.11 References

1. Theodosius Dobzhansky. "Biology, Molecular and Organismic." *American Zoologist* 4, no. 4 (1964): 449.

3.2.12 Attributions

1. Evolution: Mechanisms and Evidence adapted from Evolution and Origin of Species OpenStax, licensed CC BY 4.0
2. Kin Selection and Kin Selection in Insects adapted from Libretexts, Biology written by John W. Kimball. This content is distributed under a Creative Commons Attribution 3.0 Unported (CC BY 3.0) license and made possible by funding from The Saylor Foundation.

This page titled 3.2: Evolution - Mechanisms and Evidence is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

- **18.1: Understanding Evolution** by OpenStax is licensed CC BY 4.0.

3.3: Human Evolution

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Explain the basic trends of human evolution including bipedalism and encephalization.
2. Describe hominins and what distinguishes hominins from other members of the Primate order.
3. Describe the category of features that distinguishes proto-hominins from hominins.
4. Describe, including approximate date ranges, the evolution of the genus *Homo*, including *e*arly *Homo* species and modern humans.
5. Explain how material culture helps inform us about the psychology of ancestral *Homo sapiens* and other hominin species.

3.3.1 Overview

Trends: There are a number of trends in the evolution of the proto-hominins (primitive ape-like species regarded as possibly ancestral to modern humans) to modern Homo sapiens. These traits do not occur all at once, but over millions of years.

Proto-Hominins and Hominins: In determining what fossil features a specimen must have in order to be classified as a hominin (the term used for humans and their ancestors after the split with chimpanzees and bonobos), many different characteristics are examined, including those related to bipedalism and dental features related to chewing. Apes have a space between biting teeth and grinding teeth where the upper canine fits when the jaws close. Other characteristics such as brain and body size are also considered.

Homo Genus: The emergence of the genus Homo (our genus) marks the advent of larger brains, the emergence of material culture (e.g. stone tools), and the eventual colonization of the world outside of Africa.

Material Culture: The earliest evidence of material culture is in the form of stone tools found on sites which dated back to 2.4 million years ago (see Supplementary Content, Chapter 18, Material Culture).



Figure **3.3.1**: The original complete skull (without upper teeth and mandible) of a 2.1 million years old Australopithecus africanus specimen so-called "Mrs. Ples" (catalogue number STS 5, Sterkfontein cave, hominid fossil number 5), discovered in South Africa . Collection of the Transvaal Museum, Northern Flagship Institute, Pretoria, South Africa. (CC BY-SA 4.0; José Braga; Didier Descouens).

3.3.2 Human and Ape Evolutionary Paths: Divergence from Common Ancestors

Before we begin our discussion of human evolution, it will be helpful to take a look at the big picture. The two figures below along with Figure 3.7.16, much further below, collectively depict 20 million years of evolution leading to our species, *Homo sapiens*. Figure 3.7.2 shows the last 10 million years of humanoid evolution. Figure 3.7.3 shows the last 20 million years of humanoid evolution and so is inclusive of Figure 3.7.2, which shows greater detail beginning with the Homininae (African hominids including humans) phylogeny, 10 Mya to present. Both of these figures show our genus Homo at the top left of each diagram. Figure 3.7.16 shows an enlarged and much more detailed depiction of the genus Homo including relationships between Homo

sapiens and Neanderthals and the fact that these two species once co-existed at the same time outside of Africa about 30,000-40,000 years ago. In the first two figures immediately below, notice the great divergence into a surprisingly large number of species over millions of years. Also notice how small and how recent the genus Homo is by comparison, and how one very recent species that emerged during the Pleistocene, humans, has so quickly come to dominate the Earth, at least for now. It is hoped that by closely examining these figures, you will be better able to organize the details that follow in this section and that you will be able to understand them within the larger context of primate evolution.

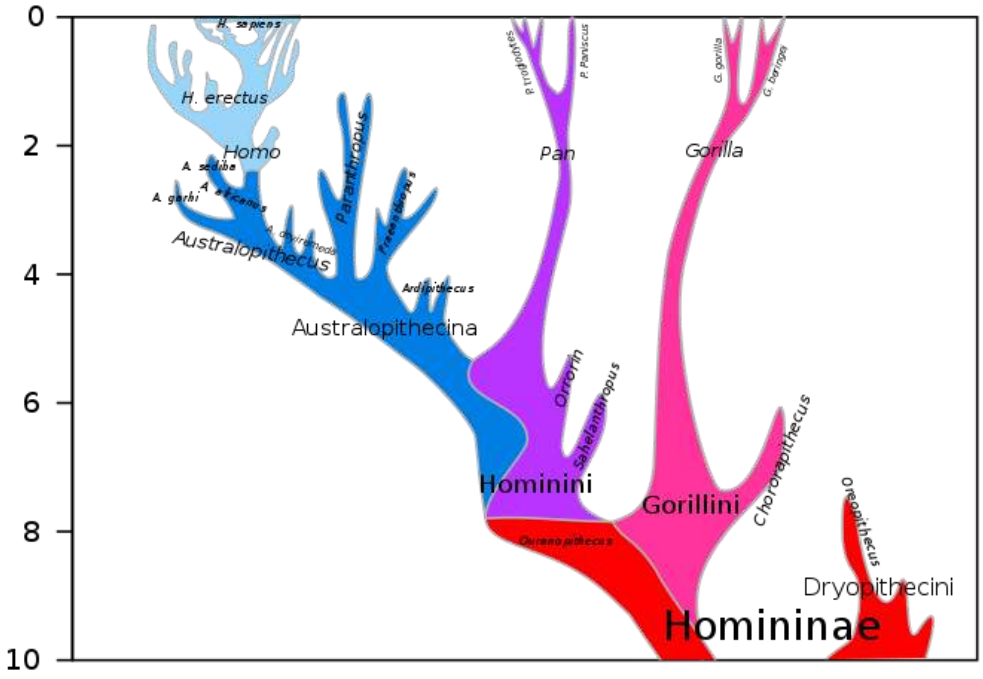


Figure **3.3.2**: Hominini (includes Homo and Pan, but excludes gorillas) and Homininae (African hominids including humans) phylogeny, 10 Mya to present; the human branch beginning with Homo, top left, is enlarged and shown in greater detail in Figure 3.7.16 closer to the bottom of this page. The Hominin ancestral line (modern humans, extinct human species and all our immediate ancestors including members of the genera *Homo*, *Australopithecus*, *Paranthropus* and *Ardipithecus*) and the ape line, including gorillas, diverged from one another about 8 Mya (million years ago) and the Hominin and Pan (chimpanzee) lines diverge from one another about 6 Mya. Homo sapiens are shown at the very top left. To go back 20 Mya, see Figure 3.7.3 below. There you can see that Pongo (orangutans) diverged from the Hominin line much earlier, about 14 million years ago (Image from Wikimedia Commons; File:Hominini lineage.svg; https://commons.wikimedia.org/wiki/F...ni_lineage.svg; by Dbachmann; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

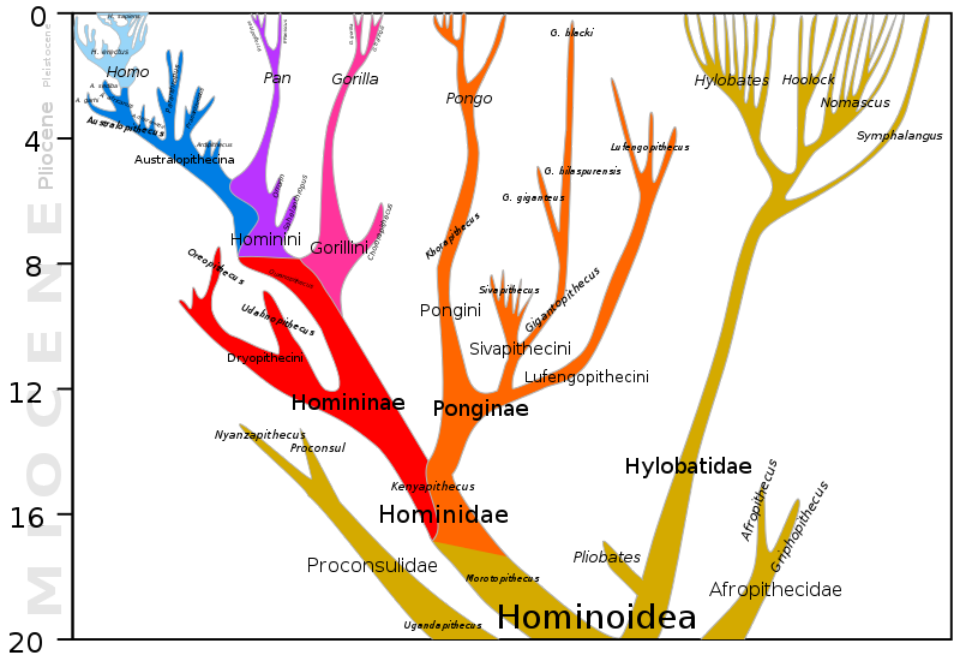


Figure **3.3.3**: Hominoidea phylogeny, 20 Mya to present. A hominoid, sometimes called an ape, is a member of the superfamily Hominoidea: extant members are the gibbons (lesser apes, family Hylobatidae) and the hominids. A hominid is a member of the family Hominidae, the great apes: orangutans (Ponginae; Pongo), gorillas (Gorillini), chimpanzees (Pan) and humans. Note that the split between the Homo line, leading to humans, from the chimpanzee and gorilla lineages occurred about six million and eight million years ago, respectively. (Image from Wikimedia Commons; File:Hominoidea lineage.svg; https://commons.wikimedia.org/wiki/F...ea_lineage.svg; by Dbachmann; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

3.3.3 Trends in Human Evolution

A number of questions about human evolution are:

- Why did our earliest ancestors stand up?
- Why did some species become extinct until only one species, *Homo sapiens*, was left?
- When, where and why did modern humans evolve?
- What was the role of the Neanderthals?
- What makes us human?

While hypotheses suggest answers, research continues to refine our understanding of human evolution.

This section provides an overview of key finds and trends in **hominin** evolution. (For additional information, see the "For Further Exploration" at the end of this page).



Figure **3.3.4**: A man contemplating his evolution.

3.3.4 Defining Hominins

It is through our study of our hominin ancestors and relatives that we are exposed to a world of “might have beens”--of other paths not taken by our species, of other ways of being human. But in order to better understand these different evolutionary trajectories, we must first define the terms we are using. If an imaginary line were drawn between ourselves and our closest relatives, the great apes, bipedalism (habitually walking upright on two feet) is where that line would be drawn. Hominin**,** then, means everyone on “our” side of that line defined by bipedalism. Thus, hominins are humans and all of our extinct bipedal ancestors and relatives since our divergence from the last common ancestor (LCA) that we shared with chimpanzees. As shown above in Figures 3.3.2 and 3.3.3, many scientists believe that this split between human and chimpanzee lineages took place about 6 million years ago. However, data from DNA using estimated mutation rates suggests that the split between the human and ape lineages may have occurred much earlier--divergence between human and chimpanzee lines taking place about 7 to 8 million and possibly as many as 13 million years ago, while the divergence between human and gorilla lines may have occurred as long ago as 8 to 19 million years ago (Gibbons, 2012). These different estimates may be confusing to students. However, this wide range of estimates serves as a good example of how our knowledge is incomplete and sometimes changing as new discoveries and methods arise. In this context, it is best to keep in mind that "[…] when all is said and done a taxonomy is just a hypothesis; it is not written on stone tablets" (Wood, 2010, p. 8908).

3.3.5 Morphological Trends in Human Evolution

There are a number of trends in the evolution of the human lineage from the proto-hominins (early human ancestral species) to modern *Homo sapiens*. These traits do not occur all at once, but over millions of years. In general, the trends include:

- the forward movement of the foramen magnum (the hole at the bottom of the skull where spinal cord enters skull to connect with brain): related to upright posture
- a reduction in the size of the canines
- an increase in the size of the molars
- an increase in cranial capacity
- flattening of the face
- rounding of the skull

Again, not all of these traits occur at the same time and there is variation among the various hominin species, but all of these morphological characteristics occur in the evolutionary line of *Homo sapiens*. Two other trends are especially important in the evolution of hominins:

- bipedalism, and
- encephalization of the brain.

These are discussed in more detail next.

3.3.5.1 Bipedalism

Bipedalism, or upright walking, was the first morphological trait on the road to humanity. Human bipedalism is the primary form of moving around (this is called habitual bipedalism). Other primates practice temporary or occasional bipedal locomotion, e.g., primates like chimpanzees may walk bipedally while they carry something in their hands. Few other animals are habitual bipeds like humans, e.g., birds and kangaroos.



Figure 3.3.5: Human skeleton walking.

There are numerous anatomical changes that evolved to make hominins efficient bipedal locomotors including changes in position of foramen magnum (see above), the vertebral column of the spine, the pelvis, femur and tibia bones, knee joint, and foot structure--many of these changes being related to improved balance and shock absorption (eLucy 2007).

The morphological changes associated with bipedalism take millions of years to evolve. They first appear 6.0-7.0 million years ago (mya), but are not fully in place until around 4.0 mya. These physical changes continue to refine until we see them as we do today in modern Homo sapiens (Jurmain et al. 2013).

3.3.5.2 Hypotheses about the Evolution of Bipedalism

Several hypotheses have been proposed over the last century or so to explain the evolution of hominins. As bipedalism is the first trait on the road to modern humans, these hypotheses focus on the emergence of habitual bipedalism. Many have been refuted as new data is discovered. The first hypothesis was the hunting hypothesis proposed by Charles Darwin. The hunting hypothesis claims that the key to human evolution was the shift from an arboreal life to a terrestrial one. He predicted that the earliest hominins would be found in Africa based on the similarities he saw between humans and African apes. He suggested that bipedalism gave the first hominins an advantage in that it freed up their hands to carry weapons used to hunt animals. Darwin also suggested that larger brains preceded bipedalism as intelligence was needed to make the tools. Now we know that habitual bipedalism predates large brains so Darwin's hypothesis is no longer considered an adequate explanation. With the discovery of new data, other hypotheses have been proposed including the patchy-forest and provisioning hypotheses.

The patchy forest hypothesis suggests that the emerging mosaic environment that began at the end of the Miocene made bipedalism advantageous. The phrase mosaic environment in this case refers to an environment that had patchy forest interspersed with grasslands that eventually became the African savannas of today. This caused food resources to become spread out over the landscape. For traveling long distances, bipedalism is more energy efficient than quadrupedalism (walking on "all fours"). Traveling bipedally freed up hands for carrying provisions and the early hominins could have easily fed from both terrestrial and arboreal resources.

The provisioning hypothesis states that having hands free to carry food allowed males to provision females and offspring. Since much of the females energy went to child-rearing, the ability of a male to provision her and her offspring would have been an attractive quality. Those males who could walk more efficiently bipedally while carrying food would have been prime mate material, allowing both the male and female to reproduce successfully. However, species such as wolves provision their females and offspring even though they are quadrupeds by consuming meat and regurgitating it to the pups.

The truth of the matter is that the origins of bipedalism are still murky. Further research will hopefully help us come closer to a determination of why bipedalism, and hence our early ancestors, evolved. In the meantime, you can explore other hypotheses on the origins of bipedalism on the NOVA web site: http://www.pbs.org/wgbh/nova/evoluti...ipedalism.html [optional].

3.3.5.3 Dental Features

Apes have a chewing complex which is good for cutting and shredding food. Over time, hominins lose this dental feature, as the canine reduces in size, and the molars increase in size (Larsen 2014).

3.3.5.4 Brain Evolution

In relation to other mammals, primates have a more expanded and elaborate brain, including expansion of the cerebral cortex. Compare the complexity of the human brain on the left to the cat brain on the right (photos are not to scale).

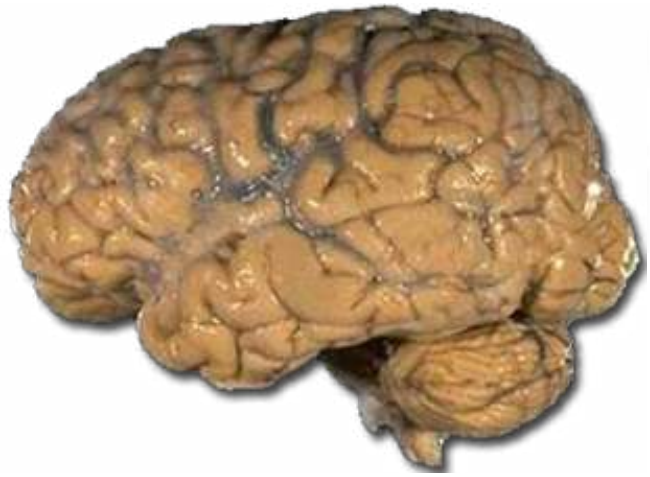


Figure **3.3.6**: Human Brain.



Figure **3.3.7**: Cat Brain.

Clearly, significant anatomical changes have taken place during the evolution of the brain in primates, in other mammals, and in animals in general. Are there generalizations that can be made about the evolution of animal brains?

Striedter (2006) has identified a number of general principles of brain evolution applicable across a wide range of species (i.e. not just primates or mammals).

1) Embryonic brains across species are more similar than adult brains, because brains tend to diversify more as they grow toward adult form;

2) relative brain size to body size in the vertebrates (animals with backbones) has tended to increase more often than decrease over evolutionary time;

3) it appears that increases in relative brain size were generally accompanied by increases in social or food foraging complexity;

4) most increases in relative brain size were accompanied by increases in absolute body size;

5) increases in absolute brain size require changes in the brain's internal connections which imply greater modularity or specialization (specialized processing modules increasing the "division of labor" as opposed to the whole brain doing all kinds of processing equally) of brain anatomy and functioning;

6) evolution generally enlarges brains by extending the period of brain development before and after birth while conserving (keeping the same) the "birth order" of different brain regions, so that big-brained animals tend to have disproportionately larger late-"born" (late-developing) regions ("late equals large"), such as cerebral cortex, leading to disproportionately more cerebral cortex (increased corticalization) in big-brained mammals (non-mammals don't have cerebral cortex); however there are exceptions to this rule, for example, at any given absolute brain size, there is more cerebral cortex in simians than in prosimians, and in parrots there is an unusually large telencephalon (forebrain) which is not accounted for by the rule;

7) changes in size proportions of brain areas, although "automatic" within the scaling (allometric) rules above, can still be adaptive and undergo natural selection;

8) as brain regions increase in absolute or proportional size, they tend to become laminated--organized into sheets of neurons--allowing point for point corresponding connections between sensory and motor maps with minimal axonal and dendritic wiring, saving space and metabolic energy;

9) as brain size increases, more regional subdivisions occur from ancestral parts subdividing into new parts, as in the dorsal thalamus (located below the cortex near the center of the brain), or, as in the case of neocortex, a new part was added onto an ancestral set of conserved (retained over evolution) brain parts;

10) a principle known as Deacon's rule is that "large equals well connected," meaning that as the relative size of a brain structure increases it tends to receive more connections and to project (send) more outputs to other structures.

Striedter (2006) adds a number of additional generalizations about the mammal and primate brains, including human brains:

11) six-layered mammalian neocortex (found only in mammals like us) probably evolved from a 3-layered reptilian precursor called dorsal cortex (something like that found in turtles) by addition of several layers of cortex;

12) aside from neocortex, the mammalian brain is similar to the reptilian brain (which also has hippocampus, for example) but even with a "fundamental scheme" of brain regions and circuitry, many minor changes in wiring can drastically change how information flows through a brain and thus how it functions--thus, the mammal brain is not just an upscale version of the reptilian brain;

13) increasing corticalization in mammals cannot be explained in terms of the above scaling (allometric) rules and involved highly specialized changes in brain anatomy presumably due to natural selection which expanded precursor sensory and motor cortical regions;

14) bird forebrains evolved along a very different path with expansion of their dorsal ventricular ridge (DVR), the major sensorimotor region of the avian telencephalon, highly similar in function to mammalian neocortex, making "many birds at least as intelligent as most mammals."

Striedter adds a number of points about the **human brain** in an attempt to identify features that make it special compared to the brains of other mammals.

15) In the six million years since bipedal apes (hominins) diverged from other apes, absolute brain size increased radically (about fourfold), not gradually, but in bursts--**from when genus *Homo* first evolved, absolute brain size doubled from 400 to 800 cubic centimeters, then remained relatively steady in *Homo erectus* during the next 1.5 million years, but then exploded again in the transition to *Homo sapien* until about 100,000 years ago, at which time absolute brain size reached its current value of about 1,200 to 1,800 cubic centimeters.** The first jump in Homo brain size was likely related to change in diet involving transition to meat and later the cooking of meat. The second leap was perhaps stimulated by competition among humans for mates and other resources;

16) the principle of "late equals large" predicts large neocortex in humans (the human neocortex to medulla ratio is twice that of chimpanzees);

17) the principle of "large equals well connected" is consistent with known expanded numbers of projections from human neocortex to motor neurons in medulla (located just above spinal cord) and spinal cord permitting greater precision of control over muscles serving hands, lips, tongue, face, jaw, respiratory muscles, and vocal folds, required for the **development of human language about 50,000 to 100,000 years ago**;

18) once human language appeared, dramatic changes in human behavior became possible without further increases in brain size;

19) increase in brain size has some disadvantages: including increased metabolic costs because the brain utilizes so much metabolic energy (20% of human metabolic energy even though it is only 2% of human body weight; being so metabolically expensive increases in brain size must be paid for by improved diet or reduction of other metabolic energy demands); decreased connectivity perhaps making the two hemispheres more independent of one another, and perhaps explaining why the two cerebral hemispheres became functionally specialized (performing different cognitive functions); and size limits of neonatal brain due to constraints imposed by size of the human mother's pelvis and birth canal. According to Striedter, these costs may explain why human brain size plateaued about 100,000 years ago;

20) within neocortex (found only in mammals), the lateral prefrontal cortex (located forward of your temples) has become relatively enlarged in the human brain, likely increasing its role in behavior (see Chapter 14 for additional discussion of the the lateral prefrontal cortex and higher cognitive functions such as thought, planning, etc.);

21) some key evolutionary changes in brain structure were not caused by increases in absolute or relative brain size, such as evolution of the neocortex of mammals, but require additional explanation; comparing distantly related species on absolute brain size alone misses important factors, for example, brains of some large whales weight 5 times as much as the human brain but whale brains have poorly laminated and thin neocortex; in cases of distantly related species, comparisons of relative brain size are more

useful, for example, human and some toothed whales (e.g. killer whales) have relative brain sizes significantly larger than average mammals of similar body size;

22) two general hypotheses about brain evolution are that individual brain systems evolve independently by natural selection (the mosaic hypothesis) or alternatively that components of such systems evolve together because of functional constraints (the concerted or constraint hypothesis), with a third view being that all brain evolution is simultaneously both mosaic and concerted.

One problem with Striedter's approach is that it doesn't explicate the forces of natural selection that may account for more specific features of brain evolution in specific species, including our own. He admits that evolution of the neocortex cannot be explained by evolution of bigger brains and that neocortex evolved independently of absolute brain size. However, aside from restating the theory that complex social life in ancestral humans, along with competition among humans for resources, stimulated the evolution of increases in the size and complexity of the neocortex, he offers little insight about what role natural selection played in the evolution of the neocortex of mammals, or in brain evolution in general. As Adkins-Regan (2006, p.12-13) states in her critique of Striedter's work, "There is relatively little discussion of tests of hypotheses about the selective pressures responsible for the origin and maintenance of traits . . . The author would seem to be experiencing symptoms of discomfort with the concept of adaptation. . .. Given that brain mechanisms are products of natural selection, a central strategy in neuroscience should be to use the methods of evolutionary biology, which have been so successful in helping us to understand the mechanisms and design of organisms generally." In Chapter 14 of this text, on Intelligence and Cognition, consistent with the critique of Adkins-Regan, you will find an extensive discussion of the role of natural selection in evolution of brain systems involved in intelligence and thinking.

3.3.5.5 Encephalization of the Brain

Encephalization of the brain refers to a couple of things: 1) the **increase in brain size** over time and 2) the **size of the brain in relation to total body mass**. The brain-size to body mass ratio does not change that much in the hominins in spite of presumed increases in intelligence, however increases in brain size becomes significant beginning with the early Homo species. Studies of living mammal species suggest that **increased number of neurons and increased density of neuron packing into the cranial cavity may be a key factor in the evolution of increased intelligence across species.** However, while there is gradual increase in brain size throughout the australopithecine lineage, it is not until early *Homo* that there is a significant increase in cranial capacity, approximately a 20% increase over australopithecines. More significant is the approximately 50% increase in brain size of *Homo erectus* and the earlier *Homo* species. It is not just the size of the brain that is important. During this process of encephalization, it has been speculated, without fossil evidence, that there was also a rewiring of the brain that coincides with the emergence of material culture such as stone tools. It is not until this occurs that hominins leave Africa, enabled greatly by cultural advances.

Non-human primate brains are symmetrical as are the brains of early hominins. **With the emergence of *Homo* we see the lateralization of the brain**--it becomes asymmetrial (right brain, left brain). We know this from endocasts. Endocasts form when minerals replace brain matter inside the cranium during the fossilization process. These endocasts allow researchers to study the cortical folds of the brain and compare it to modern humans. Based on endocasts, researchers determined that three areas of the brain began to **change in *Homo*:** the **cerebellum**, which handles learned motor activities, the **limbic system**, which processes motivation, emotion and social communication, and the **cerebral cortex**, which is responsible for sensory experiences, memory, and complex mental functions such as language, cognition, planning, imagination, and intelligence (see chapters 14 and 15 on Cognition and Intelligence and Language). It is these changes that may have allowed the early members of our genus, *Homo*, to develop **cultural adaptations to environmental pressures**.

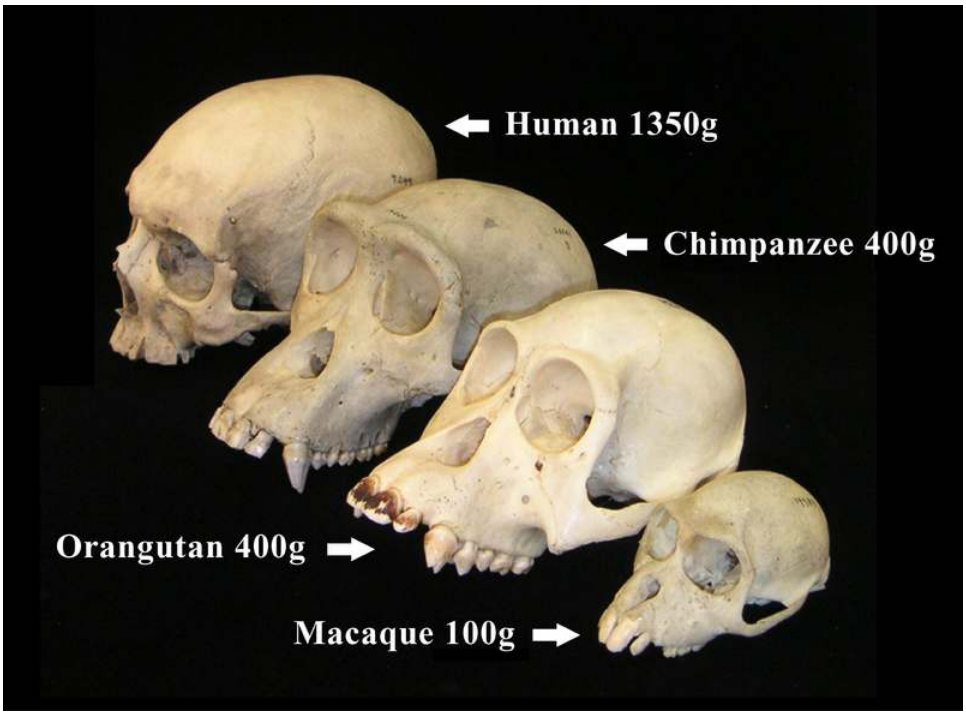


Figure 3.3.8: Comparison of cranial capacities of living primates. Primate skulls provided courtesy of the Museum of Comparative Zoology, Harvard University. (Image from Wikimedia Commons; File:Primate skull series with legend.png; https://commons.wikimedia.org/wiki/F...ith_legend.png; by Christopher Walsh, Harvard Medical School; licensed under the Creative Commons Attribution-Share Alike 2.5 Generic license).

3.3.5.6 Why did the brain change in early Homo?

The question that confronted researchers was why the brain changed. Big brains have some disadvantages:

- they require a lot of metabolic energy; approximately 25-30% of a human's metabolic energy is consumed by the brain although it comprises only 2% of total body weight
- big brains require infants to be born in an immature state with head and brain size small enough to permit birth through a relatively small birth canal, resulting in a longer period of infant dependency (the average infant brain at birth is only about 1/3 the size of an adult brain with much of brain development and behavioral and mental development occurring over a period of years after birth)
- longer infant dependency is an increased drain on maternal energy; the mother must have proper nutrition not only for herself but for the nursing infant
- it has been suggested that larger brains decrease the bipedal efficiency of females because they must have a wider pelvis and birth canal to give birth to a large brained infant

So, for large brains to become evolutionarily fixed in the *Homo* population, the advantages had to outweigh the disadvantages listed above.

One possible explanation incorporates the interaction of three different variables: group size, complex subsistence patterns (such as foraging for food or domestication of animals and plants), and the nutritional value of meat (Campbell and Loy 2000: 318). Let's address group size first.

Research suggests that brain size and size of social groups correlate positively among living primates, implying that big brains helped individuals keep track of complex social information such as dominance hierarchies, alliances, enemies, etc.

Second, a big brain allows primates to keep track of large subsistence territories and allows omnivores to develop strategies for collecting a wide variety of foods.

Third, eating meat is a relatively easy way to get the nutrition needed to run a big brain, which, as mentioned above, in modern humans takes about 1/4 to 1/3 of our daily metabolic energy. However, raw meat requires a lot of energy to digest, so the invention of cooking meat over a fire may have played a role in evolution of larger brains because cooked meat is much easier, and requires less metabolic energy, to digest. Thus, cooking of meat provided a rich source of metabolic energy, supporting evolution of larger brains.

The argument for the social brain hypothesis is laid out by Robin Dunbar (1998). Dunbar also claims that it was changes in the neocortex, the 2-4 mm thick top layer of the cerebral hemispheres, that were critical in the "hominization" (development of human cognitive abilities) of our ancestors (see chapter 14 for additional adaptive advantages associated with encephalization and increases in intelligence).

A related perspective on the evolution of the human brain emphasizes the role of culture in increasing brain size. As Geertz (2013, p. 180-181) states:

> The Australopithecine brain was approximately the same size as that of modern-day chimpanzees (some 400– 600 cm^3). The Homo sapiens brain is approximately 1200–1700 cm^3. But the expansion of the brain had begun already with the appearance of Homo habilis, the first hominine species that came out of the Australopithecine line some 2.5 million years ago (with a brain size of 500–800 cm^3) and became even more spectacular with the appearance 1.5 million years ago of the stone tool artist Homo erectus whose brain grew closer in size to modern humans at 750–1250 cm^3. It is today assumed by archeologists and paleontologists that the production and use of tools was incremental to the expansion of the brain. Because tool use is much more than simply knocking stones and bones together, and indeed, it depends very much on the recipes, rules and instructions of cultural patterns, it is widely accepted today that culture drove the expansion of the brain (along with other things, of course, such as eating meat). The causal chain, once again, however, is culture first, brain expansion afterwards. . . . 'Because tool manufacture puts a premium on manual skill and foresight, its introduction must have acted to shift selection pressures so as to favor the rapid growth of the forebrain as, in all likelihood, did the advances in social organization, communication, and moral regulation which there is reason to believe also occurred during this period of overlap between cultural and biological change. Nor were such nervous system changes merely quantitative; alterations in the interconnections among neurons and their manner of functioning may have been of greater importance than the simple increase in their number.'

The Forebrain includes the cerebral cortex (frontal, parietal, temporal, and occipital lobes) and a number of subcortical structures such as the thalamus, hypothalamus, limbic system, and basal ganglia. Different selection pressures arising from tool manufacture and use, complex social organization, communication, moral regulation, and other social and cultural factors would have likely affected different areas of the forebrain, and even different lobes of the cerebral cortex, in different ways. This follows from the fact of brain organization that different brain circuits and networks are specialized for different cognitive and behavioral functions (see Cosmides & Tooby, 2002).

3.3.5.7 Geology and Environmental Background

The Miocene period (roughly 23-5 million years ago) was geologically active in Africa. This is the period of the adaptive radiation of the apes and a period of mountain building that led to the formation of the Great Africa Rift Valley. With the emergence of the rift mountains, the rains that heretofore had moved across the continent from the Atlantic Ocean were blocked (referred to as a rain shadow), leading to the aridification of Eastern Africa. The savanna environment that evolved in Eastern Africa was and is a much more open environment than the forested environment of Western and Central Africa, leading to rise of new adaptations for plants and animals. It is in this newly emerging environment that hominin evolution takes off.

Paleoclimatic data has been correlated with speciation events in hominin evolution, but it does not seem to account for all speciation events. Nonetheless, the paleoclimatic data suggests the following:

- Grasslands spread in Africa between 10-5 million years ago during a cooling and drying phase. It is during this time frame that the common ancestor of African apes and humans lived. The common ancestor was more like a quadruped who was arboreal or at least spent a significant amount of time in the trees. In the middle of this period, approximately 7-6 million years ago, the first bipedal hominin emerged, and a few other early hominins are referred to as proto-hominins in recognition of their primitive, ape-like features.

- In the mid-Pliocene period, 3-2 million years ago, yet another cooling and drying phase is correlated with the adaptive radiation of the hominins, including the emergence of the robust australopithecines in the genus *Homo*, the same genus as modern humans.
- Near the beginning of the Pleistocene period, also known as the Ice Age, the environment continued to get drier. Open habitats spread in East Africa. During this period, *Homo ergaster* (*Homo erectus*) emerges and finally leaves the African continent.

This data has a tendency to make us think that hominin evolution was driven by environmental changes; however, geologic, climatic, and environmental changes in Africa during the Miocene, Pliocene and Pleistocene may have had little to do with the evolution of hominins, leaving open the possibility that the hypotheses discussed above, or others yet to be proposed, may be more causally related to the evolution of hominins.

3.3.6 Key Transitions in Human Evolution

On the evolutionary road to modern humans, some scientists have identified several "key transitions: (1) African ape to terrestrial bipedal ape (around 4 Ma)[Ma=million years ago]; (2) terrestrial bipedal ape (australopithecine) to ‘early Homo’ (around 2 Ma); (3) Early Homo [species] to Homo heidelbergensis (1–0.8Ma); (4) Homo heidelbergensis to larger-brained Homo (from 500 ka) [ka=thousand years ago], and (5) larger-brained Homo to H. sapiens (from 200 ka) . . . [Regarding social behavior,] "among archeologically based models, there is a total agreement that the emergence of basal hominin sociality, as part of our primate heritage, should be attributed to the very first evolutionary stages, possibly before the emergence of the Homo genus" (Anghelinu, 2013, p.13).

3.3.6.1 For Further Exploration

3.3.6.1.1 Explore Human Evolution in Print

- Boyd, Robert and Joan B. Silk. 2009. How Humans Evolved, 5th edition. New York: W. W. Norton.
- Campbell, Bernard G. and James D. Loy. 2000. Humankind Emerging, 8th edition. Boston: Allyn & Bacon.
- Johanson, Donald and Kate Wong. 2010. Lucy's Legacy: The Quest for Human Origins. New York: Harmony Books.
- Stringer, Chris and Peter Andrew. 2006. The Complete World of Human Evolution. New York: Thames & Hudson.
- Tattersall, Ian. 2008. The Fossil Trail: How We Know What We Think We Know About Human Evolution. New York: Oxford University Press.

3.3.6.1.2 Explore Human Evolution on the Web

- Becoming Human
- Talk: Origins Fossil Hominids
- Hall of Human Origins
- Science Daily: Human Evolution News
- Rediscovering Biology:Unit 9 Human Evolution
- BBC: The Evolution of Man
- Human Evolution: The Fossil Evidence in 3D

3.3.7 Proto-hominins

The oldest hominin discovered to date is from between 7.2 to 6.8 mya (million years ago).

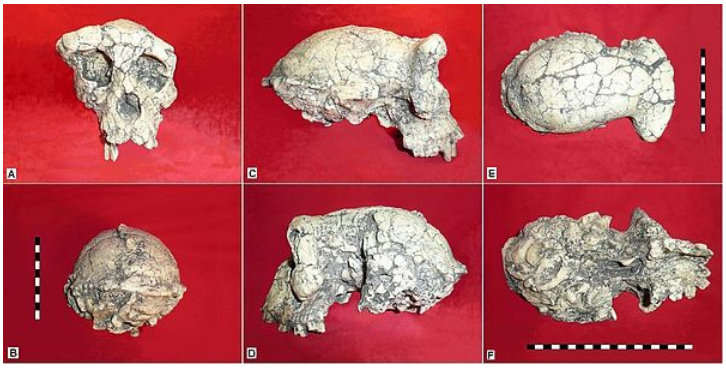


Figure **3.3.9**: *Sahelanthropus tchadensis*. The oldest known hominin.

The skull is a combination of ape-like and human-like features. Ape-like features include brain size, heavy brow ridge. Its human-like features include forward position of the foramen magnum, smaller canine teeth, and intermediate thickness of the premolar and molar enamel. Due to the pronounced brow ridge, Brunet's team suggests that the specimen is male.

There is debate among researchers as to whether this fossil is a hominin or an ape. Some suggest that the specimen belongs to that of a female ape because it is likely to find canines worn at the tips in female apes.

3.3.8 Hominins

While the hominins will be presented more or less in chronological order, do not mistake chronological order for linear evolutionary relationships; some hominins that are presented are not in the direct line to modern humans. It is also important to keep in mind that new discoveries are made each year that refine what we know about human evolution.

It is highly recommended that you begin your exploration of human evolution by watching the PBS documentary, Becoming Human (https://www.pbs.org/wgbh/nova/video/...-human-part-1/).

3.3.9 Early hominins

3.3.9.1 *Australopithecus afarensis*

Discovered in 1974 at Hadar, Ethiopia, *Australopithecus afarensis* is arguably the most well know fossil hominin species. It is dated from 3.7-3.0 mya (Scarre 2014). Over 40% of the skeleton was recovered, which allowed the team to fully reconstruct the skeleton. This fossil specimen, named Lucy, coupled with footprints found at Laetoli, Tanzania, in 1978 by Mary Leakey, confirmed that *Au. afarensis* was fully bipedal, albeit not exactly like modern humans. The footprints at Laetoli indicate that *Au. afarensis* had a short stride and a strolling gait. Since the 1970s, hundreds of specimens of Au. afarensis have been found (60 individuals at least from Hadar alone!) in Ethiopia, Kenya, and Tanzania, allowing paleoanthropologists to make "definitive statements about the locomotor pattern and stature of" (Jurmain 2013: 211) this early hominin.



Figure **3.3.10**: Reconstruction of the fossil skeleton of "Lucy" the *Australopithecus afarensis*

Au. afarensis has several primitive, or ape-like, features, including a relatively small brain in comparison to *Homo*, a u-shaped dental arcade, a flat nose, a flattened forehead (referred to as platycephaly) and prognathic face (characterized by a protruding lower jaw). Its canines, while larger than *Homo*, are smaller than earlier hominins. While its brain was larger than earlier hominins, it is still small in comparison to genus Homo. There is evidence for sexual dimorphism; *Au. afarensis* males were no taller than 5 feet and females about 3-4 feet, similar in proportion to modern African apes. *Au. afarensis* has a suite of derived traits associated with bipedalism; i.e., bowl-shaped pelvis, s-curve of the vertebral column, and knee anatomy. However, the curvature of the fingers and toes and the proportion of the arms to legs suggest to some researchers that *Au. afarensis* spent some time in the trees.

In 2006, a 3.3 million year old *Au. afarensis* child was discovered less than 4 km from where Lucy was found in 1974. It is not only the oldest juvenile fossil ever found, it is also the most complete hominin fossil found to date. Selam, as the fossil was nicknamed (or Lucy's baby or Dikika's baby), confirmed earlier suggestions that *Au. afarensis* was bipedal, yet spent time in the trees. The shoulder structure with its upward pointing shoulder joints and the bony ridge running along the shoulder blades are like that of apes, which would have facilitated arboreal movement even if they were not as fully capable as chimpanzees at moving in trees.

Paleoecological data indicates that *Au. afarensis* lived in both grassland (savanna) and woodland environments.

3.3.9.2 Australopithecus africanus



Figure **3.3.11**: *Australopithecus africanus* (Mrs. Ples)

Australopithecus africanus (see Figure 3.7.2, above) has been dated to 3-2 mya (Scarre 2014). In comparison to *Au. afarensis*, *Au. africanus* has smaller incisors and larger molars; the canines no longer have the pointed, triangular appearance seen in apes and *Au. afarensis*; however, *Au. africanus* still exhibits some protrusion of the lower jaw and has a small brain like that of *Au. afarensis*. There is general consensus that *Au. africanus* is a direct descendent of *Au. afarensis*. Its relationship to *Homo* is less clear.

3.3.9.3 *Australopithecus garhi*

Found in Bouri, Ethiopia in 1997, *Au. garhi* (see Figure 3.7.2, above) dates to 2.5 mya. Few fossil specimens have been found and those that have are relatively fragmentary. One cranium and other skull fragments were found and serve as the basis of the species identification. The size and shape of its molar teeth suggest to some researchers that *Au. garhi* is related to *Paranthropus aethiopicus* (see below), but its other features, e.g., braincase, face, and other teeth, are more like genus *Homo*. In light of this, some researchers contend it is ancestral to *Homo*.



Figure **3.3.12**: *Australopithecus garhi.*

3.3.9.4 *Australopithecus sediba*

Australopithecus sediba (Figure 3.3.13) was found in 2008 at Malapa Cave, South Africa, dating to 1.95-1.75 mya, has a mosaic of characteristics that suggest it may be transitional from the australopiths to genus *Homo*. However, this claim is controversial as the earliest dates for *Homo* predate *Au. sediba* by about 500,000 years (Becoming Human 2008).



Figure **3.3.13**: *Australopithecus sediba.*

The features that link *Au. sediba* to *Homo* include the pelvis shape, more vertical brain case, smaller cheek bones, and molar shape.

3.3.9.5 Robust Australopiths

Three robust species of hominins emerged in the Plio-Pleistocene period: *Paranthropus aethiopicus*, *Paranthropus boisei*, and *Paranthropus robustus* (see Figures 3.3.14-17). They have morphological features that suggest they were well adapted for eating hard foods that needed grinding, which led to their being identified as "robust."

> **Note**
>
> There is debate over whether the differences mentioned above qualify the robust australopiths to be in a separate genus from the australopithecines. In some anthropological works the genus *Australopithecus* is used. In others, such as this work, *Paranthropus* is used.



Figure **3.3.14**: *Paranthropus aethiopicus.*



Figure 3.3.15. *Paranthropus boisei*.

There is no evidence to suggest that *P. boisei* is ancestral to any subsequent hominin.

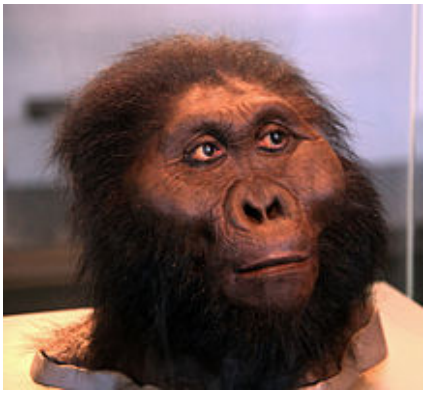


Figure 3.3.16. *Paranthropus boisei*; model of adult male (Smithsonian Museum of Natural History).



Figure 3.3.17. *Paranthropus robustus.*

3.3.10 Homo Genus

Above we described our oldest human ancestors, primarily members of the genus Australopithecus who lived between 2 million and 4 million years ago. Here we introduce the earliest members of the **genus Homo**, focusing on the species *Homo habilis* and *Homo erectus* (see Figure 3.3.19 and Figure 3.3.20, below).

3.3.10.1 Defining The Genus Homo

When grouping species into a common genus, biologists will consider criteria such as physical characteristics (morphology), evidence of recent common ancestry, and adaptive strategy (use of the environment). However, there is disagreement about which of those criteria should be prioritized, as well as how specific fossils should be interpreted in light of the criteria. There is general agreement that species classified as Homo should share characteristics broadly similar to our species. These include the following:

• a relatively large brain size, indicating a high degree of intelligence;

• a smaller and flatter face;

• smaller jaws and teeth; and

• increased reliance on culture, particularly the use of stone tools, to exploit a greater diversity of environments (adaptive zone).

Some researchers would include larger overall body size and limb proportions (longer legs/shorter arms) in this list. There is also an apparent decline in sexual dimorphism (body-size differences between males and females). While these criteria seem relatively clear-cut, evaluating them in the fossil record has proved more difficult, particularly for the earliest members of the genus. There are several reasons for this. First, many fossil specimens dating to this time period are incomplete and poorly preserved, making them difficult to evaluate. Second, early Homo fossils appear quite variable in brain size, facial features, and teeth and body size, and there is not yet consensus about how to best make sense of this diversity.

In this section, we will take several pathways toward examining the origin and evolution of the genus Homo. First, we will explore the environmental conditions of the Pleistocene epoch in which the genus Homo evolved. Next we will examine the fossil evidence for the two principal species traditionally identified as early Homo: Homo habilis and Homo erectus. Then we will use data from fossils and archeological sites to reconstruct the behavior of early members of Homo, including tool manufacture, subsistence practices, migratory patterns, and social structure. Finally, we will consider these together in an attempt to characterize the key adaptive strategies of early Homo and how they put our early ancestors on the trajectory that led to our own species, Homo sapiens.

3.3.10.2 Climate

The early hominin species covered previously, such as Ardipithecus ramidus and Australopithecus afarensis, evolved during the late Pliocene epoch. The Pliocene (5.3 million to 2.6 million years ago) was marked by cooler and drier conditions, with ice caps forming permanently at the poles. Still, Earth's climate during the Pliocene was considerably warmer and wetter than at present.

The subsequent Pleistocene epoch (2.6 million years to 11,000 years ago) ushered in major environmental change. The Pleistocene is popularly referred to as the Ice Age. Since the term "Ice Age" tends to conjure up images of glaciers and woolly mammoths, one would naturally assume that this was a period of uniformly cold climate around the globe. But this is not actually the case. Instead, climate became much more variable, cycling abruptly between warm/wet (interglacial) and cold/dry (glacial) cycles.

In Africa, paleoclimate research has determined that grasslands expanded and shrank multiple times during this period, even as they expanded over the long term, becoming increasingly common during the Pleistocene. One solution adopted by some hominins was to specialize in feeding on the new types of plants growing in this landscape. The robust australopithecines probably developed their large molar teeth with thick enamel in order to exploit this particular dietary niche.

Members of the genus Homo took a different route. Faced with the unstable African climate and shifting landscape, they evolved bigger brains that enabled them to rely on cultural solutions such as crafting stone tools that opened up new foraging opportunities. This strategy of behavioral flexibility served them well during this unpredictable time and led to new innovations such as increased meat-eating, cooperative hunting, and the exploitation of new environments outside Africa, including Europe and Asia.

3.3.11 The Emergence of Homo (our Genus)

The emergence of the genus *Homo* marks the advent of larger brains, the emergence of material culture (at least material culture that survives in the archeological record), and the eventual colonization of the world outside of Africa. The earliest *Homo* species are contemporaneous with several australopiths: *Au. africanus*, *Au. garhi*, *Au. sediba*, and all of the *Paranthropus* species. Africa was flush with hominins (Figure **3.3.16**). There are several trends we see in the evolution of the earliest *Homo* species to humans, *Homo sapiens*:

- Rounding of the cranium
- Enlargement and rewiring of the brain (judged from endocasts)
- Smaller faces and teeth
- Decreasing prognathism
- Tallness
- Diversity of cultural traits

As with the overview on early hominins, we will continue taking the "lumper" approach - a tendency to lump specimens into fewer species rather than tending to differentiate and hypothesizing a greater number of species. Several *Homo* species will not be discussed in detail, but may be mentioned in passing, e.g., *Homo rudolfensis*, *Homo ergaster*, *Homo gautengensis*, *Homo antecessor*, *Homo cepranensis*, *Homo rhodesiensis*, *Homo tsaichangensis*. We will first examine the morphological characteristics of various *Homo* species. Cultural traits will be addressed separately.

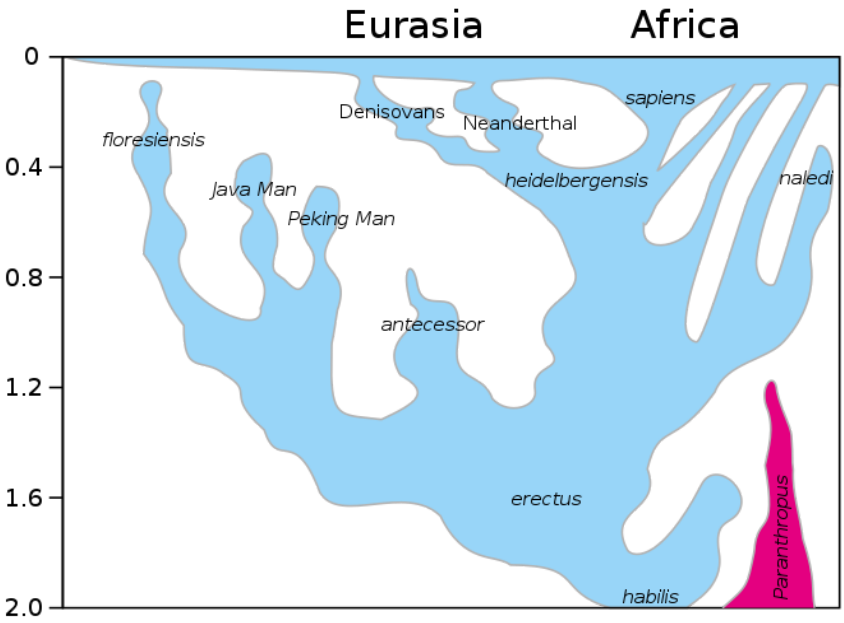


Figure **3.3.18**: A model of the evolution of the genus Homo over the last 2 million years (millions of years, Mya, is on the vertical axis). The rapid "Out of Africa" expansion of H. sapiens is indicated at the top of the diagram by the lateral expansion of blue all the way across the top of the diagram, with admixture indicated with Neanderthals, Denisovans, and unspecified archaic African hominins. The late survival of robust australopithecines (Paranthropus) alongside Homo until 1.2 Mya is indicated in purple.

- H. heidelbergensis is shown as the link between Neanderthals, Denosiovans and H. sapiens
- Division of Asian H. erectus into Java Man and Peking Man
- H. antecessor shown as a branch of H. erectus reaching Europe
- After H. sapiens emerge from Africa some 60 kya they spread across the globe and interbred with other descendants of H. heidelbergensis, and Neanderthals, Denisovans

(Image from Wikimedia Commons; File:Homo lineage 2017update.svg; https://commons.wikimedia.org/wiki/F...2017update.svg; by User:Conquistador, User:Dbachmann; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

3.3.12 Homo Habilis: The Earliest Members of our Genus

Homo habilis has traditionally been considered the earliest species placed in the genus Homo. However, as we will see, there is substantial disagreement among paleoanthropologists about the fossils classified as Homo habilis, including whether they come from a single or multiple species, or even whether they should be part of the genus Homo at all.

Compared to the australopithecines, **Homo habilis has a somewhat larger brain size–an average of 650 cubic centimeters (cc) compared to less than 500 cc for Australopithecus.** Additionally, the skull is more rounded and the face less prognathic. However, the postcranial remains show a body size and proportions similar to Australopithecus. Known dates for fossils identified as Homo habilis range from about 2.5 million years ago to 1.7 million years ago.

Summary features of Homo habilis (adapted by Kenneth A. Koenigshofer, PhD, Chaffey College, from *Explorations: An Open Invitation to Biological Anthropology;* https://explorations.americananthro.....php/chapters/; Chapter 10, "Early Members of the

Genus Homo" by Bonnie Yoshida-Levine Ph.D.; Beth Shook, Katie Nelson, Kelsie Aguilera, and Lara Braff, Eds.; licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, except where otherwise noted).

Hominin: Homo habilis
Dates 2.5 million years ago to 1.7 million years ago
Region(s): East and South Africa
Famous Discoveries: Olduvai Gorge, Tanzania; Koobi Fora, Kenya; Sterkfontein, South Africa
Brain Size: 650 cc average (range from 510 cc to 775 cc)
Dentition: Smaller teeth with thinner enamel compared to Australopithecus; parabolic dental arcade shape
Cranial Features: Rounder cranium and less facial prognathism than Australopithecus
Postcranial Features: Small stature; similar body plan to Australopithecus
Culture: Oldowan tools; oldest of known stone tool type; scrapers, choppers; indication of changing cognitive abilities (see Module 3.8)

Stone tools almost certainly predated Homo habilis (possibly by Australopithecus garhi or the species responsible for the tools from Kenya dating to 3.7 million years ago). However, stone tools become more frequent at sites dating to about 2 million years ago, the time of Homo habilis (Roche, Blumenschine, and Shea 2009). This suggests that these hominins were increasingly reliant on stone tools to make a living.

Stone tools are assigned a good deal of importance in the study of human origins. **Studying the form of the tools, the raw materials selected, and how they were made and used can provide insight into the thought processes of early humans** and how they modified their environment in order to survive (see Supplementary Content, Chapter 18, Material Culture). Paleoanthropologists have traditionally classified collections of stone tools into industries, based on their form and mode of manufacture. There is not an exact correspondence between a tool industry and a hominin species; however, some general associations can be made between tool industries and particular hominins, locations, and time periods. The names for the four primary tool industries in human evolution (from oldest to most recent) are the Oldowan, Acheulean, Mousterian, and Upper Paleolithic. The oldest stone tool industry is the Oldowan, named after the site of Olduvai Gorge where the tools were first discovered. The time period of the Oldowan is generally considered to last from about 2.5 mya to 1.6 mya. The tools of this industry are described as “flake and chopper” tools—the choppers consisting of stone cobbles with a few flakes struck off them. To a casual observer, these tools might not look much different from randomly broken rocks. However, they are harder to make than their crude appearance suggests. The rock selected as the core must be struck by the rock serving as a hammerstone at just the right angle so that one or more flat flakes are removed. This requires selecting rocks that will fracture predictably instead of chunking, as well as the ability to plan ahead and envision the steps needed to create the finished product. The process leaves both the core and the flakes with sharp cutting edges that can be used for a variety of purposes.

What were the hominins doing with the tools? One key activity seems to have been butchering animals. Animal bones with cutmarks start appearing at sites with Oldowan tools. Studies of animal bones at the site show leg bones are often cracked open, suggesting that they were extracting the marrow from the bone cavities. It is interesting to consider whether the hominins hunted these animals or acquired them through other means. The butchered bones come from a variety of African mammals, ranging from small antelope to animals as big as wildebeest and elephants! It is difficult to envision slow, small-bodied Homo habilis with their Oldowan tools bringing down such large animals. One possibility is that the hominins were scavenging carcasses from lions and other large cats. Regardless of how they were acquiring the meat, all these activities suggest an important dietary shift from the way that the australopithecines were eating. The Oldowan toolmakers were exploiting a new ecological niche that provided them with more protein and calories. Overall, increasing use of stone tools allowed hominins to expand their ecological niche and exert more control over their environment. As we’ll see shortly, this pattern continued and became more pronounced with Homo erectus.

3.3.12.1 Discovery of *Homo habilis*

Homo habilis was first discovered by Louis and Mary Leakey at Olduvai Gorge, Tanzania in 1960. Associated with stone tools (Oldowan), the Leakeys named their discovery “handy man.” *H. habilis* fossils have been found in Tanzania, Kenya, Ethiopia, and South Africa, although there is some debate as to whether the South Africa specimens should be included in the species. Some researchers contend that there was another early *Homo* species, *Homo rudolfensis*, which dates back to 2.4-2.5 mya. The *H. rudolfensis* fossils are slightly larger than those of *H. habilis*, leading some researchers to suggest the *H. habilis* exhibited sexual dimorphism and what we are seeing are male and female specimens of *H. habilis*. Others claim the size differences are significant enough to warrant the two species designations (O’Neil 1999-2012). In 2013 a *Homo* mandible was discovered in the Ledi-Geraru research area, Afar, Ethiopia. Dated to 2.8 to 2.75mya, the mandible exhibits an *Australopithecus*-like chin and *Homo*-like teeth (Villmoare et al. 2015). While still early in the research process, this discovery and further research may push back the date of the origin of *Homo* and help to resolve the debate between the *H. rudolfensis* and *H. habilis* fossils. For our purpose, we will consider them all *H. habilis*, making the approximate date range for this hominin 2.5 to 1.4mya or less conservatively to 1.7mya.



Figure **3.3.19**: *Homo habilis.*

Morphologically, *H. habilis* has a larger brain than the australopiths, about 35% larger (O’Neil c1999-2012). You will recall from the section on trends in human evolution that it is speculated that the brain also began to rewire at this point. *H. erectus* exhibits less prognathism (protrusion of the lower jaw) and platycephaly (flattening of the back of the head) than early hominins. The brow ridge is also smaller. All of these traits together make the face smaller than the australopiths. Postcranially, *H. habilis* exhibit a mix of primitive and derived traits. Primitive traits connecting it to an australopith ancestor are the longer forearms and the size of the finger bones along with how the tendons attach to the wrist bones. The tips of the finger bones are broad like humans. Smaller

teeth, a dental arcade in the shape of a parabolic arch, foot morphology, and a more rounded skull complete the human-like traits. Microanalysis of tooth wear indicates that *H. habilis* was omnivorous.

Table 3.3.2

Key Homo habilis fossil locations and the corresponding fossils and dates (adapted by Kenneth A. Koenigshofer, PhD, Chaffey College, from *Explorations: An Open Invitation to Biological Anthropology;* https://explorations.americananthro.....php/chapters/; Chapter 10, "Early Members of the Genus Homo" by Bonnie Yoshida-Levine Ph.D.; Beth Shook, Katie Nelson, Kelsie Aguilera, and Lara Braff, Eds.; licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, except where otherwise noted).

Location of Fossils Dates (mya = millions of years ago) Description
Ledi-Gararu, Ethiopia 2.8 mya Partial lower jaw with evidence of both Australopithecus and Homo traits; tentatively considered oldest Early Homo fossil evidence.
Olduvai Gorge, Tanzania 1.7 mya to 1.8 mya. Several different specimens classified as Homo habilis, including the type specimen found by Leakey, a relatively complete foot, and a skull with a cranial capacity of about 600 cc.
Koobi Fora, Lake Turkana Basin, Kenya 1.9 mya. Several fossils from Lake Turkana basin show considerable size differences, leading some experts to classify the larger specimen as a separate species, *Homo rudolfensis*.
Sterkfontein and other possible about 1.7 mya South African caves have yielded fragmentary remains identified as Homo habilis, but South African cave sites secure dates and specifics about the fossils are lacking.

3.3.13 *Homo erectus*

About 1.9 mya, a new species of Homo appeared. Known as *Homo erectus*, the prevailing scientific view was that this species was much more like us. These hominins were equipped with bigger brains and large bodies with limb proportions similar to our own. Perhaps most importantly, their way of life is now one that is recognizably human, with more advanced tools, hunting, use of fire, and colonizing new environments outside of Africa.

Compared to Homo habilis, Homo erectus had a larger brain size (average of about 900 cc compared to 650 cc to 750 cc for habilis). Instead of having a rounded shape like our skulls have, the erectus skull was long and low like a football, with a receding forehead, and a horizontal ridge called an occipital torus that gave the back of the skull a squared-off appearance. The cranial bones are thicker than those of modern humans, and some Homo erectus skulls have a slight thickening along the sagittal suture called a sagittal keel. Large, shelf-like brow ridges hang over the eyes. As noted above, the climate was increasingly arid and the forest canopy in parts of Africa was being replaced with a more open grassland environment, resulting in increased sun exposure for our ancestors. Compared to the earlier australopithecines, members of the genus Homo were also developing larger bodies and brains, starting to obtain meat by hunting or scavenging carcasses, and crafting sophisticated stone tools. For purposes of cooling, H. erectus may have had little body hair accompanied by darkened skin to protect against sun exposure. It is generally agreed that Homo erectus was the first hominin to migrate out of Africa and colonize Asia and later Europe (although recent discoveries in Asia may challenge this view).

3.3.13.1 Homo erectus Fossils

Based on current fossil data, *Homo erectus* existed between 1.9 million (mya) to 25 thousand years ago (Jurmain 2013). *H. erectus*, literally “upright human,” fossils have been found in Java (Indonesia), Africa, China, Europe, and Israel. Based on morphological differences in the cranium, some scientists identify two species, *H. erectus* in Asia and *H. ergaster* in Africa with the African specimens being smaller than the Asian; however, we will use the *H. erectus* designation for both.



Figure **3.3.20**: *Homo erectus*, Turkana Boy.

In 1984, a nearly complete *H. erectus* skeleton was found along a river in northern Kenya. Potassium-argon dating places “Turkana Boy” between 1.64 and 1.33 million years ago (mya). Aging and sexing of the fossil remains indicate that the individual was a male about age eight. He stood about 5’3” tall. Recent studies indicate that Turkana Boy followed the growth pattern of apes, so would have been near his adult height at the time of his death (Jurmain et al. 2013).

Homo erectus has a long history in Indonesia; fossils from Java were dated by argon dating to about 1.6 million to 1.8 million years. *H. erectus* fossils from the site of Ngandong in Java have yielded very recent dates of 43,000 years, although a more recent study with different dating methods concluded that they were much older—between 140,000 and 500,000 years old.

The pattern of increased brain size continued with *H. erectus*; its brain is up to 50% larger than its predecessor, *H. habilis* (O’Neil c1999-2012). This large brain was supported by a diet heavy in meat and other proteins. Its distinguishing characteristics include its sagittal keeling (a thickening of bone that runs from front to back on top of the skull; likely for muscle attachment) massive brow ridges (supraorbital tori) and bony prominences on the back of the skull. Postcranially, *H. erectus*’ bones are thicker than *H. habilis*, as were its jaws and face bones, and the proportion of arms to legs is like that of modern humans, causing some to suggest that its bipedal gait was like ours. The length of its leg bones indicates that *H. erectus* would have been an efficient long-distance runner, allowing them to “run down small and even medium size game animals on the tropical savannas of East Africa” (O’Neil c1999-2012). If so, as mentioned above, it is likely that *H. erectus* had much less body hair than its predecessors, as they would have needed to be able to sweat efficiently to cool the body. It is possible that *H. erectus* had little body hair (NOVA 2011). During the Pleistocene, the climate was increasingly arid and the forest canopy in parts of Africa was being replaced with a more open grassland environment, resulting in increased sun exposure for our ancestors.

Compared to the earlier australopithecines, members of the genus *Homo* were also developing larger bodies and brains, starting to obtain meat by hunting or scavenging carcasses, and crafting sophisticated stone tools. According to Nina Jablonski, an expert on the evolution of human skin, the loss of body hair and increased sweating capacity are part of the package of traits characterizing the genus *Homo*. While larger brains and long-legged bodies made it possible for humans to cover long distances while foraging, this new body form had to cool itself effectively to handle a more active lifestyle. Preventing the brain from overheating was especially critical. The ability to keep cool may have also enabled hominins to forage during the hottest part of the day, giving them an advantage over savanna predators, like lions, that typically rest during the heat of the day.

As noted above, scientists generally agree that *H. erectus* was the first hominin to leave Africa. As mentioned previously, fossils have been found in Africa, northern China, Indonesia, Europe, and Israel. In the Republic of Georgia, fossils were found in strata dated to 1.7mya suggesting that *H. erectus* left Africa soon after it evolved. A recent report (Dembo et al. 2015) posits that *H. habilis* was the first hominin to leave Africa, not *H. erectus*. Should this contention be supported with more data, it can still be argued that *H. erectus* was quite successful in colonizing the Old World (Africa, Europe, and Asia), helped, no doubt, by its advanced cultural behaviors.

Now we can address the question of why *Homo erectus* traveled such vast distances to these far-flung regions. To do this, we have to consider what we have learned about the biology, culture, and environmental circumstances of *Homo erectus.* The larger brain and body size of *Homo erectus* were fueled by a diet consisting of more meat, and longer more powerful legs made it possible to walk and run longer distances to acquire food. Since they were eating higher on the food chain, it was necessary for them to extend their home range to find sufficient game. Cultural developments including better stone tools and new technology such as fire gave them greater flexibility in adapting to different environments. Finally, the major Pleistocene climate shift discussed earlier in the chapter certainly played a role. Changes in air temperature, precipitation, access to water sources, and other habitat alteration had far-reaching effects on animal and plant communities; this included *Homo erectus*. If hominins were relying more on hunting, the migration patterns of their prey could have led them increasingly long distances.

There is evidence of *Homo erectus* in China from several regions and time periods. *Homo erectus* fossils from northern China, collectively known as “Peking Man,” are some of the most famous human fossils in the world, dated to about 400,000–700,000 years ago. The discovery of stone tools from China dating to 2.1 million years, older than any *Homo erectus* fossils anywhere, opens up the intriguing possibility that hominins earlier than *Homo erectus* could have migrated out of Africa.

At this time, researchers aren’t in agreement as to whether the first Europeans belonged to *Homo erectus* proper or to a later descendent species.

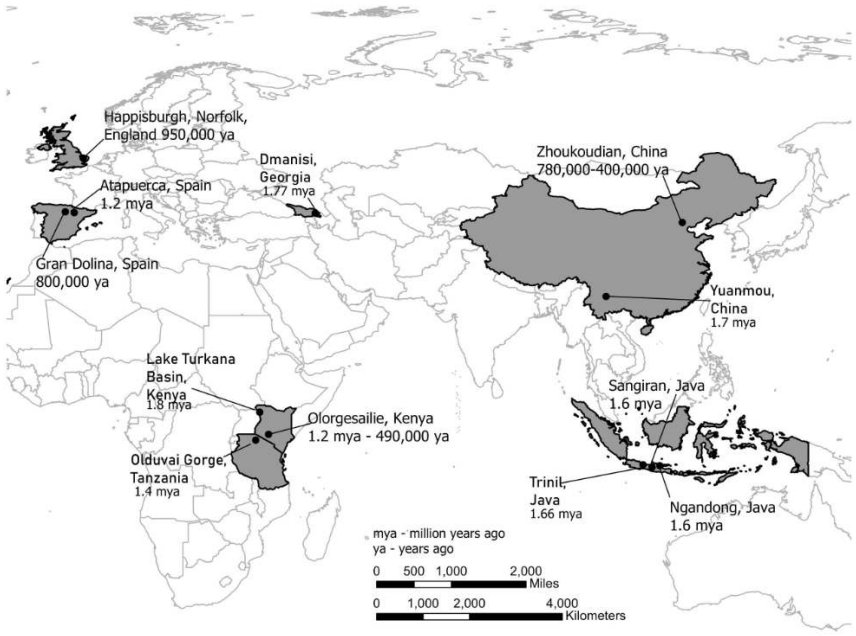


Figure **3.3.21**: Map showing the locations of Homo erectus fossils around Africa and Eurasia (adapted by Kenneth A. Koenigshofer, PhD, Chaffey College, from *Explorations: An Open Invitation to Biological Anthropology;* https://explorations.americananthro.....php/chapters/; Chapter 10, "Early Members of the Genus Homo" by Bonnie Yoshida-Levine Ph.D.; Beth Shook, Katie Nelson, Kelsie Aguilera, and Lara Braff, Eds.; licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, except where otherwise noted).

3.3.13.2 Homo erectus Culture

Homo erectus shows significant cultural innovations in diet, technology, life history, environments occupied, and perhaps even social organization, some that you will probably recognize as more “human-like” than any of the hominins previously covered.

About 1.5 million years ago, some Homo erectus populations began making different forms of tools (known as "material culture;" see Section 3.14). These tools–classified together as constituting the Acheulean tool industry–are more complex in form and more consistent in their manufacture. Unlike the Oldowan tools, which were cobbles modified by striking off a few flakes, Acheulean toolmakers carefully shaped both sides of the tool. This type of technique, known as bifacial flaking, requires more planning and skill on the part of the toolmaker; he or she would need to be aware of principles of symmetry when crafting the tool. One of the most common tool forms of this period was the handaxe. Besides handaxes, forms such as scrapers, cleavers, and flake tools are present at Homo erectus sites. One striking aspect of Acheulean tools is their uniformity. They are more standardized in form and mode of manufacture than the earlier Oldowan tools. For example, the aforementioned handaxes vary in size, but they are remarkably consistent in regard to their shape and proportions. They were also an incredibly stable tool form over time—lasting well over a million years with little change.

Recently, newer methods—including microscopic analysis of burned rock and bone—have revealed clear evidence of fire use at Koobi Fora, Kenya, dating to 1.5 million years ago (Hlubik et al. 2017).

There is general consensus that *H. erectus* evolved from *H. habilis* and *Homo heidelbergensis* evolved from *H. erectus* in Africa, eventually supplanting *H. erectus* populations in the Old World--Africa, Asia, and Europe (Figure **3.3.18**).

3.3.13.3 Tool Use and Cognitive Abilities of Homo erectus

What (if anything) do the Acheulean tools tell us about the mind of Homo erectus? Clearly, they took a fair amount of skill to manufacture. Apart from the actual shaping of the tool, other decisions made by toolmakers can reveal their use of foresight and planning. Did they just pick the most convenient rocks to make their tools, or did they search out a particular raw material that would be ideal for a particular tool? Analysis of Acheulean stone tools suggest that at some sites, the toolmakers selected their raw materials carefully—traveling to particular rock outcrops to quarry stones and perhaps even removing large slabs of rock at the quarries to get at the most desirable material. Such complex activities would require advanced planning. They also likely required cooperation and communication with other individuals, as such actions would be difficult to carry out solo. However, other Homo erectus sites lack evidence of such selectivity; instead of traveling even a short distance for better raw material, the hominins tended to use what was available in their immediate area (Shipton et al. 2018). In contrast to Homo erectus tools, the tools of early modern Homo sapiens during the Upper Paleolithic display tremendous diversity across regions and time periods. Additionally, Upper Paleolithic tools and artifacts communicate information such as status and group membership. Such innovation and social signaling seem to have been absent in Homo erectus, suggesting that they had a different relationship with their tools than did Homo sapiens (Coolidge and Wynn 2017). Some scientists assert that these contrasts in tool form and manufacture may signify key cognitive differences between the species, such as the ability to use a complex language.

Table 3.3.3. Characteristics of Homo erectus. (adapted by Kenneth A. Koenigshofer, PhD, Chaffey College, from *Explorations: An Open Invitation to Biological Anthropology;* https://explorations.americananthro.....php/chapters/; Chapter 10, "Early Members

of the Genus Homo" by Bonnie Yoshida-Levine Ph.D.; Beth Shook, Katie Nelson, Kelsie Aguilera, and Lara Braff, Eds.; licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, except where otherwise noted).

Hominin	*Homo erectus*
Dates	1.8 million years ago to about 200,000 years ago
Region(s)	East and South Africa; West Eurasia; China and Southeast Asia
Famous Discoveries	Lake Turkana, Olorgesailie, Kenya; Zhoukoudian, China; Dmanisi, Republic of Georgia
Brain Size	Average 900 cc; range between 650 cc and 1,100 cc
Dentition	Smaller teeth than *Homo habilis*
Cranial Features	Long, low skull with robust features including thick cranial vault bones and large brow ridge, sagittal keel, and occipital torus
Postcranial Features	Larger body size compared to *Homo habilis*; body proportions (longer legs and shorter arms) similar to *Homo sapiens*
Culture	Acheulean tools (in Africa) (see Module 3.8); evidence of increased hunting and meat-eating; use of fire; migration out of Africa

3.3.14 THE BIG PICTURE OF EARLY *HOMO*

We are discovering that the evolution of the genus *Homo* is more complex than what was previously thought. The earlier prevailing view of a simple progression from *Australopithecus* to *Homo habilis* to *Homo erectus* as clearly delineated stages in human evolution just doesn't hold up anymore.

3.3.14.1 Variability in the Fossil Record of Early Homo

As is apparent from the information presented here, there is tremendous variability during this time. While fossils classified as *Homo habilis* show many of the characteristics of the genus *Homo*, such as brain expansion and smaller tooth size, the small body size and long arms are more akin to australopithecines. There is also tremendous variability within the fossils assigned to *Homo habilis*, so there is no consensus on whether it is a single or multiple species of *Homo*, a member of the genus *Australopithecus*, or even a yet-to-be-defined new genus.

What does this diversity mean for how we should view early *Homo*? First, there isn't an abrupt break between *Australopithecus* and *Homo habilis* or even between *Homo habilis* and *Homo erectus*. Characteristics we define as *Homo* don't appear as a unified package; they appear in the fossil record at different times. This is known as mosaic evolution.

We can consider several explanations for the diversity we see within early *Homo* from about 2.5 million to 1.5 million years ago. One possibility is the existence of multiple contemporaneous species of early *Homo* during this period. In light of the pattern of environmental instability discussed earlier, it shouldn't be surprising to see fossils from different parts of Africa and Eurasia display tremendous variability. Multiple hominin forms could also evolve in the same region, as they diversified in order to occupy different ecological niches. However, even the presence of multiple species of hominin does not preclude their interacting and interbreeding with one another.

Diversity of brain and body sizes could also reflect developmental plasticity—short-term adaptations within a lifetime (Anton, Potts, and Aiello, 2014). These have the advantage of being more flexible than genetic natural selection, which could only occur over many generations. For example, among human populations today, different body sizes are thought to be adaptations to different climate or nutritional environments.

3.3.14.2 **Trends in the Behavior of Early** Homo

New discoveries are also questioning old assumptions about the behavior of *Homo habilis* and *Homo erectus*. Just as the fossil evidence doesn't neatly separate *Australopithecus* and *Homo*, evidence of the lifeways of early *Homo* show similar diversity. For example, one of the traditional dividing lines between *Homo* and *Australopithecus* was thought to be stone tools: *Homo* made them; *Australopithecus* didn't. However, the recent discovery of stone tools from Kenya dating to 3.3 million years ago challenges this point of view. Similarly, the belief that *Homo erectus* was the first species to settle outside Africa may now come into question with the report of 2.1 million-year-old stone tools from China. If this find is supported by additional evidence, it may cause a reevaluation of *Homo erectus* being first to leave. Instead, there could have been multiple earlier migrations of hominins such as *Homo habilis* or even *Australopithecus* species.

Rather than obvious demarcations between species and their corresponding behavioral advancements, it now looks like many behaviors were shared among species. Despite the haziness dominating the early *Homo* narrative, we can identify some overall trends for the million-year period associated with early *Homo.* These trends include brain expansion, a reduction in facial prognathism, smaller jaw and tooth size, larger body size, and evidence of full terrestrial bipedalism. These traits are associated with a key behavioral shift that emphasizes culture as a flexible strategy to adapt to unpredictable environmental circumstances. Included in this repertoire are the creation and use of stone tools to process meat obtained by scavenging and later hunting, a utilization of fire and cooking, and the roots of the human life history pattern of prolonged childhood, cooperation in child raising, and the practice of skilled foraging techniques. In fact, it's apparent that the cultural innovations are driving the biological changes, and vice versa, fueling a feedback loop that will continue during the later stages of human evolution.

3.3.15 *Homo heidelbergensis* : the common ancestor of *Homo neanderthalensis* in Europe and *Homo sapiens* in Africa.

Some publications, e.g. Larsen 2014, refer to *H. heidelbergensis* as archaic *Homo sapiens*, but for our purposes, we will use the *H. heidelbergensis* designation. Otto Schoetensack found the first *Homo heidelbergensis* fossils in 1907 in Mauer, Germany. Since then *H. heidelbergensis* fossils have been found in Africa, Europe, and Asia. The date range for the species is 800 kya (thousand years ago) to 350 kya. Primitive traits (traits like the ancestors) include its large supraorbital tori (brow ridge), sagittal keeling, and low frontal bone. Derived traits (traits different from the ancestor) include separate supraorbital tori over each eye orbit, a more vertical posterior cranial vault, wide parietal bones in relation to the cranial base, and a larger cranial capacity than *H. erectus* (Becoming Human c2008). Additionally, they exhibit sexual dimorphism similar to that of modern humans.



Figure **3.3.22**: *Homo heidelbergensis* skull on mirrored steel stand showing reflection of palate and dentition in upper jaw.

There is regional variation in the morphology of *H. heidelbergensis*. European specimens found at Atapuerca, (Spain), Petralona (Greece), Seinheim (Germany), and Swancombe (England) show that they had compact bodies, which could have been a response to living in the cold climates of the north as it would help to conserve heat. Additionally, the cranium is a mosaic of *H. erectus* traits

and derived traits. In Asia, data from sites such as Zhoukoudian, Jinniushan, and Dali (China) show a mix of *H. erectus* and *H. sapiens* traits; the latter includes large cranial capacity and thin braincase walls. African specimens from Kabwe (Zambia), Florisbad (South Africa), Laetoli (Tanzania), and Bodo (Ethiopia) also show a combination of *H. erectus* and *H. sapiens* traits. It shares the massive supraorbital tori and prominent occipital torus with *H. erectus* and thin cranial vault bones, less angulated occipital, and cranial base with *H. sapiens*.

Significantly, *H. heidelbergensis* is the common ancestor of *Homo neanderthalensis* in Europe and *Homo sapiens* in Africa.

3.3.16 *Homo neanderthalensis*

Numerous Neanderthal fossils have been recovered since its discovery in 1856 in the Neander Valley, Germany. Neanderthals have been the speculation of scientists and the general public ever since. Some anthropologists classify Neanderthals as a subspecies of *Homo sapiens*, *Homo sapiens neanderthalensis*, while others interpret the morphological differences as significant enough to warrant classifying them as a different species, *Homo neanderthalensis*. Here, we will use the latter designation.



Figure **3.3.23**: *Homo neanderthalensis.*

True Neanderthals first appear in the fossil record about 200,000 years ago (200 kya), with fossils exhibiting Neanderthal-like characteristics appearing as early as 400 kya. Recent research indicates that Neanderthals went extinct only about 40,000 years ago, between 41 kya and 39 kya (Higham et al., 2014). Molecular research denotes that some Neanderthal DNA lives on in modern humans, approximately 2% of the DNA of “people who descend from Europeans, Asians, and other non-Africans is Neanderthal” (Callaway, 2014). The Neanderthal genes are involved in fighting infections, dealing with ultraviolet radiation (Callaway, 2014), and living at high altitudes (Callaway, 2015). Neanderthal DNA has recently been linked with depression, obesity, and certain skin disorders, e.g., lesions caused by sun exposure (Callaway, 2015).

Neanderthals are the only hominin to originally evolve in a glacial environment, leading to some characteristics adapted for the cold climate. Many scientists contend that the midface prognathism (protruding midface) allowed for enlarged sinuses that functioned to warm and add moisture to the cold, dry air before entering the lungs. Small holes below the eye orbits, called the infraorbital foramina, are larger in the European Neanderthals that modern humans, suggesting that the blood vessels were larger, which allowed for more blood flow to the face. This would have helped keep the face warmer. They are relatively short and stocky, males averaging 5’ 5” and females 5’1”, with shorter appendages, both of which would help to conserve heat by providing less surface area from which to radiate heat. Their leg bones are thick and dense, suggesting that they frequently walked and ran, most likely in food procurement activities. Some postulate that some Neanderthals had pale skin that would have helped increase Vitamin D synthesis by increasing the amount of UV radiation to be absorbed by the body (O’Neil, c1999-2014). The width of their body trunks and short tibias fit the predictions for cold weather adaptation as proposed by Christopher Ruff (Larsen, 2013). The brain size of Neanderthals may also be related to cold weather. Averaging a cranial capacity between 1300 to 1400 cm3, they have the largest brains of all hominins, including our own species, Homo sapiens. The size may be associated with increased metabolic efficiency in cold weather, which is similar to modern Inuit peoples today who have a larger brain size than other human populations (Jurmain et al., 2013).

The presence of an occipital bun (a prominent bulge at the lower back of the skull) is one of the characteristics used to identify *H. neanderthalensis* specimens, although it should be noted that this characteristics persists in a small percentage of modern human populations. The occipital bun may have evolved to counterbalance Neanderthal’s heavy face when running; it prevents the head from making huge horizontal accelerations (NOVA 2002). The occipital bun also makes the skull elongated in comparison to most *Homo sapiens*. They had heavy brow ridges like those of *H. heidelbergensis*. Neanderthal skulls recovered from Amud and Tabun

in Israel exhibit more *H. sapiens*-like cranial characteristics, including lack of an occipital bun, smaller eye orbits, tall and wide nasal openings, and smaller teeth (Larsen 2014).

3.3.17 Denisovans

The Denisovans were archaic humans closely related to Neanderthals, whose populations overlapped geographically with the ancestors of modern-day humans who existed at the same time. In 2010, scientists announced the discovery and DNA analysis of a finger bone and two teeth found in Denisova Cave, Siberia (Reich et al., 2010). The artifacts were recovered from a deposit dated to 50 kya to 30 kya. Data suggest that the remains were from an individual who shared a common origin with Neanderthals, but was not a Neanderthal, nor was it a modern human. The Denisovan individual(s) share 4-6% of its genetic material with modern peoples living in New Guinea, Bougainville Islands, and China. Further studies (Cooper and Stringer, 2013) indicate that the Denisovans crossed Eurasia and interbred with modern humans, but genetically the Denisovans were more closely related to Neanderthals than modern humans (Meyer et al., 2012). While further finds will shed more light on the Denisovans, it is clear that there was more genetic variability during the Pleistocene than previously thought (Larsen, 2014), but that gene flow between the various populations had implications for the emergence of modern humans (Pääbo, 2015).

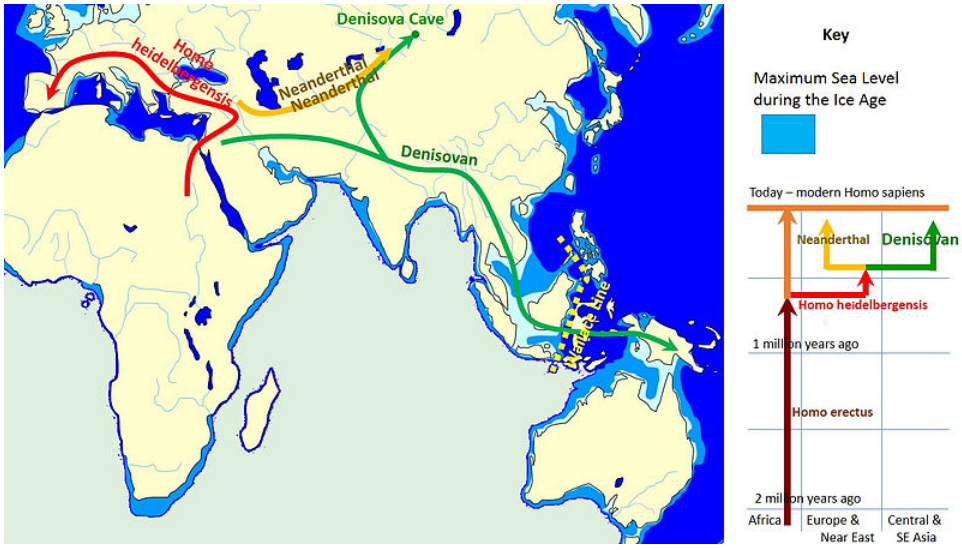


Figure **3.3.24**: Spread and Evolution of Denisovans.

3.3.18 *Homo floresiensis*

In 2003, a team of scientists discovered unusual fossils in Liang Bua cave, Flores Island, Indonesia. The small stature of the individual led to the naming of a new species, *Homo floresiensis*. Since then twelve individuals have been found ranging in time from 74 kya to 17 kya. *H. floresiensis* had a small brain and stood only about 3.5' tall as an adult. They had receding foreheads, no chins, shrugged-forward shoulders, and large feet relative to their short legs. They share some characteristics with *H. sapiens*, including smaller dentition, separate brow ridges, and a non-prognathic face (flattened face).



Figure **3.3.25**: *Homo floresiensis.*

Several hypotheses have been proposed to explain the appearance of *H. floresiensis*. One suggests that we are seeing island dwarfism, which is an evolutionary process resulting from long-term isolation on a small island. Limited resources and lack of predators selects for smaller bodied individuals who need fewer resources then large-bodied individuals. Another hypothesis claims that *H. floresiensis*is is not a separate species, but *H. sapiens* exhibiting microcephaly or some other developmental deficiency such as hyperthyroid cretinism (Oxnard et al. 2010) because the "…cranial features…are within the modern range of variation seen in living populations from the larger region [of Indonesia]" (Larsen 2014: 409), lending support to the second hypothesis. How *H. floresiensis* fits into the evolutionary picture is unclear based on the current data.

3.3.19 *Homo sapiens*

At present, only one hominin species inhabits planet Earth, *Homo sapiens*. All other *Homo* species became extinct by approximately 30,000 to 40,000 years ago. Many scientists believe that these extinctions were driven by competition from *Homo sapiens*. Morphologically, *H. sapiens* characteristics include the presence of a chin, a large brain, flat face, rounded or globular cranium, and a continuous, reduced brow ridge. Its bones are gracile (slender) in comparison to earlier hominins, although the earliest *H. sapiens* are more robust than modern populations and show none of the cold weather adaptations of Neanderthals. The more gracile nature lends credence to the hypothesis that modern humans evolved first in Africa. The leaner body proportions are more adaptive to the tropical African environments as there would be more body surface area to radiate heat. DNA evidence also supports an African origin. Modern African populations have more genetic diversity than any other modern human population, inferring that they have been evolving longer (Becoming Human c2008).



Figure **3.3.26**: *Homo sapiens.*

The oldest *Homo sapiens* fossils were found in Africa at Omo Kibish, Ethiopia, dated to almost 200,000 years ago (195 kya; kya = 1,000 years ago). *Homo sapiens* fossils found at Herto, Ethiopia (160 kya–154 kya), Klasies River Mouth and Border Cave, South Africa (120 kya – 89 kya), Skhūl, Israel (130 kya–100 kya), and Qafzeh, Israel (120 kya–92 kya), add further support to the age of *Homo sapiens*. Sites in Asia, e.g., in China and Borneo, Indonesia date to 40 kya or younger and in Europe date no earlier then 31 kya (Mladeč, Czech Republic). Modern humans somehow made it to Australia over open ocean by 55 kya; however, no human remains have been found on the Australian continent dating earlier than 30 kya at Lake Mungo. This data lends little support for the Multiregional Hypothesis, a model that proposed that modern *H. sapiens* evolved in separate places in the Old World from local archaic *H. sapiens* populations. This model suggested that gene flow between the populations led to the genetic similarity of modern humans. Most anthropologists support the model of an African origin. There are two variants of this model. The Out of Africa model claims that modern humans arose in Africa and spread to Europe after 50 kya, replacing non-*H. sapiens* population with no gene flow. An alternative is the Assimilation model that also claimed modern human first arose in Africa and spread to Europe and Asia. The primary difference between the Out of Africa and Assimilation models is that the latter does claim gene flow between *H. sapiens* and *H. neanderthalensis*. Taking into account the recent DNA analyses of Neanderthals discussed above, multiple lines of evidence support the Assimilation model.

The case of the disappearing Neanderthals

We know that some Neanderthal genes are in the genome of modern humans, so in essence, some part of the Neanderthals survives today. However, there are no Neanderthals walking around in the modern world. So, what happened to the Neanderthals? Scientists suggest that *H. sapiens* out-competed Neanderthals with their more diverse diet and "sophisticated and cognitive abilities" (Becoming Human c2008), which may have included superior ability to form visual imagery in imagination to forecast outcomes of future actions, thereby enhancing planning and inventiveness (Koenigshofer, 2017). These traits allowed *H. sapiens* to readily adapt to rapidly changing climatic conditions during the Upper Paleolithic.

A recent hypothesis suggests that Neanderthals died out during a period of powerful volcanic activity in western Eurasia. Excavations at Mezmaiskaya Cave in the Caucasus Mountains in southern Russia have recovered a plethora of Neanderthal bones, stone tools, and remains of prey animals. The Neanderthal remains and artifacts appear in strata above a layer of volcanic ash and below a second layer of volcanic ash. No Neanderthal bones or artifacts have been found above the second level, suggesting that Neanderthals no longer occupied the area. Both of the ash layers contain levels of pollen associated with cooler, drier climates. "The ash layers correspond chronologically to what is known as the Campanian Ignimbrite super-eruption, which occurred around 40,000 years ago in modern day Italy, and a smaller eruption thought to have occurred around the same time in the Caucasus Mountains" (University of Chicago Press Journals 2010). The ensuing volcanic winter caused a dramatic climate shift that led to the demise of the Neanderthals. This data correlates with Higham et al.'s (2014) research that claims Neanderthals went extinct between 41,000 and 39,000 years ago.

3.3.20 Brain Evolution

Many scientists (Cajal, 1937; Crick & Koch, 1990; Edelman, 2004) believe that the human brain is the most complex machine in the known universe. Its complexity points to one undeniable fact—that it has evolved slowly over time from simpler forms. Evolution of the brain is intriguing not only because we can marvel at this complicated biological structure, but also because it reveals a long history of many less complex nervous systems (Figure 3.3.25), suggesting a record of adaptive behaviors in life forms other than humans. Thus, evolutionary study of the nervous system is important, and it is the first step in understanding its design, its workings, and its functional interface with the environment.

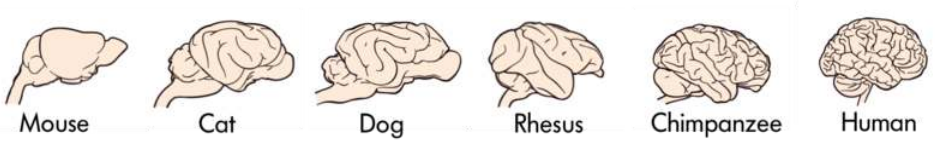


Figure **3.3.27**: The brains of various mammals (not to scale). The brains shown are artificially depicted as approximately equal in size. Though the vast differences in the actual size of these brains is not shown, it is still possible to see the increasing degree of folding of the cerebral cortex as indicated by the greater number of fissures and gyri (valleys and hills on the surface of the cortex, indicated by the number of lines) on the surface of the chimpanzee and human brains compared to the others. Note that the mouse cortex is smooth, showing no folding at all indicating less corticalization--cortical development. The greatest folding of the cerebral cortex occurs in humans revealing a trend in mammalian and primate evolution toward increasing corticalization along with increasing brain size. For more on trends in brain evolution in mammals and primates, see section 18.13.

The brains of some animals, other mammals, like apes, monkeys, and rodents, are structurally similar to humans (Figure 3.3.27), while others are not (e.g., invertebrates such as the octopus). Does anatomical similarity of these brains suggest that behaviors that emerge in these species are also similar? Indeed, many animals display behaviors that are similar to humans, e.g., apes use nonverbal communication signals with their hands and arms that resemble nonverbal forms of communication in humans (Gardner & Gardner, 1969; Goodall, 1986; Knapp & Hall, 2009). If we study very simple physiological responses made by individual neurons (nerve cells), the neural responses of invertebrates such as the sea slug (Kandel & Schwartz, 1982) look very similar to the neural responses of single neurons in mammals, including humans, suggesting that from time immemorial such basic functions of individual nerve cells in the brains of many simple animal forms have been conserved over the course of evolution of more complex animals and that in fact these fundamental features of nerve cells are the foundation of more complex behaviors in animals that evolved later (Bullock, 1984).

Although basic physiological processes in nerve cells are very similar across species, nevertheless even at the micro-anatomical level, we note that individual neurons differ in complexity in some respects across animal species. Human neurons exhibit more intricate complexity than neurons in other animals; for example, neuronal processes (dendrites, which receive input from other

neurons) in humans have many more branch points, branches, and dendritic spines associated with connections between neurons (see Chapter 10 on Learning and Memory).

Complexity in the structure of the nervous system, both at the macro- and micro-levels, gives rise to complex behaviors. We can observe similar movements of the limbs, as in nonverbal communication, in apes and humans, but the variety and intricacy of nonverbal behaviors using hands in humans surpasses apes. This is related to the fact that humans have more neurons for the control of movement of the hands than apes do. Deaf individuals who use American Sign Language (ASL) express themselves in English nonverbally; they use this language with such fine gradation that many accents of ASL exist (Walker, 1987). Complexity of behavior with increasing complexity of the nervous system, especially the cerebral cortex, can be observed in the genus *Homo* (Figures 3.3.27 and 3.3.28). If we compare sophistication of material culture in **Homo habilis** (2 million years ago; brain volume ~650 cm3) and **Homo sapiens** (300,000 years to now; brain volume ~1400 cm3), the evidence shows that *Homo habilis* used crude stone tools in stark contrast with modern tools used by *Homo sapiens* to erect cities, develop written languages, create complex industries, and embark on space travel. All of this is due to increasing complexity of the brain.

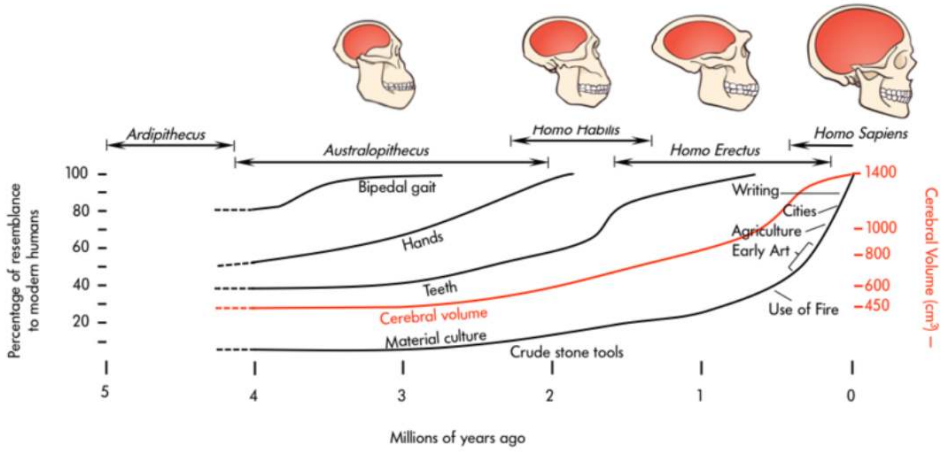


Figure **3.3.28**: Increasing brain volume and other anatomical and behavioral features including development of material culture (including stone tools) over hominin evolution. Note that both brain volume and cultural advances accelerate after first use of fire. Missing from this diagram is the 1,400 cubic centimeter brain of Neanderthals, equal or slightly larger than the brain of modern humans.

What led to the complexity of the brain and nervous system during evolution, and the accompanying refinement of behavior and cognition? Darwin (1859, 1871) proposed two forces: natural selection and sexual selection as work engines driving these changes. As noted previously, Darwin prophesied, “psychology will be based on a new foundation, that of the necessary acquirement of each mental power and capacity by gradation.” In other words, Darwin predicted that psychology will be based on evolution (Rosenzweig, Breedlove, & Leiman, 2002).

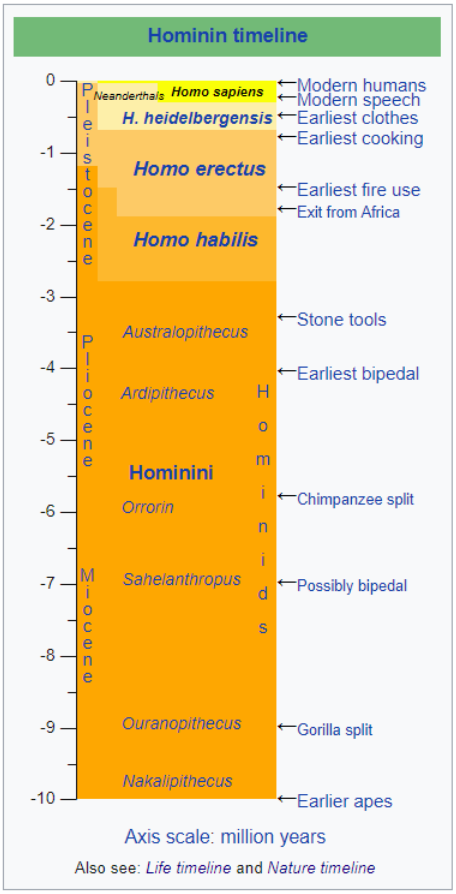


Figure **3.3.29**: Hominin timeline highlighting evolutionary splits (divergences) from gorilla and chimpanzee genetic lines, major behavioral/cultural innovations, and migration out of Africa. Not shown is migration of Homo sapiens into Europe about 45,000 years ago where they co-existed with Neanderthals for thousands of years (Image from Wikimedia Commons; File:HumanTimeline-TemplateImage-20181222.png; https://commons.wikimedia.org/wiki/F...e-20181222.png; by Drbogdan; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

3.3.21 Tool Use, Evolution of Intelligence, and the Brain

As the discussion above attests, tool construction and use is ancient in the genus *Homo* dating back millions of years at least to Australopithecus. Cachel (2012, p. 222) states, "Human tool behavior is species-specific. It remains a diagnostic feature of humans." According to Osiurak and Massen (2014, p.1107),"Tool use has also been often viewed as an important step during evolution (van Schaik et al., 1999) or even as a marker of the evolution of human intelligence." Vaesen (2012, p. 203) argues that human tool making and tool use provide evidence of "unique higher cognitive ability" in humans and mark "a major cognitive discontinuity" between humans and apes. According to Vaesen (p. 203), there are "nine cognitive capacities deemed crucial to tool use: enhanced hand-eye coordination, body schema plasticity, causal reasoning, function representation, executive control, social learning, teaching, social intelligence, and language. Since striking differences between humans and great apes stand firm in eight out of nine of these domains, [he] conclude[s] that human tool use still marks a major cognitive discontinuity between us and our closest relatives." He also "address[es] the evolution of human technologies. In particular, [he] show[s] how the cognitive traits reviewed help to explain why technological accumulation evolved so markedly in humans, and so modestly in apes." Technological accumulation refers to the accumulation of technological innovations over generations leading to ever increasing technological complexity. This is closely tied to the great capacity of humans for efficient cultural transmission of learned behaviors over successive generations, an ability that occurs in apes only to a relatively limited degree. Vaesen (p. 213) explains: "The complexity of human technologies is tightly linked to our remarkable ability for **cumulative culture** [bold added]: Humans have been able to build complex systems by accumulating modifications over successive generations, gradually improving on previous achievements." This requires "sophisticated mechanisms for social learning and active teaching," in which language

certainly played a part, but it is "likely that other cognitive capacities need to be added; for example, . . . causal thought is part and parcel of human social learning, and hence," is also needed for "high-fidelity cultural transmission" (p. 214). Vaesen further emphasizes this point: "causal thought (whenever it emerged) is a plausible explanans for why cumulative culture evolved so markedly in humans and so modestly in apes" (p.214). Cachel (2012, p. 222) goes further and states that experiments demonstrate the "great gulf that exists between human causal reasoning and the reasoning abilities of chimpanzees. In fact, chimpanzees do not seem to have a “theory of how the world works.” If this is so, then it epitomizes the unique character of human tool behavior – that it is based on the unconscious, sophisticated knowledge of energy, movement, objects, and the interaction of objects that Povinelli et al. (2000) label 'folk physics'.” Causal reasoning lowers individual learning costs, making learning by individuals more efficient, "by making individual learning reasoned rather than random" (Vaesen, 2012, p.215). Vaesen also argues that executive control involving inhibition of current drives and focus of attention contributes in a similar way.

Vaesen (2012) describes the work of other researchers who have identified brain areas involved in tool use. Vaesen (p.204) writes "Orban and colleagues (2006) identified a set of functional regions in the **dorsal intraparietal sulcus** (IPS) of the human brain that is involved in representations of the central visual field and in the extraction of three-dimensional form from motion. Crucially, these brain regions were not found in the brains of monkeys. The regions subserve, the authors conjectured, the enhanced visual analysis necessary for the precision with which humans manipulate tools. Second, Stout and Chaminade (2007) found that parts of these regions were indeed recruited when modern human subjects engaged in Oldowan-like tool making. Importantly, no increased activation was observed when the human subjects were asked just to strike cobbles together without intending to produce flakes. Human dorsal IPS, thus, may allow for better identification of suitable targets on the core, and as such, explain in part why humans outperform other primates in matters of tool use."

Additional insights regarding tool use and the brain come from study of brain-damaged individuals. Osiurak and Massen (2014, p. 1107) review results of studies of "patients with tool use disorders, also called apraxia of tool use. When asked to light a candle, for example, those patients can light the candle correctly but then put it to the mouth in an attempt to smoke it. Such observations have led traditional cognitive models of apraxia to assume that tool use is supported by sensorimotor knowledge about tool manipulation . . . This sensory-motor hypothesis assumes that manipulation knowledge stored within **inferior fronto-parietal** areas [of the cerebral cortex] is critical to tool use skills." This may be the case because "the **left inferior parietal lobe** might rather support technical reasoning, namely, the ability to reason about physical object properties. . . . the ability to understand mechanical actions might be the specificity of the **anterior** portions of the **inferior parietal lobe** (particularly the supramarginal gyrus) while the **posterior parietal cortex** might be involved in the planning of the grasping and reaching components of both tool-use and non-tool-use actions."

File:ParietCapts lateral.png

Figure **3.3.30**: The human right cerebral cortex highlighting parts of the parietal lobe. Folds on the surface of the cortex form hills (gyri, singular, gyrus) and valleys (fissures or sulci; singular, sulcus). The central sulcus divides the frontal lobe from the parietal lobe behind it. The parietal lobe, especially on the right side of the brain, processes information about space and spatial relations. Regions of the parietal lobe are involved in tool construction and tool use, important in human evolution. See text for details. (Image from Wikimedia Commons; File:ParietCapts lateral.png; https://commons.wikimedia.org/wiki/F...ts_lateral.png; author, Sebastian023; licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license. Caption by Kenneth A. Koenigshofer, PhD).

Tool construction and use, although critically important in human evolution as indicated by the discussion above, are just a single narrow aspect of the complex of abilities that make up human intelligence. In Chapter 14 on Intelligence and Cognition, we continue our discussion of this topic by examining the brain mechanisms which underlie other cognitive abilities that we collectively refer to as intelligence. In the next section of this chapter, David Buss from the University of Texas examines evolutionary theories in psychology.

3.3.22 References

Alemseged Z, Spoor F, Kimbel WH, Bobe R, Geraads D, Reed D, Wynn JG. (2006). A juvenile early hominin skeleton from Dikika, Ethiopoa. Nature 443 (Sep 21): 296-301. Available from: http://www.nature.com.offcampus.lib....ture05047.html. doi:10.1038/nature05047.

Anghelinu, M. (2013). Fitting the ladder to the tree: A common sense view on the cognitive evolution of the Pleistocene human lineage. *Arheologia Moldovei*, *36* (1), 7-24.

Bailey SE. (2006). Sahelanthropus tchadensis. In: Encyclopedia of anthropology, Vol. 5. Thousand Oaks (CA): SAGE Reference. p. 2044-2045.

Barras C. (2014). Human 'missing link' fossils may be a jumble of species. New Sci [Internet] [cited 2015 Aug 13]; 222(2964). Available from:https://www.newscientist.com/article/mg22229643-200-human-missing-link-fossils-may-be-jumble-of-species/

Becoming Human. c2008. The human lineage through time. Institute of Human Origins [Internet] [cited 2015 Aug 3]. Available from: http://www.becominghuman.org/node/human-lineage-through-time

Cachel, S. (2012). Human tool behavior is species-specific and remains unique. *Behavioral and Brain Sciences*, *35*(4), 222.

Callaway E. (2014). Modern human genomes reveal our inner Neanderthal. Nature [Internet] [cited 2015 Aug 21]. Available from: http://www.nature.com/news/modern-hu...erthal-1.14615

Callaway E. (2015). Neanderthals had outsize effect on human biology. Nature [Internet] [cited 2015 Aug 21]; 523 (7562): 512-513. Available from: http://www.nature.com.offcampus.lib....iology-1.18086. doi: 10.1038/523512a

Campbell BG. (2000). *Humankind Emerging*, 8th edition. Needham Heights (MA): Allyn&Bacon.

Cooper A, Stringer CB. (2013). Did the Denisovans cross Wallace's Line? Science [Internet] [cited 2015 Aug 21}; 342 (6156): 321-323. Available from: http://www.sciencemag.org.offcampus....d-963a99b700ea. doi: 10.1126/science.1244869.

Cosmides, L., & Tooby, J. (2002). Unraveling the enigma of human intelligence: Evolutionary psychology and the multimodular mind. *The evolution of intelligence*, 145-198.

Dembo M, Matzke NJ, Mooers AØ, Collard M. (2015). Bayesian analysis of a morphological supermatrix sheds light on controversial hominin relationships. Proc R Soc B [Internet] [cited 2015 Aug 20]; 282(1812). Available from: http://rspb.royalsocietypublishing.o...0943.e-letters. doi: 10.1098/rspb.2015.0943.

Domínguez-Rodrigo M, Pickering TR, Baquedano E, Mabulla A, Mark DF, Musiba C, et al. (2013). First partial skeleton of a 1.34-million-year-old Paranthropus boisei from Bed II, Olduvai Gorge, Tanzania. PLoS ONE 8(12): e80347. Available from: http://journals.plos.org/plosone/art...l.pone.0080347. doi:10.1371/journal.pone.0080347

Dunbar, R. I. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, *6* (5), 178-190.

eFossils [Internet] [cited 2015 Aug 10]. Department of Anthropology, The University of Texas at Austin. Available from: http://efossils.org/

eLucy: Step-by-step: the evolution of bipedalism [Internet]. c2007. Austin (TX): Department of Anthropology, University of Texas-Austin. [cited 2015 Aug 3]. Available from: elucy.org/Main/LessonOverview.html

Geertz, A. W. (2013). The meaningful brain: Clifford Geertz and the cognitive science of culture. In: *Mental Culture: Classical Social Theory and the Cognitive Science of Religion.* Edited by Dimitris Xygalatas and William W. McCorkle Jr., 176-196.

Gibbons, A. (2012). Generation Gaps Suggest Ancient Human-Ape Split. *Science*, August 13, 2012. American Association for the Advancement of Science. https://www.science.org/content/arti...%20years%20ago. Retrieved July 4, 2022.

Hawks J. (2005). The Great Rift Valley. John Hawks weblog [Internet] [cited 2015 Aug 2]. Available from:http://johnhawks.net/weblog/topics/geology/rift/rift_valley_overview.html

Harmand S, Lewis JE, Feibel CS, Lepre CJ, Prat S, Lenoble A, Boës X, Quinn RL, Brenet M, Arroyo A, Taylor N, Clément S, Daver G, Brugal JP, Leakey L, Mortlock RA, Wright JD, Lokorodi S, Kirwa C, Kent DV, Roche H. 2015. 3.3-million-year-old tools from Lomekwi 3, West Turkana, Kenya. Nature 521 (May 21): 310-315. Available from: http://www.nature.com.offcampus.lib....ture14464.html. doi:10.1038/nature14464.

Hunt KD. Australopithecines. In: Encyclopedia of anthropology, Vol. 1. Thousand Oaks (CA): SAGE Reference, 2006. p.311-317.

Higham T, Douka K, Wood R et al. (2014). The timing and spatiotemporal patterning of Neanderthal disappearance. Nature [Internet] [cited 2015 Aug 21]; 512 (7514): 306-309. Available from: http://www.nature.com.offcampus.lib....ture13621.html.

doi:10.1038/nature13621

Jurmain R, Kilgore L, Trevathan W. (2013). Essentials of physical anthropology, 9th edition. Belmont (CA): Wadsworth Cengage Learning.

Koenigshofer, K. A. (2017). General Intelligence: Adaptation to Evolutionarily Familiar Abstract Relational Invariants, Not to Environmental or Evolutionary Novelty. *The Journal of Mind and Behavior*, *38* (2), 119-153.

Larsen, CS. (2014). Our origins: discovering physical anthropology. New York (NY): W. W. Norton & Company, Inc.

Lebatard AE, Bouriès DL, Duriner P, Jolivet M, Braucjer R, Carcaillet J, Schuster M, Arnaud N, Monié P, Lihoreau F, Likius A, Macaye HT, Vignaud P, Brunet M. 2008. Cosmogenic nuclide dating of Sahelanthropus tchadensis and Australopithecus bahrelghazali: Mio-Pliecene hominids from Chad. Proc Natl Acad Sci U S A [Internet] [cited 2015 Aug 3]; 105(9): 3226-3231. Available from: http://www.pnas.org/content/105/9/3226.full. doi: 10.1073/pnas.0708015105

Lice and human evolution [Internet]. 2011 Feb 6. NOVAscienceNow. [cited 2015 Aug 20]. Available from: www.pbs.org/wgbh/nova/evolution/lice.html

Meyer M, Kircher M, Gansauge MT, et al. (2012). A high-coverage genome sequence from an archaic Denisovan individual. Science [Internet] [cited 2015 Aug 21]; 338 (6104): 222-226. Available from: http://www.sciencemag.org.offcampus..../6104/222.full. doi: 10.1126/science.1224344.

NOVA [Internet]. 2002 Jan 22. Neanderthals on trial. PBS [cited 2015 Aug 21]. Available from: http://www.pbs.org/wgbh/nova/transcr...nderthals.html

O'Neil D. c1999-2012. Early human evolution: a survey of the biological and cultural evolution of Homo habilis and Homo erectus. Behavioral Sciences Department, Palomar College [Internet] [cited 2015 Aug 16]. Available from: anthro.palomar.edu/homo/Default.htm

O'Neil D. c1999-2014. Evolution of modern humans: a survey of the biological and cultural evolution of archaic and modern Homo sapiens. Behavioral Sciences Department, Palomar College [Internet] [cited 2015 Aug 16]. Available from: anthro.palomar.edu/homo2/default.htm

Osiurak, F., & Massen, C. (2014). The cognitive and neural bases of human tool use. *Frontiers in psychology*, *5*, 1107.

Oxnard, C., Obendorf, P.J., Kefford, B.J. (2010). Post-cranial skeletons of hyperthyroid cretins show a similar anatomical mosaic as Homo floesiensis. PLoS One [Internet] [cited 2015 Aug 21]; 5 (9). Available from: http://www.sciencedaily.com/releases...0928025514.htm. doi: 10.1371/journal.pone.0013018.

Pääbo S. 2015. The diverse origins of the human gene pool. Nat Rev Genet [Internet] [cited 2015 Aug 21]; 16: 313-314. Available from: http://www.nature.com.offcampus.lib....l/nrg3954.html. doi:10.1038/nrg3954

Reich D, Green RE, Kircher M, et al. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature [Internet] [cited 2015 Aug 21]; 468 (7327). Available from: http://www.nature.com/nature/journal...ture09710.html

Richmond BG, Jungers WL. (2008). Orrorin tugenensis femoral morphology and the evolution of hominin bipedalism. Science [Internet] [cited 2015 Aug 3]; 319(5870): 1662-1665. Available from: www.jstor.org/stable/20053635

Scarre C. (2005).*The Human Past: World Prehistory&the Development of Human Societies*. London (UK): Thames & Hudson Ltd.

Scarre C. (2013). The human past. London (UK): Thames & Hudson.

Sci-News.com [Internet]. 2013 Dec 6. Paranthropus boisei: 1.34-million-year-old hominin found in Tanzania. [cited 2015 Aug 12]. Available from: http://www.sci-news.com/othersciences/anthropology/science-paranthropus-boisei-hominin-tanzania-01603.html

Smithsonian Institution [Internet]. 2015 Aug 4. What does it mean to be human? [cited 2015 Aug 11]. Available from: http://humanorigins.si.edu/evidence/human-fossils/species

Su DF. (2013). The earliest hominins: Sahelanthropus, Orrorin, and Ardipithecus. The Nature Education Knowledge Project [Internet] [cited 2015 Aug 3]. Available from: http://www.nature.com/scitable/knowl...hecus-67648286

Vaesen, K. (2012). The cognitive bases of human tool use. *Behavioral and brain sciences*, *35* (4), 203-218.

Ward CV, Manthi FK, Plavcan JM. (2013). New fossils of Australopithecus anamensis from Kanapoi, West Turkana, Kenya (2003-2008). J Hum Evol 65(5): 501-524. Available from doi:10.1016/j.jhevol.2013.05.006

Wong K. 2006. Special report: Lucy's baby: an extraordinary new human fossil comes to light. Sci Am [Internet] [cited 2015 Aug 10]; Sep 20. Available from: http://www.scientificamerican.com/ar...rt-lucys-baby/

Wood, B. (2010). Reconstructing human evolution: Achievements, challenges, and opportunities. *Proceedings of the National Academy of Sciences*, *107* (supplement_2), 8902-8909.

University of Chicago Press Journals [Internet]. Volcanoes wiped out Neanderthals, new study suggests. ScienceDaily 2010 Oct 7 [cited 2015 Aug 21]. Available from: http://www.sciencedaily.com/releases...1006094057.htm

References from Brain Evolution:

American Association for the Advancement of Science (AAAS). (1880). Dr. Paul Broca. *Science*, 1(8), 93. http://www.jstor.org/stable/2900242

Bullock, T. H. (1984). Comparative neuroscience holds promise for quiet revolutions. *Science*, 225(4661), 473–478.

Cachel, S. (2012). Human tool behavior is species-specific and remains unique. *Behavioral and Brain Sciences*, *35* (4), 222.

Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, 2, 263–275.

Darwin, C. (1871). *The descent of man, and the selection in relation to sex*. London: J. Murray.

Darwin, C. (1859). On the origins of species by means of natural selection, or, *The preservation of favoured races in the struggle for life*. London, UK: J. Murray.

Demonet, J. F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J. L., Wise, R., . . . Frackowiak, R. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain*, 115(6), 1753–1768.

Edelman, G. (2004). *Wider than the sky: The phenomenal gift of consciousness*. New Haven, CT: Yale University Press.

Gardner, R. A., & Gardner, B. T. (1969). Teaching sign language to a chimpanzee. *Science*, 165(3894), 664–672.

Goodall, J. (1986). *The chimpanzees of Gombe: Patterns of behavior*. Cambridge, MA: Harvard University Press.

Hubel, D. H. (1995). *Eye, brain, and vision*. Freeman & Co., NY: Scientific American Library/Scientific American Books.

Knapp, M. L., & Hall, J. A. (2009). *Nonverbal communication in human interaction*. Boston, MA: Wadsworth Cengage Learning.

Osiurak, F., & Massen, C. (2014). The cognitive and neural bases of human tool use. *Frontiers in psychology*, *5*, 1107.

Ramón y Cahal, S. (1937). Recollections of my life. *Memoirs of the American Philosophical Society*, 8, 1901–1917.

Rolls, E. T., & Cowey, A. (1970). Topography of the retina and striate cortex and its relationship to visual acuity in rhesus monkeys and squirrel monkeys. *Experimental Brain Research*, 10(3), 298–310.

Rosenzweig, M. R., Breedlove, S. M., & Leiman, A. L. (2002). *Biological psychology* (3rd ed.). Sunderland, MA: Sinauer Associates.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, 283(5408), 1657–1661.

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive psychology*, 33, 5–42.

Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: An integrated systems perspective. *Science*, *255*(5043), 419–423.

Walker, L. A. (1987). *A loss for words: The story of deafness in a family*. New York, NY: Harper Perennial.

3.3.22.1 Attributions

"Human Evolution" text and figures not attributed to other sources adapted by Kenneth A. Koenigshofer, PhD, Chaffey College, from Biological Anthropology (Saneda and Field), chapter: Human Evolution; LibreTexts by Tori Saneda & Michelle Field, Professors (Anthropology) at Cascadia Community College (via Wikieducator)

"Defining Hominins" text and figures not attributed to other sources were adapted and modified by Kenneth A. Koenigshofer, Ph.D., Chaffey College, from *Explorations: An Open Invitation to Biological Anthropology;* https://explorations.americananthro.....php/chapters/; Chapter 9: "Early Hominins," Editors: Beth Shook, Katie Nelson, Kelsie Aguilera and Lara Braff, American Anthropological Association Arlington, VA 2019; CC BY-NC 4.0 International, except where otherwise noted.

"Homo Genus," "Defining the Genus Homo," "Climate," "Homo habilis: The Earliest Members of Our Genus," "Homo erectus," "Homo erectus Culture," "Tool Use and Cognitive Abilities of Homo erectus" and "The Big Picture of Early Homo," text and figures not attributed to other sources adapted by Kenneth A. Koenigshofer, PhD, Chaffey College, from *Explorations: An Open Invitation to Biological Anthropology;* https://explorations.americananthro.....php/chapters/; Chapter 10, "Early Members of the Genus Homo" by Bonnie Yoshida-Levine Ph.D.; Beth Shook, Katie Nelson, Kelsie Aguilera, and Lara Braff, Eds.; licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, except where otherwise noted.

"Brain Evolution" text and figures not attributed to other sources adapted by Kenneth A. Koenigshofer, PhD, Chaffey College, from The Nervous System by Aneeq Ahmad, licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Permissions beyond the scope of this license may be available in our Licensing Agreement.

Table 3.3: "Summary features of Homo habilis" adapted by Kenneth A. Koenigshofer, PhD, Chaffey College, from *Explorations: An Open Invitation to Biological Anthropology;* https://explorations.americananthro.....php/chapters/; Chapter 10, "Early Members of the Genus Homo" by Bonnie Yoshida-Levine Ph.D.; Beth Shook, Katie Nelson, Kelsie Aguilera, and Lara Braff, Eds.; licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, except where otherwise noted.

Creative Commons License



This page titled 3.3: Human Evolution is shared under a mixed license and was authored, remixed, and/or curated by Keneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

- **3.1: Trends** by Tori Saneda & Michelle Field is licensed CC BY-SA 4.0. Original source: https://wikieducator.org/Biological_Anthropology.
- **2.3: Modern Primates** by Tori Saneda & Michelle Field is licensed CC BY-SA 4.0. Original source: https://wikieducator.org/Biological_Anthropology.
- **3.2: Proto-Hominins** by Tori Saneda & Michelle Field is licensed CC BY-SA 4.0. Original source: https://wikieducator.org/Biological_Anthropology.
- **3.3: Homo Genus** by Tori Saneda & Michelle Field is licensed CC BY-SA 4.0. Original source: https://wikieducator.org/Biological_Anthropology.

3.4: Evolutionary Theories in Psychology

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Explain the meaning of “evolution.”
2. Describe the primary mechanisms by which evolution takes place.
3. Describe the two major classes of adaptations.
4. Explain sexual selection and its two primary processes.
5. Describe gene selection theory.
6. Discuss psychological adaptations.
7. Explain the core premises of sexual strategies theory.
8. Explain the core premises of error management theory, and provide two empirical examples of adaptive cognitive biases.

3.4.1 Overview

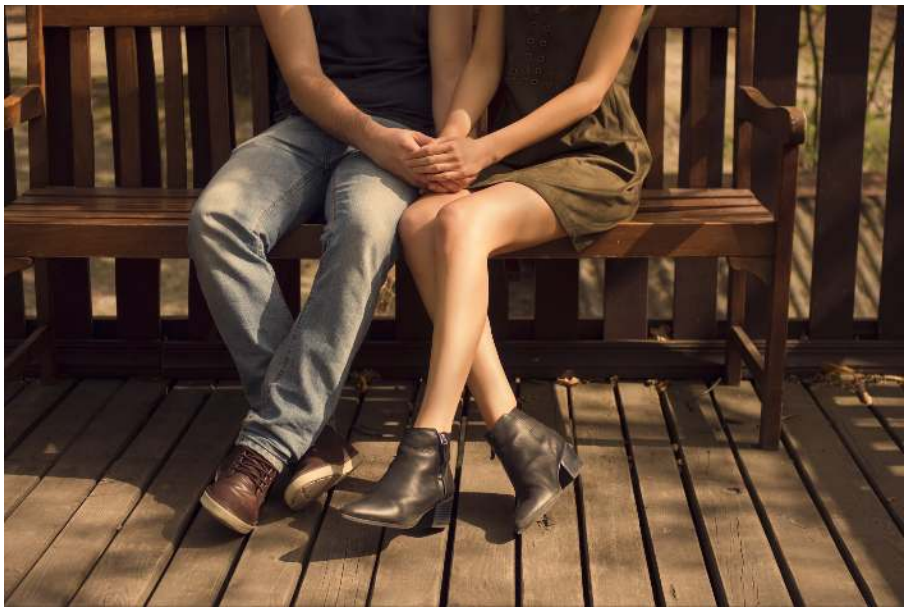


Figure **3.4.1**: It may seem like just a casual date, but don't doubt that the forces of evolution are hard at work below the surface. [Image: Best Couples, https://goo.gl/aBMY6W, CC BY-SA 2.0, https://goo.gl/jSSrcO]

If you have ever been on a first date, you’re probably familiar with the anxiety of trying to figure out what clothes to wear or what perfume or cologne to put on. In fact, you may even consider flossing your teeth for the first time all year. When considering why you put in all this work, you probably recognize that you’re doing it to impress the other person. But how did you learn these particular behaviors? Where did you get the idea that a first date should be at a nice restaurant or someplace unique? It is possible that we have been taught these behaviors by observing others. It is also possible, however, that these behaviors—the fancy clothes, the expensive restaurant—are biologically programmed into us. That is, just as peacocks display their feathers to show how attractive they are, or some lizards do push-ups to show how strong they are, when we style our hair or bring a gift to a date, we’re trying to communicate to the other person: “Hey, I’m a good mate! Choose me! Choose me!"

However, we all know that our ancestors hundreds of thousands of years ago weren’t driving sports cars or wearing designer clothes to attract mates. So how could someone ever say that such behaviors are “biologically programmed” into us? Well, even though our ancestors might not have been doing these specific actions, these behaviors are the result of the same driving force: the powerful influence of evolution (Links to an external site.) (Links to an external site.). Yes, evolution—certain traits and behaviors

developing over time because they are advantageous to our survival. In the case of dating, doing something like offering a gift might represent more than a nice gesture. Just as chimpanzees will give food to mates to show they can provide for them, when you offer gifts to your dates, you are communicating that you have the money or "resources" to help take care of them. And even though the person receiving the gift may not realize it, the same evolutionary forces are influencing his or her behavior as well. The receiver of the gift evaluates not only the gift but also the gift-giver's clothes, physical appearance, and many other qualities, to determine whether the individual is a suitable mate. But because these evolutionary processes are hardwired into us, it is easy to overlook their influence.

To broaden your understanding of evolutionary processes, this section will present some of the most important elements of evolution as they impact psychology. Evolutionary theory helps us piece together the story of how we humans have prospered. It also helps to explain why we behave as we do on a daily basis in our modern world: why we bring gifts on dates, why we get jealous, why we crave our favorite foods, why we protect our children, and so on. Evolution may seem like a historical concept that applies only to our ancient ancestors but, in truth, it is still very much a part of our modern daily lives.

3.4.2 Basics of Evolutionary Theory

Evolution simply means change over time. Many think of evolution as the development of traits and behaviors that allow us to survive this "dog-eat-dog" world, like strong leg muscles to run fast, or fists to punch and defend ourselves. However, physical survival is only important if it eventually contributes to successful reproduction. That is, even if you live to be a 100-year-old, if you fail to mate and produce children, your genes will die with your body. Thus, reproductive success, not survival success, is the engine of evolution by natural selection. Every mating success by one person means the loss of a mating opportunity for another. Yet every living human being is an evolutionary success story. Each of us is descended from a long and unbroken line of ancestors who triumphed over others in the struggle to survive (at least long enough to mate) and reproduce. However, in order for our genes to endure over time—to survive harsh climates, to defeat predators—we have inherited adaptive, psychological processes designed to ensure success.

At the broadest level, we can think of organisms, including humans, as having two large classes of adaptations —or traits and behaviors that evolved over time to increase our reproductive success. The first class of adaptations are called survival adaptations: mechanisms that helped our ancestors handle the "hostile forces of nature." For example, in order to survive very hot temperatures, we developed sweat glands to cool ourselves. In order to survive very cold temperatures, we developed shivering mechanisms (the speedy contraction and expansion of muscles to produce warmth). Other examples of survival adaptations include developing a craving for fats and sugars, encouraging us to seek out particular foods rich in fats and sugars that keep us going longer during food shortages. Some threats, such as snakes, spiders, darkness, heights, and strangers, often produce fear in us, which encourages us to avoid them and thereby stay safe. These are also examples of survival adaptations. However, all of these adaptations are for physical survival, whereas the second class of adaptations are for reproduction, and help us compete for mates. These adaptations are described in an evolutionary theory proposed by Charles Darwin, called sexual selection theory.

3.4.3 Sexual Selection Theory

Darwin noticed that there were many traits and behaviors of organisms that could not be explained by "survival selection." For example, the brilliant plumage of peacocks should actually lower their rates of survival. That is, the peacocks' feathers act like a neon sign to predators, advertising "Easy, delicious dinner here!" But if these bright feathers only lower peacocks' chances at survival, why do they have them? The same can be asked of similar characteristics of other animals, such as the large antlers of male stags or the wattles of roosters, which also seem to be unfavorable to survival. Again, if these traits only make the animals less likely to survive, why did they develop in the first place? And how have these animals continued to survive with these traits over thousands and thousands of years? Darwin's answer to this conundrum was the theory of sexual selection: the evolution of characteristics, not because of survival advantage, but because of mating advantage.

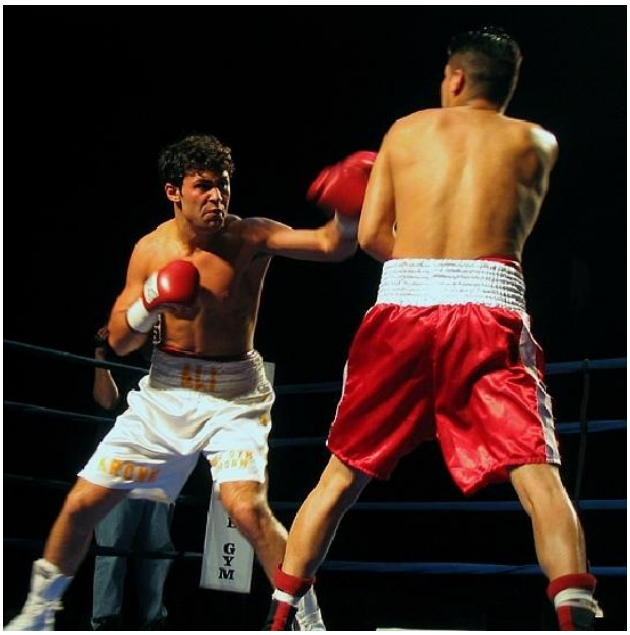


Figure **3.4.2**: Modern sports like boxing can be seen as modified/stylized versions of the evolutionary behavior of intrasexual competition. [Image: Dave Hogg, https://goo.gl/fL5U2Z, CC BY 2.0, https://goo.gl/9uSnqN]

Sexual selection occurs through two processes. The first, intrasexual competition, occurs when members of one sex compete against each other, and the winner gets to mate with a member of the opposite sex. Male stags, for example, battle with their antlers, and the winner (often the stronger one with larger antlers) gains mating access to the female. That is, even though large antlers make it harder for the stags to run through the forest and evade predators (which lowers their survival success), they provide the stags with a better chance of attracting a mate (which increases their reproductive success). Similarly, human males sometimes also compete against each other in physical contests: boxing, wrestling, karate, or group-on-group sports, such as football. Even though engaging in these activities poses a "threat" to their survival success, as with the stag, the victors are often more attractive to potential mates, increasing their reproductive success. Thus, whatever qualities lead to success in intrasexual competition are then passed on with greater frequency due to their association with greater mating success.

The second process of sexual selection is preferential mate choice, also called intersexual selection. In this process, if members of one sex are attracted to certain qualities in mates—such as brilliant plumage, signs of good health, or even intelligence—those desired qualities get passed on in greater numbers, simply because their possessors mate more often. For example, the colorful plumage of peacocks exists due to a long evolutionary history of peahens' (the term for female peacocks) attraction to males with brilliantly colored feathers.

In all sexually-reproducing species, adaptations in both sexes (males and females) exist due to survival selection and sexual selection. However, unlike other animals where one sex has dominant control over mate choice, humans have "mutual mate choice." That is, both women and men typically have a say in choosing their mates. And both mates value qualities such as kindness, intelligence, and dependability that are beneficial to long-term relationships—qualities that make good partners and good parents.

3.4.4 Gene Selection Theory

In modern evolutionary theory, all evolutionary processes boil down to an organism's genes. Genes are the basic "units of heredity," or the information that is passed along in DNA that tells the cells and molecules how to "build" the organism and how that organism should behave. Genes that are better able to encourage the organism to reproduce, and thus replicate themselves in the organism's offspring, have an advantage over competing genes that are less able. For example, take female sloths: In order to attract a mate, they will scream as loudly as they can, to let potential mates know where they are in the thick jungle. Now, consider two types of genes in female sloths: one gene that allows them to scream extremely loudly, and another that only allows them to scream moderately loudly. In this case, the sloth with the gene that allows her to shout louder will attract more mates—increasing reproductive success—which ensures that her genes are more readily passed on than those of the quieter sloth.

Essentially, genes can boost their own replicative success in two basic ways. First, they can influence the odds for survival and reproduction of the organism they are in (individual reproductive success or fitness—as in the example with the sloths). Second, genes can also influence the organism to help other organisms who also likely contain those genes—known as "genetic relatives"—to survive and reproduce (which is called inclusive fitness). For example, why do human parents tend to help their own kids with the financial burdens of a college education and not the kids next door? Well, having a college education increases one's attractiveness to other mates, which increases one's likelihood for reproducing and passing on genes. And because parents' genes are in their own children (and not the neighborhood children), funding their children's educations increases the likelihood that the parents' genes will be passed on.

Understanding gene replication is the key to understanding modern evolutionary theory. It also fits well with many evolutionary psychological theories. However, for the time being, we'll ignore genes and focus primarily on actual adaptations that evolved because they helped our ancestors survive and/or reproduce.

3.4.5 Evolutionary Psychology

Evolutionary psychology aims the lens of modern evolutionary theory on the workings of the human mind. It focuses primarily on psychological adaptations: mechanisms of the mind that have evolved to solve specific problems of survival or reproduction. These kinds of adaptations are in contrast to physiological adaptations, which are adaptations that occur in the body as a consequence of one's environment. One example of a physiological adaptation is how our skin makes calluses. First, there is an "input," such as repeated friction to the skin on the bottom of our feet from walking. Second, there is a "procedure," in which the skin grows new skin cells at the afflicted area. Third, an actual callus forms as an "output" to protect the underlying tissue—the final outcome of the physiological adaptation (i.e., tougher skin to protect repeatedly scraped areas). On the other hand, a psychological adaptation is a development or change of a mechanism in the mind. For example, take sexual jealousy. First, there is an "input," such as a romantic partner flirting with a rival. Second, there is a "procedure," in which the person evaluates the threat the rival poses to the romantic relationship. Third, there is a behavioral output, which might range from vigilance (e.g., snooping through a partner's email) to violence (e.g., threatening the rival).

Evolutionary psychology is fundamentally an interactionist framework, or a theory that takes into account multiple factors when determining the outcome. For example, jealousy, like a callus, doesn't simply pop up out of nowhere. There is an "interaction" between the environmental trigger (e.g., the flirting; the repeated rubbing of the skin) and the initial response (e.g., evaluation of the flirter's threat; the forming of new skin cells) to produce the outcome.

In evolutionary psychology, culture also has a major effect on psychological adaptations. For example, status within one's group is important in all cultures for achieving reproductive success, because higher status makes someone more attractive to mates. In individualistic cultures, such as the United States, status is heavily determined by individual accomplishments. But in more collectivist cultures, such as Japan, status is more heavily determined by contributions to the group and by that group's success. For example, consider a group project. If you were to put in most of the effort on a successful group project, the culture in the United States reinforces the psychological adaptation to try to claim that success for yourself (because individual achievements are rewarded with higher status). However, the culture in Japan reinforces the psychological adaptation to attribute that success to the whole group (because collective achievements are rewarded with higher status). Another example of cultural input is the importance of virginity as a desirable quality for a mate. Cultural norms that advise against premarital sex persuade people to ignore their own basic interests because they know that virginity will make them more attractive marriage partners. Evolutionary psychology, in short, does not predict rigid robotic-like "instincts." That is, there isn't one rule that works all the time. Rather, evolutionary psychology studies flexible, environmentally-connected and culturally-influenced adaptations that vary according to the situation.

Psychological adaptations are hypothesized to be wide-ranging, and include food preferences, habitat preferences, mate preferences, and specialized fears. These psychological adaptations also include many traits that improve people's ability to live in groups, such as the desire to cooperate and make friends, or the inclination to spot and avoid frauds, punish rivals, establish status hierarchies, nurture children, and help genetic relatives. Research programs in evolutionary psychology develop and empirically test predictions about the nature of psychological adaptations. Below, we highlight a few evolutionary psychological theories and their associated research approaches.

3.4.6 Sexual Strategies Theory

Sexual strategies theory is based on sexual selection theory. It proposes that humans have evolved a list of different mating strategies, both short-term and long-term, that vary depending on culture, social context, parental influence, and personal mate value (desirability in the "mating market").

In its initial formulation, sexual strategies theory focused on the differences between men and women in mating preferences and strategies (Buss & Schmitt, 1993). It started by looking at the minimum parental investment needed to produce a child. For women, even the minimum investment is significant: after becoming pregnant, they have to carry that child for nine months inside of them. For men, on the other hand, the minimum investment to produce the same child is considerably smaller—simply the act of sex.



Figure **3.4.3**: Because women bear responsibility for pregnancy, they may use different sexual selection strategies than men do. [Image: CC0 Public Domain, https://goo.gl/m25gce]

These differences in parental investment have an enormous impact on sexual strategies. For a woman, the risks associated with making a poor mating choice is high. She might get pregnant by a man who will not help to support her and her children, or who might have poor-quality genes. And because the stakes are higher for a woman, wise mating decisions for her are much more valuable. For men, on the other hand, the need to focus on making wise mating decisions isn't as important. That is, unlike women, men 1) don't biologically have the child growing inside of them for nine months, and 2) do not have as high a cultural expectation to raise the child. This logic leads to a powerful set of predictions: In short-term mating, women will likely be choosier than men (because the costs of getting pregnant are so high), while men, on average, will likely engage in more casual sexual activities (because this cost is greatly lessened). Due to this, men will sometimes deceive women about their long-term intentions for the benefit of short-term sex, and men are more likely than women to lower their mating standards for short-term mating situations.

An extensive body of empirical evidence supports these and related predictions (Buss & Schmitt, 2011). Men express a desire for a larger number of sex partners than women do. They let less time elapse before seeking sex. They are more willing to consent to sex with strangers and are less likely to require emotional involvement with their sex partners. They have more frequent sexual fantasies and fantasize about a larger variety of sex partners. They are more likely to regret missed sexual opportunities. And they lower their standards in short-term mating, showing a willingness to mate with a larger variety of women as long as the costs and risks are low.

However, in situations where both the man and woman are interested in long-term mating, both sexes tend to invest substantially in the relationship and in their children. In these cases, the theory predicts that both sexes will be extremely choosy when pursuing a long-term mating strategy. Much empirical research supports this prediction, as well. In fact, the qualities women and men generally look for when choosing long-term mates are very similar: both want mates who are intelligent, kind, understanding, healthy, dependable, honest, loyal, loving, and adaptable.

Nonetheless, women and men do differ in their preferences for a few key qualities in long-term mating, because of somewhat distinct adaptive problems. Modern women have inherited the evolutionary trait to desire mates who possess resources, have qualities linked with acquiring resources (e.g., ambition, wealth, industriousness), and are willing to share those resources with them. On the other hand, men more strongly desire youth and health in women, as both are cues to fertility. These male and female differences are universal in humans. They were first documented in 37 different cultures, from Australia to Zambia (Buss, 1989), and have been replicated by dozens of researchers in dozens of additional cultures (for summaries, see Buss, 2012).

As we know, though, just because we have these mating preferences (e.g., men with resources; fertile women), people don't always get what they want. There are countless other factors which influence who people ultimately select as their mate. For example, the sex ratio (the percentage of men to women in the mating pool), cultural practices (such as arranged marriages, which inhibit individuals' freedom to act on their preferred mating strategies), the strategies of others (e.g., if everyone else is pursuing short-term sex, it's more difficult to pursue a long-term mating strategy), and many others all influence who we select as our mates.

Sexual strategies theory—anchored in sexual selection theory— predicts specific similarities and differences in men and women's mating preferences and strategies. Whether we seek short-term or long-term relationships, many personality, social, cultural, and ecological factors will all influence who our partners will be.

3.4.7 Error Management Theory



Figure **3.4.4**: If you were walking in the woods and heard a sound in the bushes you might be startled and act on the worst case scenario—such as the threat of a wild animal—by moving in the opposite direction. This is evolutionary psychology at work, keeping you safe so you can survive and reproduce. [Image: Nicholas T, https://goo.gl/gZ3zEL, CC BY 2.0, https://goo.gl/9uSnqN]

Error management theory (EMT) deals with the evolution of how we think, make decisions, and evaluate uncertain situations—that is, situations where there's no clear answer how we should behave (Haselton & Buss, 2000; Haselton, Nettle, & Andrews, 2005). Consider, for example, walking through the woods at dusk. You hear a rustle in the leaves on the path in front of you. It could be a snake. Or, it could just be the wind blowing the leaves. Because you can't really tell why the leaves rustled, it's an uncertain situation. The important question then is, what are the costs of errors in judgment? That is, if you conclude that it's a dangerous snake so you avoid the leaves, the costs are minimal (i.e., you simply make a short detour around them). However, if you assume the leaves are safe and simply walk over them—when in fact it is a dangerous snake—the decision could cost you your life.

Now, think about our evolutionary history and how generation after generation was confronted with similar decisions, where one option had low cost but great reward (walking around the leaves and not getting bitten) and the other had a low reward but high cost (walking through the leaves and getting bitten). These kinds of choices are called "cost asymmetries." If during our evolutionary history we encountered decisions like these generation after generation, over time an adaptive bias would be created: we would make sure to err in favor of the least costly (in this case, least dangerous) option (e.g., walking around the leaves). To put it another way, EMT predicts that whenever uncertain situations present us with a safer versus more dangerous decision, we will psychologically adapt to prefer choices that minimize the cost of errors.

EMT is a general evolutionary psychological theory that can be applied to many different domains of our lives, but a specific example of it is the visual descent illusion. To illustrate: Have you ever thought it would be no problem to jump off of a ledge, but as soon as you stood up there, it suddenly looked much higher than you thought? The visual descent illusion (Jackson & Cormack, 2008) states that people will overestimate the distance when looking down from a height (compared to looking up) so that people will be especially wary of falling from great heights—which would result in injury or death. Another example of EMT is the auditory looming bias: Have you ever noticed how an ambulance seems closer when it's coming toward you, but suddenly seems far away once it's immediately passed? With the auditory looming bias, people overestimate how close objects are when the sound is moving toward them compared to when it is moving away from them. From our evolutionary history, humans learned, "It's better to be safe than sorry." Therefore, if we think that a threat is closer to us when it's moving toward us (because it seems louder), we will be quicker to act and escape. In this regard, there may be times we ran away when we didn't need to (a false alarm), but wasting that time is a less costly mistake than not acting in the first place when a real threat does exist.

EMT has also been used to predict adaptive biases in the domain of mating. Consider something as simple as a smile. In one case, a smile from a potential mate could be a sign of sexual or romantic interest. On the other hand, it may just signal friendliness. Because of the costs to men of missing out on chances for reproduction, EMT predicts that men have a sexual overperception bias: they often misread sexual interest from a woman, when really it's just a friendly smile or touch. In the mating domain, the sexual overperception bias is one of the best-documented phenomena. It's been shown in studies in which men and women rated the sexual interest between people in photographs and videotaped interactions. As well, it's been shown in the laboratory with participants engaging in actual "speed dating," where the men interpret sexual interest from the women more often than the women actually intended it (Perilloux, Easton, & Buss, 2012). In short, EMT predicts that men, more than women, will over-infer sexual interest based on minimal cues, and empirical research confirms this adaptive mating bias.

3.4.8 Summary

Sexual strategies theory and error management theory are two evolutionary psychological theories that have received much empirical support from dozens of independent researchers. But, there are many other evolutionary psychological theories, such as social exchange theory for example, that also make predictions about our modern day behavior and preferences, too. The merits of each evolutionary psychological theory, however, must be evaluated separately and treated like any scientific theory. That is, we should only trust their predictions and claims to the extent they are supported by scientific studies. However, even if the theory is scientifically grounded, just because a psychological adaptation was advantageous in our history, it doesn't mean it's still useful today. For example, even though women may have preferred men with resources in generations ago, our modern society has advanced such that these preferences are no longer apt or necessary. Nonetheless, it's important to consider how our evolutionary history has shaped our automatic or "instinctual" desires and reflexes of today, so that we can better shape them for the future ahead.

3.4.9 Outside Resources

FAQs

http://www.anth.ucsb.edu/projects/human/evpsychfaq.html (Links to an external site.) (Links to an external site.)

Web: Articles and books on evolutionary psychology

http://homepage.psy.utexas.edu/homepage/Group/BussLAB/ (Links to an external site.) (Links to an external site.)

Web: Main international scientific organization for the study of evolution and human behavior, HBES

http://www.hbes.com/ (Links to an external site.) (Links to an external site.)

3.4.10 Discussion Questions

1. How does change take place over time in the living world?
2. Which two potential psychological adaptations to problems of survival are not discussed in this module?
3. What are the psychological and behavioral implications of the fact that women bear heavier costs to produce a child than men do?
4. Can you formulate a hypothesis about an error management bias in the domain of social interaction?

3.4.11 Vocabulary

- Adaptations: Evolved solutions to problems that historically contributed to reproductive success.
- Error management theory (EMT): A theory of selection under conditions of uncertainty in which recurrent cost asymmetries of judgment or inference favor the evolution of adaptive cognitive biases that function to minimize the more costly errors.
- Evolution: Change over time. Is the definition changing?
- Gene Selection Theory: The modern theory of evolution by selection by which differential gene replication is the defining process of evolutionary change.
- Intersexual selection: A process of sexual selection by which evolution (change) occurs as a consequences of the mate preferences of one sex exerting selection pressure on members of the opposite sex.
- Intrasexual competition: A process of sexual selection by which members of one sex compete with each other, and the victors gain preferential mating access to members of the opposite sex.

- Natural selection: Differential reproductive success as a consequence of differences in heritable attributes.
- Psychological adaptations: Mechanisms of the mind that evolved to solve specific problems of survival or reproduction; conceptualized as information processing devices.
- Sexual selection: The evolution of characteristics because of the mating advantage they give organisms.
- Sexual strategies theory: A comprehensive evolutionary theory of human mating that defines the menu of mating strategies humans pursue (e.g., short-term casual sex, long-term committed mating), the adaptive problems women and men face when pursuing these strategies, and the evolved solutions to these mating problems.

3.4.12 References

1. Buss, D. M. (2012). Evolutionary psychology: The new science of the mind (4th ed.). Boston, MA: Allyn & Bacon.
2. Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. Behavioral & Brain Sciences, 12, 1–49.
3. Buss, D. M., & Schmitt, D. P. (2011). Evolutionary psychology and feminism. Sex Roles, 64, 768–787.
4. Buss, D. M., & Schmitt, D. P. (1993). Sexual strategies theory: An evolutionary perspective on human mating. Psychological Review, 100, 204–232.
5. Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. Journal of Personality and Social Psychology, 78, 81–91.
6. Haselton, M. G., Nettle, D., & Andrews, P. W. (2005). The evolution of cognitive bias. In D. M. Buss (Ed.), The handbook of evolutionary psychology (pp. 724–746). New York, NY: Wiley.
7. Jackson, R. E., & Cormack, J. K. (2008). Evolved navigation theory and the environmental vertical illusion. Evolution and Human Behavior, 29, 299–304.
8. Perilloux, C., Easton, J. A., & Buss, D. M. (2012). The misperception of sexual interest. Psychological Science, 23, 146–151.

3.4.13 Attributions

Evolutionary Theories in Psychology by David M. Buss via Noba, licensed CC BY-NC-SA 4.0.

This page titled 3.4: Evolutionary Theories in Psychology is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

3.5: DNA, Chromosomes, and Gene-Environment Correlations

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Explain the role chromosomes play in carrying genetic information.
2. Explain the relationship between genes and environment including types of gene-environment correlations.
3. Discuss the role of twin and adoption studies in investigating genetic influences on mind and behavior.

3.5.1 Overview

Chromosomes are structures in the nucleus of a cell containing DNA, histone protein, and other structural proteins. Chromosomes also contain genes, most of which are made up of DNA and RNA. But our genes and chromosomes interact with environmental factors to determine our actual physical and behavioral traits (our phenotype). Teasing out the relative effects of genetic and environmental influences on the phenotype is often difficult because of the complex interactions between them. All traits, including psychological and behavioral traits, are determined by complex interactions between genetic and environmental factors.

3.5.2 DNA and the Genetic Code

DNA, or deoxyribonucleic acid, determines whether our eyes are blue or brown, how tall we will be, and even our dispositions for certain types of behavior. DNA holds our "genetic code;" it is shaped like a double helix, made of sequences of nucleic acids attached to a sugar phosphate backbone. Genes are subsections of DNA molecules linked together that encode a particular characteristic.

Each chromosome is made up of a single DNA molecule coiled around histone proteins. Research dating back to the 1800s shows that every living creature has a specific set of chromosomes in the nucleus of each of its cells.

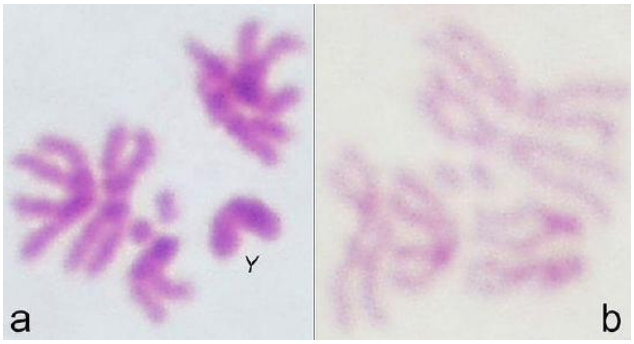


Figure 3.5.1: Human chromosome structure: Chromosomes are made up of a variety of gene sequences. By studying chromosomes and genes, scientists are able to determine the genetic basis for many diseases.

Human chromosomes are divided into two types—autosomes and sex chromosomes. Some genetic traits are linked to a person's sex and therefore passed on by the sex chromosomes. The autosomes contain the remainder of a person's genetic information. All human beings have 23 pairs of chromosomes by which genetic material is developed and characteristically demonstrated; 22 of these are autosomes, while the remaining pair (either XX, female, or XY, male) represents a person's sex chromosomes. These 23 pairs of chromosomes work together to create the person we ultimately become.

Chromosomal abnormalities can occur during fetal development if something goes wrong during the replication of the cells. Common abnormalities include Down syndrome (caused by an extra chromosome #21), Klinefelter syndrome (caused by an extra X chromosome), and Turner syndrome (caused by a missing X chromosome). Genetic counseling is available for families in order to determine if any abnormalities exist that may be passed along to offspring. Many chromosomal abnormalities are of psychological importance, with substantial impacts on mental processes; for example, Down syndrome can cause mild to moderate intellectual disabilities.

As science advances, the ability to manipulate genes and chromosomes is becoming increasingly sophisticated. Cloning is an example of taking chromosomal and genetic material and creating a new animal, and was first successfully achieved in the famous example of Dolly the sheep. There is much controversy surrounding the manipulation of chromosomes in human beings, with many people believing it to be unethical.



Figure **3.5.2**: Dolly the sheep: The first successfully cloned animal.

3.5.3 Gene-Environment Correlations: Nature or Nurture

Our genetic destiny is not necessarily written in stone. Genetic expression can be influenced by the environment, including various social factors, as well as physical environmental factors, ranging from light and temperature to exposure to chemicals (see Section 3.7 on Epigenetics).

The environment in which a person is raised can trigger the expression of behavior for which a person is genetically predisposed, while the same person raised in a different environment may exhibit different behavior.

Long-standing debates have taken place over which factor is more important, genes or environment. Is a person destined to have a particular outcome in life because of his or her genetic makeup, or can the environment (and the people in it) work to change what might be considered “bad” genes? Today, it is generally agreed that neither genes nor environment work alone; rather, the two work in tandem to create the people we ultimately become. However, because of gene-environment correlations, discriminating the effects of environment from the effects of genes can be very difficult.

We now know that genes can be turned on and off. Environmental elements like light and temperature have been shown to induce certain changes in genetic expression; additionally, exposure to drugs and chemicals can significantly affect how genes are expressed. People often inherit sensitivity to the effects of various environmental risk factors, and therefore different individuals may be differently affected by exposure to the same environment in medically significant ways. For example, some people become very ill from exposure to peanuts, while others are entirely unaffected. Another example is exposure to sunlight. Sunlight exposure has a much stronger influence on skin cancer risk in fair-skinned humans than in individuals with an inherited tendency for darker skin. The color of a person’s skin is largely genetic, but the environment also affects these genes in different ways.

3.5.4 Gene-environment Correlations by Type

Gene-environment correlations, represented as rGE, occur when genetic factors influence environmental exposure. Genes can affect environmental exposure indirectly via behavior. For example, high IQ parents may have more books in their homes than lower IQ parents leading to greater exposure to books and ideas in homes with high IQ parents compared to homes with lower IQ parents. In such a case, children of high IQ parents may read earlier and have greater facility with a wider range of ideas partly indirectly through the influence of genes for higher IQ on the selection of environments which facilitate exposure to books and ideas during childhood. However, the relation between early exposure to books and intellectual ability later in life may be misinterpreted as evidence of a purely environmental relationship and therefore support the view that nurture is the predominant influence. Such an interpretation would be mistaken because it missed the significant influence of the genes of the parents on selection of the home environment, along with direct genetic effects on intelligence, that led to later intellectual abilities of their children.

Gene-environment correlations, rGEs, can be causally related or non-causally related and can be categorized as passive, evocative, or active.

3.5.4.1 Passive Gene-Environment Correlations

In passive gene-environment correlation, an association exists between a person's genetic makeup and the environment in which he or she is raised. In other words, the person's environment, particularly in the case of children, is significantly determined by the parent's genetic characteristics. Parents create a home environment that is influenced by their own heritable characteristics. When the children's own genotype influences their behavior or cognitive outcomes, the result can be misinterpreted as a relationship between environment and psychological outcome. For example, an intelligent parent is likely to create a home environment rich in educational materials and experience. Since intelligence is moderately heritable, it can be argued that intelligence in the child is inherited rather than a factor of the home environment created by the parents. It is relatively unclear whether the genetic or environmental factors had more to do with the child's development because both genes and environment affected the child's development and many aspects of the child's environment were affected by the parents' genes when the home environment was created by the parents. Thus, correlations between environment and a child's cognitive characteristics may be misinterpreted as evidence for environmental influence that is really due more to genetic factors involved in selection of the environment, factors which have been missed and therefore left out of the analysis of causation.

3.5.4.2 Evocative Gene-Environment Correlations

Evocative gene-environment correlation happens when an individual's (heritable) behavior evokes an environmental response. For example, the association or correlation between marital conflict (arguing) and depression may reflect or be due to the tensions that arise from interacting with a depressed spouse rather than a causal effect of marital conflict on risk for depression. In other words, marital conflict may not cause increased risk for depression, but instead a genetic disposition for depression in one marriage partner may cause tensions between partners that lead to marital conflict. Therefore, what first appears as an environmental factor in depression may actually have genetic roots. Genes cause a disposition for depression, and depression in one partner leads to increased marital conflict. Therefore a correlation between marital conflict and depression may not be evidence for the influence of environment on depression, but the influence of genetics for depression on the environment (marital conflict).

3.5.4.3 Active Gene-Environment Correlations

In active gene-environment correlation, the person's genetic makeup may lead them to select particular environments. For example, a shy person is likely to choose quiet activities and less boisterous environments than an extroverted individual. He or she may be more likely to spend time at the library than at a dance club. Thus, a correlation between spending time at libraries and shyness may not indicate the influence of libraries, an environmental factor, on shyness, but instead reflects the effect of genes for shyness on selection of libraries as a place to spend time.

3.5.5 Adoption and Twin Studies in the Nature vs. Nurture Debate

Adoption and twin studies can help make sense of the influence of genes and the environment. Studies of adult twins are used to investigate which traits are heritable. Identical twins share the same genotype, meaning their genetic makeup is the same. Identical twins raised apart tend to be similar in intelligence and, in some cases, life events and circumstance, when studied years later, even when raised separately. However, researchers have discovered that the phenotype (or the observable expression of a gene) of identical twins grows apart as they age.

In adoption studies, identical twins raised by different families can give insight into the nature-versus-nurture debate. Since the child is being raised by parents who are genetically different from his or her biological parents, the influence of the environment shows in how similar the child is to his or her adoptive parents or adoptive siblings vs. how similar the child is to his or her biological parents and siblings. Adoption studies and twin studies make a strong case for genetic influence. However, it is now generally accepted among scientists, that all traits are influenced by both genetic and environmental differences among individuals.

3.5.6 KEY TAKEAWAYS

3.5.6.1 Key Points

- Chromosomes are structures in the nucleus of a cell containing DNA coiled around histone proteins.
- All animals have some number of chromosomes, which transmit genetic material. Human beings have 46 chromosomes (23 pairs).
- Humans have two types of chromosomes: autosomes and sex chromosomes.
- Chromosomal abnormalities can result in genetic conditions such as Down syndrome.
- Today it is generally accepted that nature and nurture work in tandem to create the people we ultimately become.
- Adoption and twin studies show that both nature and nurture are factors in human development.
- The environment in which a person is raised can trigger expressions of behavior for which that person is genetically predisposed; genetically identical people raised in different environments may exhibit different behavior.
- Three types of gene -environment correlations (rGE) exist: passive (ambiguous correlation), evocative (one factor invokes a response in the other), and active (one factor influences a preference for another).

3.5.6.2 Key Terms

- **chromosome**: A structure in the cell nucleus that contains DNA, histone protein, and other structural proteins.
- **gene**: A unit of heredity; a segment of DNA or RNA transmitted from one generation to the next, carrying genetic information such as the sequence of amino acids for a protein.
- **autosome**: Any chromosome that is not a sex chromosome.
- **gene-environment correlation**: A relationship in which exposure to environmental conditions correlates with an individual's genotype.
- **phenotype**: The observable expression of a gene.

3.5.7 Attributions

Adapted by Kenneth A. Koenigshofer, Ph.D., Chaffey College, from Genetics and Behavior by Boundless.com. **License**: *CC BY-SA: Attribution-ShareAlike*

3.5.8

This page titled 3.5: DNA, Chromosomes, and Gene-Environment Correlations is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

3.6: The Nature-Nurture Question

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Discuss the nature–nurture debate and why the problem fascinates us.
2. Explain why nature–nurture questions are difficult to study empirically.
3. Describe the major research designs that can be used to study nature–nurture questions.
4. Discuss the complexities of nature–nurture and why questions that seem simple turn out not to have simple answers.
5. Explain the main claim of sociobiology and the reasons why it is controversial.

3.6.1 **Overview**

People have a deep intuition about what has been called the “nature–nurture question.” Some aspects of our behavior feel as though they originate in our genetic makeup, while others feel like the result of our upbringing or our own hard work. The scientific field of behavior genetics attempts to study these differences empirically, either by examining similarities among family members with different degrees of genetic relatedness, or, more recently, by studying differences in the DNA of people with different behavioral traits. The scientific methods that have been developed are ingenious, but often inconclusive. Many of the difficulties encountered in the empirical science of behavior genetics turn out to be conceptual, and our intuitions about nature and nurture get more complicated the harder we think about them. In the end, it is an oversimplification to ask how “genetic” some particular behavior is. Genes and environments always combine to produce behavior, and the real science is in the discovery of how they combine for a given behavior.

3.6.2 Genes and Environment

It may seem obvious that we are born with certain characteristics while others are acquired, and yet in the history of psychology, no other question has caused so much controversy and offense: We are so concerned with nature–nurture because our very sense of moral character seems to depend on it. While we may admire the athletic skills of a great basketball player, we think of his height as simply a gift, a payoff in the “genetic lottery.” For the same reason, no one blames a short person for his height or someone’s congenital disability on poor decisions: To state the obvious, it’s “not their fault.” But we do praise the concert violinist (and perhaps her parents and teachers as well) for her dedication, just as we condemn cheaters, slackers, and bullies for their bad behavior.

The problem is, most human characteristics aren’t usually as clear-cut as height or instrument-mastery, affirming our nature–nurture expectations strongly one way or the other. In fact, even the great violinist might have some inborn qualities—perfect pitch, or long, nimble fingers—that support and reward her hard work. And the basketball player might have eaten a diet while growing up that promoted his genetic tendency for being tall. When we think about our own qualities, they seem under our control in some respects, yet beyond our control in others. And often the traits that don’t seem to have an obvious cause are the ones that concern us the most and are far more personally significant. What about how much we drink or worry? What about our honesty, or religiosity, or sexual orientation? They all come from that uncertain zone, neither fixed by nature nor totally under our own control.



Figure **3.6.1**: Researchers have learned a great deal about the nature-nurture dynamic by working with animals. But of course many of the techniques used to study animals cannot be applied to people. Separating these two influences in human subjects is a greater research challenge. [Image: Sebastián Dario, https://goo.gl/OPiIWd, CC BY-NC 2.0, https://goo.gl/FIlc2e]

One major problem with answering nature-nurture questions about people is, how do you set up an experiment? In nonhuman animals, there are relatively straightforward experiments for tackling nature–nurture questions. Say, for example, you are interested in aggressiveness in dogs. You want to test for the more important determinant of aggression: being born to aggressive dogs or being raised by them. You could mate two aggressive dogs—angry Chihuahuas—together, and mate two nonaggressive dogs—happy beagles—together, then switch half the puppies from each litter between the different sets of parents to raise. You would then have puppies born to aggressive parents (the Chihuahuas) but being raised by nonaggressive parents (the Beagles), and vice versa, in litters that mirror each other in puppy distribution. The big questions are: Would the Chihuahua parents raise aggressive beagle puppies? Would the beagle parents raise nonaggressive Chihuahua puppies? Would the puppies' nature win out, regardless of who raised them? Or... would the result be a combination of nature and nurture? Much of the most significant nature–nurture research has been done in this way (Scott & Fuller, 1998), and animal breeders have been doing it successfully for thousands of years. In fact, it is fairly easy to breed animals for behavioral traits.

For example, Koenigshofer and Nachman (1974) selectively bred rats for their ability to learn taste aversions. When a novel flavor such as sugar water is followed by mild poisoning by injection 15 minutes later, rats learn to avoid that flavor in the future after a single pairing of flavor and illness (see section 10.2, Specialized Forms of Learning). However, rats varied in how well they learned aversion to the sugar water. A few rats when tested 4 days later showed that they had learned a strong aversion (they consumed very little of the flavor that had been previously followed by poison-induced illness) while a few others learned a weak aversion (they consumed a large amount of the flavor previously paired with illness), while most of the conditioned rats learned a moderate aversion to the flavor previously paired with illness (they consumed an intermediate amount of the flavor previously paired with poison-induced illness). After this first phase of the experiment, rats were selectively bred. Those males and females that learned a strong aversion were bred together, while male and female rats that learned a weak aversion were bred together. Those that learned an intermediate aversion were not bred. This process, involving the pairing of taste and illness followed by selective breeding for strong or weak learning, was repeated for several generations. After just five generations of this selective breeding procedure for taste aversion learning, two strains of rats resulted, strong learners and weak learners, which showed no overlap at all in the strength of taste aversion that they learned after just five generations of selective breeding. The worst learners of the strong learner strain were still much better learners than the best learners of the weak learner strain. The effect of selective breeding in these experiments was greater than the effects of brain lesions on learning ability. This suggests that learning abilities have a significant genetic component--a proposition that is consistent with the hypothesis that at least some differences among humans in learning ability involve genetic differences. That learning abilities have a genetic components suggests that learning abilities have evolved over time by selection (i.e. natural selection). If learning is an evolved genetic trait, how can any behavioral trait, even those that are learned, not have a genetic component?

Although we can experiment with animals, with people, however, we can't assign babies to parents at random, or select parents with certain behavioral characteristics to mate, merely in the interest of science (though history does include horrific examples of such practices, in misguided attempts at "eugenics," the shaping of human characteristics through intentional breeding). In typical human families, children's biological parents raise them, so it is very difficult to know whether children act like their parents due to

genetic (nature) or environmental (nurture) reasons. Nevertheless, despite our restrictions on setting up human-based experiments, we do see real-world examples of nature-nurture at work in the human sphere—though they only provide partial answers to our questions.

The science of how genes and environments work together to influence behavior is called behavioral genetics. The easiest opportunity we have to observe this is the **adoption study**. When children are put up for adoption, the parents who give birth to them are no longer the parents who raise them. This setup isn't quite the same as the experiments with dogs (children aren't assigned to random adoptive parents in order to suit the particular interests of a scientist) but adoption still tells us some interesting things, or at least confirms some basic expectations. For instance, if the biological child of tall parents were adopted into a family of short people, do you suppose the child's growth would be affected? What about the biological child of a Spanish-speaking family adopted at birth into an English-speaking family? What language would you expect the child to speak? And what might these outcomes tell you about the difference between height and language in terms of nature-nurture?



Figure **3.6.2**: Studies focused on twins have led to important insights about the biological origins of many personality characteristics.

Another option for observing nature-nurture in humans involves **twin studies**. There are two types of twins: **monozygotic (MZ)** and **dizygotic (DZ)**. Monozygotic twins, also called "identical" twins, result from a single zygote (fertilized egg) and have the same DNA. They are essentially clones. Dizygotic twins, also known as "fraternal" twins, develop from two zygotes (fertilized eggs) and share 50% of their DNA. Fraternal twins are ordinary siblings who happen to have been born at the same time. To analyze nature–nurture using twins, we compare the similarity of MZ and DZ pairs. Sticking with the features of height and spoken language, let's take a look at how nature and nurture apply: Identical twins, unsurprisingly, are almost perfectly similar for height. The heights of fraternal twins, however, are like any other sibling pairs: more similar to each other than to people from other families, but hardly identical. This contrast between twin types gives us a clue about the role genetics plays in determining height. Now consider spoken language. If one identical twin speaks Spanish at home, the co-twin with whom she is raised almost certainly does too. But the same would be true for a pair of fraternal twins raised together. In terms of spoken language, fraternal twins are just as similar as identical twins, so it appears that the genetic match of identical twins doesn't make much difference. What language you speak is determined by environment; having more genes in common (MZ vs DZ) doesn't make any difference.

Twin and adoption studies are two instances of a much broader class of methods for observing nature-nurture called **quantitative genetics** , the scientific discipline in which similarities among individuals are analyzed based on how biologically related they are.

We can do these studies with siblings and half-siblings, cousins, twins who have been separated at birth and raised separately (Bouchard, Lykken, McGue, & Segal, 1990 ; such twins are very rare and play a smaller role than is commonly believed in the science of nature–nurture), or with entire extended families (see Plomin, DeFries, Knopik, & Neiderhiser, 2012 , for a complete introduction to research methods relevant to nature–nurture).

For better or for worse, contentions about nature–nurture have intensified because quantitative genetics produces a number called a **heritability coefficient** , varying from 0 to 1, that is meant to provide a single measure of genetics' influence on a trait. In a general way, a heritability coefficient **measures how strongly differences among individuals are related to differences among their genes**. But beware: Heritability coefficients, although simple to compute, are deceptively difficult to interpret. Nevertheless, numbers that provide simple answers to complicated questions tend to have a strong influence on the human imagination, and a great deal of time has been spent discussing whether the heritability of intelligence or personality or depression is equal to one number or another.

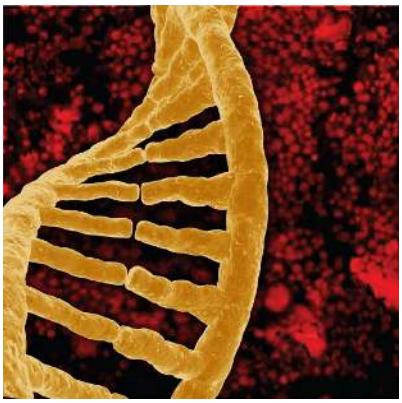


Figure **3.6.3**: Artist's depiction of double helix of DNA. (Copyright; author via source)

One reason nature–nurture continues to fascinate us so much is that we live in an era of great scientific discovery in genetics, comparable to the times of Copernicus, Galileo, and Newton, with regard to astronomy and physics. Every day, it seems, new discoveries are made, new possibilities proposed. When Francis Galton first started thinking about nature–nurture in the late-19th century he was very influenced by his cousin, Charles Darwin, but genetics per se was unknown. Mendel's famous work with peas, conducted at about the same time, went undiscovered for 20 years; quantitative genetics was developed in the 1920s; DNA was discovered by Watson and Crick in the 1950s; the human genome was completely sequenced at the turn of the 21st century; and we are now on the verge of being able to obtain the specific DNA sequence of anyone at a relatively low cost. No one knows what this new genetic knowledge will mean for the study of nature–nurture, but as we will see in the next section, answers to nature–nurture questions have turned out to be far more difficult and mysterious than anyone imagined.

3.6.3 What have we learned about nature-nurture?

It would be satisfying to be able to say that nature–nurture studies have given us conclusive and complete evidence about where traits come from, with some traits clearly resulting from genetics and others almost entirely from environmental factors, such as childrearing practices and personal will; but that is not the case. Instead, **everything has turned out to have some footing in genetics**. The more genetically-related people are, the more similar they are—for everything: height, weight, intelligence, personality, mental illness, etc. Sure, it seems like common sense that some traits have a genetic bias. For example, adopted children resemble their biological parents even if they have never met them, and identical twins are more similar to each other than are fraternal twins. And while certain psychological traits, such as personality or mental illness (e.g., schizophrenia), seem reasonably influenced by genetics, it turns out that the same is true for political attitudes, how much television people watch (Plomin, Corley, DeFries, & Fulker, 1990), and whether or not they get divorced (McGue & Lykken, 1992).



Figure **3.6.4**: Research over the last half century has revealed how central genetics are to behavior. The more genetically related people are the more similar they are not just physically but also in terms of personality and behavior. [Image: Paul Altobelli, https://goo.gl/SWLwm2, CC BY 2.0, https://goo.gl/9uSnqN]

It may seem surprising, but genetic influence on behavior is a relatively recent discovery. In the middle of the 20th century, psychology was dominated by the doctrine of behaviorism, which held that behavior could only be explained in terms of environmental factors. Psychiatry concentrated on psychoanalysis, which probed for roots of behavior in individuals' early life-histories. The truth is, neither behaviorism nor psychoanalysis is incompatible with genetic influences on behavior, and neither Freud nor Skinner was naive about the importance of organic processes in behavior. Nevertheless, in their day it was widely thought that children's personalities were shaped entirely by imitating their parents' behavior, and that schizophrenia was caused by certain kinds of "pathological mothering." Whatever the outcome of our broader discussion of nature–nurture, the basic fact that the best predictors of an adopted child's personality or mental health are found in the biological parents he or she has never met, rather than in the adoptive parents who raised him or her, presents a significant challenge to purely environmental explanations of personality or psychopathology. The message is clear: **You can't leave genes out of the equation. But keep in mind, no behavioral traits are completely inherited, so you can't leave the environment out altogether, either.**

Trying to untangle the various ways nature-nurture influences human behavior can be messy, and often common-sense notions can get in the way of good science. One very significant contribution of behavioral genetics that has changed psychology for good can be very helpful to keep in mind: When your subjects are biologically-related, no matter how clearly a situation may seem to point to environmental influence, it is never safe to interpret a behavior as wholly the result of nurture without further evidence. For example, when presented with data showing that children whose mothers read to them often are likely to have better reading scores in third grade, it is tempting to conclude that reading to your kids out loud is important to success in school; this may well be true, but the study as described is inconclusive, because there are genetic as well as environmental pathways between the parenting practices of mothers and the abilities of their children. This is a case where "correlation does not imply causation," as they say. To establish that reading aloud causes success, a scientist can either study the problem in adoptive families (in which the genetic pathway is absent) or by finding a way to randomly assign children to oral reading conditions.

The outcomes of nature–nurture studies have fallen short of our expectations (of establishing clear-cut bases for traits) in many ways. The most disappointing outcome has been the inability to organize traits from more-to less-genetic. As noted earlier, **everything has turned out to be at least somewhat heritable (passed down), yet nothing has turned out to be absolutely heritable**, and there hasn't been much consistency as to which traits are more heritable and which are less heritable once other considerations (such as how accurately the trait can be measured) are taken into account (Turkheimer, 2000). The problem is

conceptual: The heritability coefficient, and, in fact, the whole quantitative structure that underlies it, does not match up with our nature–nurture intuitions. We want to know how "important" the roles of genes and environment are to the development of a trait, but in focusing on "important" maybe we're emphasizing the wrong thing. First of all, genes and environment are both crucial to every trait; without genes the environment would have nothing to work on, and too, genes cannot develop in a vacuum. Even more important, because nature–nurture questions look at the differences among people, the cause of a given trait depends not only on the trait itself, but also on the differences in that trait between members of the group being studied.

The classic example of the heritability coefficient defying intuition is the trait of having two arms. No one would argue against the development of arms being **a biological, genetic process**. But fraternal twins are just as similar for "two-armedness" as identical twins, resulting in a **heritability coefficient of zero** for the trait of having two arms. Normally, according to the heritability model, this result (coefficient of zero) would suggest all nurture, no nature, but we know that's not the case. The reason this result is not a tip-off that arm development is less genetic than we imagine is **because people do not vary in the genes related to arm development**—which essentially upends the heritability formula. In fact, in this instance, the opposite is likely true: the extent that people differ in arm number is likely the result of accidents and, therefore, environmental. For reasons like these, we always have to be very careful when asking nature–nurture questions, especially when we try to express the answer in terms of a single number. The heritability of a trait is not simply a property of that trait, but a property of the trait in a particular context of relevant genes and environmental factors.

Another issue with the heritability coefficient is that it divides traits' determinants into two portions—genes and environment—which are then calculated together for the total variability. This is a little like asking how much of the experience of a symphony comes from the horns and how much from the strings; the ways instruments or genes integrate is more complex than that. It turns out to be the case that, **for many traits, genetic differences affect behavior under some environmental circumstances but not others—a phenomenon called gene-environment interaction**, or G x E. In one well-known example, Caspi et al. (2002) showed that among maltreated children, those who carried a particular allele of the MAOA gene showed a predisposition to violence and antisocial behavior, while those with other alleles did not. Whereas, in children who had not been maltreated, the gene had no effect. Making matters **even more complicated** are very recent studies of what is known as **epigenetics** (see module, "Epigenetics" http://noba.to/37p5cb8v), a process **in which the DNA itself is modified by environmental events, and those genetic changes transmitted to children.**



Figure **3.6.5**: The answer to the nature–nurture question has not turned out to be as straightforward as we would like. The many questions we can ask about the relationships among genes, environments, and human traits may have many different answers, and the answer to one tells us little about the answers to the others. [Image and caption: Sundaram Ramaswamy, https://goo.gl/Bv8lp6, CC BY 2.0, https://goo.gl/9uSnqN]

Some common questions about nature–nurture are, how susceptible is a trait to change, how malleable is it, and do we "have a choice" about it? These questions are much more complex than they may seem at first glance. For example, phenylketonuria is an inborn error of metabolism caused by a single gene; it prevents the body from metabolizing phenylalanine. Untreated, it causes intellectual disability and death. But it can be treated effectively by a straightforward environmental intervention: avoiding foods containing phenylalanine. Height seems like a trait firmly rooted in our nature and unchangeable, but the average height of many populations in Asia and Europe has increased significantly in the past 100 years, due to changes in diet and the alleviation of

poverty. Even the most modern genetics has not provided definitive answers to nature–nurture questions. When it was first becoming possible to measure the DNA sequences of individual people, it was widely thought that we would quickly progress to finding the specific genes that account for behavioral characteristics, but that hasn't happened. There are a few rare genes that have been found to have significant (almost always negative) effects, such as the single gene that causes Huntington's disease, or the Apolipoprotein gene that causes early onset dementia in a small percentage of Alzheimer's cases. Aside from these rare genes of great effect, however, the genetic impact on behavior is broken up over many genes, each with very small effects. For most behavioral traits, the effects are so small and distributed across so many genes that we have not been able to catalog them in a meaningful way. In fact, the same is true of environmental effects. We know that extreme environmental hardship causes catastrophic effects for many behavioral outcomes, but fortunately extreme environmental hardship is very rare. Within the normal range of environmental events, those responsible for differences (e.g., why some children in a suburban third-grade classroom perform better than others) are much more difficult to grasp.

The difficulties with finding clear-cut solutions to nature–nurture problems bring us back to the other great questions about our relationship with the natural world: the mind-body problem and free will. Investigations into what we mean when we say we are aware of something reveal that consciousness is not simply the product of a particular area of the brain, nor does choice turn out to be an orderly activity that we can apply to some behaviors but not others. So it is with nature and nurture: What at first may seem to be a straightforward matter, able to be indexed with a single number, becomes more and more complicated the closer we look. The many questions we can ask about the intersection among genes, environments, and human traits—how sensitive are traits to environmental change, and how common are those influential environments; are parents or culture more relevant; how sensitive are traits to differences in genes, and how much do the relevant genes vary in a particular population; does the trait involve a single gene or a great many genes; is the trait more easily described in genetic or more-complex behavioral terms?—may have different answers, and the answer to one tells us little about the answers to the others.

It is tempting to predict that the more we understand the wide-ranging effects of genetic differences on all human characteristics—especially behavioral ones—our cultural, ethical, legal, and personal ways of thinking about ourselves will have to undergo profound changes in response. Perhaps criminal proceedings will consider genetic background. Parents, presented with the genetic sequence of their children, will be faced with difficult decisions about reproduction. These hopes or fears are often exaggerated. In some ways, our thinking may need to change—for example, when we consider the meaning behind the fundamental American principle that all men are created equal. Human beings differ, and like all evolved organisms they differ genetically. The Declaration of Independence predates Darwin and Mendel, but it is hard to imagine that Jefferson—whose genius encompassed botany as well as moral philosophy—would have been alarmed to learn about the genetic diversity of organisms. One of the most important things modern genetics has taught us is that almost all human behavior is too complex to be nailed down, even from the most complete genetic information, unless we're looking at identical twins. The science of nature and nurture has demonstrated that genetic differences among people are vital to human moral equality, freedom, and self-determination, not opposed to them. As Mordecai Kaplan said about the role of the past in Jewish theology, genetics gets a vote, not a veto, in the determination of human behavior. We should indulge our fascination with nature–nurture while resisting the temptation to oversimplify it.

3.6.4 Outside Resources

Web: Institute for Behavioral Genetics

http://www.colorado.edu/ibg/

3.6.5 Vocabulary

Adoption study: **A behavior genetic research method that involves comparison of adopted children to their adoptive and biological parents.**

***Behavioral genetics*: The empirical science of how genes and environments combine to generate behavior.**

***Heritability coefficient*: An easily misinterpreted statistical construct that purports to measure the percentage of differences among individuals on a trait due to genetics.**

***Quantitative genetics*: Scientific and mathematical methods for inferring genetic and environmental processes based on the degree of genetic and environmental similarity among organisms.**

***Twin studies*: A behavior genetic research method that involves comparison of the similarity of identical (monozygotic; MZ) and fraternal (dizygotic; DZ) twins.**

3.6.6 Discussion Questions

1. Is your personality more like one of your parents than the other? If you have a sibling, is his or her personality like yours? In your family, how did these similarities and differences develop? What do you think caused them?
2. Can you think of a human characteristic for which genetic differences would play almost no role? Defend your choice.
3. Do you think the time will come when we will be able to predict almost everything about someone by examining their DNA on the day they are born?
4. Identical twins are more similar than fraternal twins for the trait of aggressiveness, as well as for criminal behavior. Do these facts have implications for the courtroom? If it can be shown that a violent criminal had violent parents, should it make a difference in culpability or sentencing?

3.6.7 **Sociobiology**

Sociobiology is an interdisciplinary science originally popularized by social insect researcher E.O. Wilson in the 1970s. Wilson defined the science as "the extension of population biology and evolutionary theory to social organization." The main thrust of sociobiology is that animal and human behavior, including aggressiveness and other social interactions, can be explained almost solely in terms of genetics and natural selection (Wilson, 1975).

This science is controversial; some have criticized the approach for ignoring the environmental effects on behavior. This is another example of the "nature versus nurture" debate of the role of genetics versus the role of environment in determining an organism's characteristics.

Sociobiology also links genes with behaviors and has been associated with "biological determinism," the belief that all behaviors are hardwired into our genes. No one disputes that certain behaviors can be inherited and that natural selection plays a role in their organization. However, it is the application of such principles to human behavior that ruffles feathers and sparks this controversy, which remains active today.

3.6.8 **Key Points**

- Sociobiology argues that all animal and human behavior, including aggressiveness and other social interactions, can be explained almost solely in terms of genetics and natural selection.
- Sociobiology is controversial: some have criticized the approach for ignoring the environmental effects on behavior and for being similar to "biological determinism," or the belief that all behaviors are hardwired into our genes.

3.6.9 **Key Terms**

- **biological determinism**: also known as genetic determinism is the belief that most human traits, physical and psychological, are innate and determined by genes.
- **sociobiology**: the science that applies the principles of evolutionary biology to the study of social behaviour in both humans and animals, suggesting that social behavior in animals and humans is shaped primarily by genes and genetic evolution.

3.6.10 References

Bouchard, T. J., Lykken, D. T., McGue, M., & Segal, N. L. (1990). Sources of human psychological differences: The Minnesota study of twins reared apart. Science, 250(4978), 223–228.

Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., Taylor, A. & Poulton, R. (2002). Role of genotype in the cycle of violence in maltreated children. Science, 297(5582), 851–854.

Koenigshofer, K. A. & Nachman, M. (1974). Selective Breeding for Taste Aversion Learning in Rats. Paper presented at the 54th Annual Convention of the Western Psychological Association. April 28, 1974. San Francisco.

McGue, M., & Lykken, D. T. (1992). Genetic influence on risk of divorce. Psychological Science, 3(6), 368–373.

Plomin, R., Corley, R., DeFries, J. C., & Fulker, D. W. (1990). Individual differences in television viewing in early childhood: Nature as well as nurture. Psychological Science, 1(6), 371–377.

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2012). Behavioral genetics. New York, NY: Worth Publishers.

Scott, J. P., & Fuller, J. L. (1998). Genetics and the social behavior of the dog. Chicago, IL: University of Chicago Press.

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. Current Directions in Psychological Science, 9(5), 160–164.

Wilson, E. O. (1975). *Sociobiology: The New Synthesis*. Cambridge, MA: Belknap Press.

3.6.11 Attribution

The Nature-Nurture Question by Eric Turkheimer is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Sociobiology adapted by Kenneth A. Koenigshofer, Ph.D., Chaffey College, from Sociobiology. Provided by: Wiktionary. Located at: en.wiktionary.org/wiki/sociobiology. License: ***CC BY-SA: Attribution-ShareAlike***

LICENSES AND ATTRIBUTIONS

CC LICENSED CONTENT, SHARED PREVIOUSLY

- Curation and Revision. **Provided by**: Boundless.com. **License**: ***CC BY-SA: Attribution-ShareAlike***

CC LICENSED CONTENT, SPECIFIC ATTRIBUTION

- OpenStax College, Biology. October 17, 2013. **Provided by**: OpenStax CNX. **Located at**: **http://cnx.org/content/m44879/latest...ol11448/latest**. **License**: ***CC BY: Attribution***
- OpenStax College, Biology. October 23, 2013. **Provided by**: OpenStax CNX. **Located at**: **http://cnx.org/content/m44879/latest...ol11448/latest**. **License**: ***CC BY: Attribution***

3.6.12

This page titled 3.6: The Nature-Nurture Question is shared under a mixed license and was authored, remixed, and/or curated by Keneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

3.7: Epigenetics in Psychology

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Explain what the term epigenetics means and the molecular machinery involved.
2. Describe important neural and developmental pathways that are regulated by epigenetic factors, and discuss examples of epigenetic effects on personality traits and cognitive behavior.
3. Explain how misregulation of epigenetic mechanisms can lead to disease states, and discuss examples.
4. Describe how epigenetic machinery can be targets for therapeutic agents, and discuss examples.
5. Discuss examples of genetic diseases and their patterns of inheritance.

3.7.1 Overview

Early life experiences exert a profound and long-lasting influence on physical and mental health throughout life. The efforts to identify the primary causes of this have significantly benefited from studies of the epigenome—a dynamic layer of information associated with DNA that differs between individuals and can be altered through various experiences and environments. The epigenome has been heralded as a key “missing piece” of the etiological puzzle for understanding how development of psychological disorders may be influenced by the surrounding environment, in concordance with the genome. Understanding the mechanisms involved in the initiation, maintenance, and heritability of epigenetic states is thus an important aspect of research in current biology, particularly in the study of learning and memory, emotion, and social behavior in humans. Moreover, epigenetics in psychology provides a framework for understanding how the expression of genes is influenced by experiences and the environment to produce individual differences in behavior, cognition, personality, and mental health. In this module, we survey recent developments revealing epigenetic aspects of mental health and review some of the challenges of epigenetic approaches in psychology to help explain how nurture shapes nature.

3.7.2 Introduction



Figure **3.7.1**: DNA stands for Deoxyribonucleic Acid, and although each person’s DNA is unique to that individual, it is 99.9% similar to every other human on the planet. [Image: CC0 Public Domain, https://goo.gl/m25gce]

Early childhood is not only a period of physical growth; it is also a time of mental development related to changes in the anatomy, physiology, and chemistry of the nervous system that influence mental health throughout life. Cognitive abilities associated with learning and memory, reasoning, problem solving, and developing relationships continue to emerge during childhood. Brain development is more rapid during this critical or sensitive period than at any other, with more than 700 neural connections created each second. Herein, complex gene –environment interactions (or genotype–environment interactions, G×E) serve to increase the number of possible contacts between neurons, as they hone their adult synaptic properties and excitability. Many weak connections

form to different neuronal targets; subsequently, they undergo remodeling in which most connections vanish and a few stable connections remain. These structural changes (or plasticity) may be crucial for the development of mature neural networks that support emotional, cognitive, and social behavior. The generation of different morphology, physiology, and behavioral outcomes from a single genome in response to changes in the environment forms the basis for “phenotypic plasticity,” which is fundamental to the way organisms cope with environmental variation, navigate the present world, and solve future problems.

The challenge for psychology has been to integrate findings from genetics and environmental (social, biological, chemical) factors, including the quality of infant–mother attachments, into the study of personality and our understanding of the emergence of mental illness. These studies have demonstrated that common DNA sequence variation and rare mutations account for only a small fraction (1%–2%) of the total risk for inheritance of personality traits and mental disorders (Dick, Riley, & Kendler, 201 0; Gershon, Alliey-Rodriguez, & Liu, 2011). Additionally, studies that have attempted to examine the mechanisms and conditions under which DNA sequence variation influences brain development and function have been confounded by complex cause-and-effect relationships (Petronis, 2010). The large unaccounted heritability of personality traits and mental health suggests that additional molecular and cellular mechanisms are involved.

Epigenetics has the potential to provide answers to these important questions and refers to the transmission of phenotype in terms of gene expression in the absence of changes in DNA sequence—hence the name epi- (Greek: επί- over, above) genetics (Waddington, 1942 ; Wolffe & Matzke, 1999). The advent of high-throughput techniques such as sequencing-based approaches to study the distributions of regulators of gene expression throughout the genome led to the collective description of the “epigenome.” In contrast to the genome sequence, which is static and the same in almost all cells, the epigenome is highly dynamic, differing among cell types, tissues, and brain regions (Gregg et al., 2010). Recent studies have provided insights into epigenetic regulation of developmental pathways in response to a range of external environmental factors (Dolinoy, Weidman, & Jirtle, 2007). These environmental factors during early childhood and adolescence can cause changes in expression of genes conferring risk of mental health and chronic physical conditions. Thus, the examination of genetic–epigenetic–environment interactions from a developmental perspective may determine the nature of gene misregulation in psychological disorders.

This module will provide an overview of the main components of the epigenome and review themes in recent epigenetic research that have relevance for psychology, to form the biological basis for the interplay between environmental signals and the genome in the regulation of individual differences in physiology, emotion, cognition, and behavior.

3.7.3 Molecular control of gene expression: the dynamic epigenome



Figure **3.7.2**: Identical twins are the perfect example of epigenetics. Although they share exactly the same DNA, their unique experiences in life will cause some genes (and not others) to express themselves. This is why, over time, identical twins come to look and behave differently. [Image: M., https://goo.gl/VU5iJv, CC BY-NC-SA 2.0, https://goo.gl/Toc0ZF]

Almost all the cells in our body are genetically identical, yet our body generates many different cell types, organized into different tissues and organs, and expresses different proteins. Within each type of mammalian cell, about 2 meters of genomic DNA is divided into nuclear chromosomes. Yet the nucleus of a human cell, which contains the chromosomes, is only about 2 μm in diameter. To achieve this 1,000,000-fold compaction, DNA is wrapped around a group of 8 proteins called histones. This combination of DNA and histone proteins forms a special structure called a “nucleosome,” the basic unit of chromatin, which represents a structural solution for maintaining and accessing the tightly compacted genome. These factors alter the likelihood that

a gene will be expressed or silenced. Cellular functions such as gene expression, DNA replication, and the generation of specific cell types are therefore influenced by distinct patterns of chromatin structure, involving covalent modification of both histones (Kadonaga, 1998) and DNA (Razin, 1998).

Importantly, epigenetic variation also emerges across the lifespan. For example, although identical twins share a common genotype and are genetically identical and epigenetically similar when they are young, as they age they become more dissimilar in their epigenetic patterns and often display behavioral, personality, or even physical differences, and have different risk levels for serious illness. Thus, understanding the structure of the nucleosome is key to understanding the precise and stable control of gene expression and regulation, providing a molecular interface between genes and environmentally induced changes in cellular activity.

3.7.4 The primary epigenetic mark: DNA modification

DNA methylation is the best-understood epigenetic modification influencing gene expression. DNA is composed of four types of naturally occurring nitrogenous bases: adenine (A), thymine (T), guanine (G), and cytosine (C). In mammalian genomes, DNA methylation occurs primarily at cytosine residues in the context of cytosines that are followed by guanines (CpG dinucleotides), to form 5-methylcytosine in a cell-specific pattern (Goll & Bestor, 2005 ; Law & Jacobsen, 2010 ; Suzuki & Bird, 2008). The enzymes that perform DNA methylation are called DNA methyltransferases (DNMTs) , which catalyze the transfer of a methyl group to the cytosine (Adams, McKay, Craig, & Burdon, 1979). These enzymes are all expressed in the central nervous system and are dynamically regulated during development (Feng, Chang, Li, & Fan, 2005 ; Goto et al., 1994). The effect of DNA methylation on gene function varies depending on the period of development during which the methylation occurs and location of the methylated cytosine. Methylation of DNA in gene regulatory regions (promoter and enhancer regions) usually results in gene silencing and reduced gene expression (Ooi, O'Donnell, & Bestor, 2009 ; Suzuki & Bird, 2008 ; Sutter and Doerfler, 1980 ; Vardimon et al., 1982). This is a powerful regulatory mechanism that ensures that genes are expressed only when needed. Thus DNA methylation may broadly impact human brain development, and age-related misregulation of DNA methylation is associated with the molecular pathogenesis of neurodevelopmental disorders.

3.7.5 Histone modification and the histone code

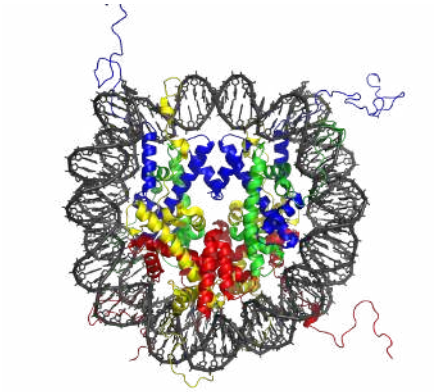


Figure **3.7.3**: Life experiences, like a stressful event in childhood, can cause the modification of histone proteins (pictured) to help adapt to one's environment. For example, in response to a stressful event, histone modification of one's DNA might occur to encourage a more cautions personality—in order to avoid future, stressful encounters. [Image: Zephyris, https://goo.gl/gGrSQd, CC BY-SA 3.0, https://goo.gl/kB1Ogc]

The modification of histone proteins comprises an important epigenetic mark related to gene expression. One of the most thoroughly studied modifications is histone acetylation, which is associated with gene activation and increased gene expression (Wade, Pruss, & Wolffe, 1997). Acetylation on histone tails is mediated by the opposing enzymatic activities of histone acetyltransferases (HATs) and histone deacetylases (HDACs) (Kuo & Allis, 1998). For example, acetylation of histone in gene regulatory regions by HAT enzymes is generally associated with DNA demethylation, gene activation, and increased gene expression (Hong, Schroth, Matthews, Yau, & Bradbury, 1993 ; Sealy & Chalkley, 1978). On the other hand, removal of the acetyl group (deacetylation) by HDAC enzymes is generally associated with DNA methylation, gene silencing, and decreased gene expression (Davie & Chadee, 1998). The relationship between patterns of histone modifications and gene activity provides evidence for the existence of a "histone code" for determining cell-specific gene expression programs (Jenuwein & Allis, 2001).

Interestingly, recent research using animal models has demonstrated that histone modifications and DNA methylation of certain genes mediates the long-term behavioral effects of the level of care experienced during infancy.

3.7.6 Early childhood experience

The development of an individual is an active process of adaptation that occurs within a social and economic context. For example, the closeness or degree of positive attachment of the parent (typically mother)–infant bond and parental investment (including nutrient supply provided by the parent) that define early childhood experience also program the development of individual differences in stress responses in the brain, which then affect memory, attention, and emotion. In terms of evolution, this process provides the offspring with the ability to physiologically adjust gene expression profiles contributing to the organization and function of neural circuits and molecular pathways that support (1) biological defensive systems for survival (e.g., stress resilience), (2) reproductive success to promote establishment and persistence in the present environment, and (3) adequate parenting in the next generation (Bradshaw, 1965).

3.7.7 Parental investment and programming of stress responses in the offspring

The most comprehensive study to date of variations in parental investment and epigenetic inheritance in mammals is that of the maternally transmitted responses to stress in rats. In rat pups, maternal nurturing (licking and grooming) during the first week of life is associated with long-term programming of individual differences in stress responsiveness, emotionality, cognitive performance, and reproductive behavior (Caldji et al., 1998 ; Francis, Diorio, Liu, & Meaney, 1999 ; Liu et al., 1997 ; Myers, Brunelli, Shair, Squire, & Hofer, 1989 ; Stern, 1997). In adulthood, the offspring of mothers that exhibit increased levels of pup licking and grooming over the first week of life show increased expression of the glucocorticoid receptor in the hippocampus (a brain structure associated with stress responsivity as well as learning and memory) and a lower hormonal response to stress compared with adult animals reared by low licking and grooming mothers (Francis et al., 1999 ; Liu et al., 1997). Moreover, rat pups that received low levels of maternal licking and grooming during the first week of life showed decreased histone acetylation and increased DNA methylation of a neuron-specific promoter of the glucocorticoid receptor gene (Weaver et al., 2004). The expression of this gene is then reduced, the number of glucocorticoid receptors in the brain is decreased, and the animals show a higher hormonal response to stress throughout their life. The effects of maternal care on stress hormone responses and behaviour in the offspring can be eliminated in adulthood by pharmacological treatment (HDAC inhibitor trichostatin A, TSA) or dietary amino acid supplementation (methyl donor L-methionine), treatments that influence histone acetylation, DNA methylation, and expression of the glucocorticoid receptor gene (Weaver et al., 2004 ; Weaver et al., 2005). This series of experiments shows that histone acetylation and DNA methylation of the glucocorticoid receptor gene promoter is a necessary link in the process leading to the long-term physiological and behavioral sequelae of poor maternal care. This points to a possible molecular target for treatments that may reverse or ameliorate the traces of childhood maltreatment.



Figure 3.7.4: Parental care during one's childhood has important and consequential effects on the development of an individual, effects that persist even into adulthood. [Image: The White Ribbon Alliance, https://goo.gl/KgY6N5, CC BY-NC-SA 2.0, https://goo.gl/Toc0ZF]

Several studies have attempted to determine to what extent the findings from model animals are transferable to humans. Examination of post-mortem brain tissue from healthy human subjects found that the human equivalent of the glucocorticoid

receptor gene promoter (NR3C1 exon 1F promoter) is also unique to the individual (Turner, Pelascini, Macedo, & Muller, 2008). A similar study examining newborns showed that methylation of the glucocorticoid receptor gene promoter maybe an early epigenetic marker of maternal mood and risk of increased hormonal responses to stress in infants 3 months of age (Oberlander et al., 2008). Although further studies are required to examine the functional consequence of this DNA methylation, these findings are consistent with our studies in the neonate and adult offspring of low licking and grooming mothers that show increased DNA methylation of the promoter of the glucocorticoid receptor gene, decreased glucocorticoid receptor gene expression, and increased hormonal responses to stress (Weaver et al., 2004). Examination of brain tissue from suicide victims found that the human glucocorticoid receptor gene promoter is also more methylated in the brains of individuals who had experienced maltreatment during childhood (McGowan et al., 2009). These finding suggests that DNA methylation mediates the effects of early environment in both rodents and humans and points to the possibility of new therapeutic approaches stemming from translational epigenetic research. Indeed, similar processes at comparable epigenetic labile regions could explain why the adult offspring of high and low licking/grooming mothers exhibit widespread differences in hippocampal gene expression and cognitive function (Weaver, Meaney, & Szyf, 2006).

However, this type of research is limited by the inaccessibility of human brain samples. The translational potential of this finding would be greatly enhanced if the relevant epigenetic modification can be measured in an accessible tissue. Examination of blood samples from adult patients with bipolar disorder, who also retrospectively reported on their experiences of childhood abuse and neglect, found that the degree of DNA methylation of the human glucocorticoid receptor gene promoter was strongly positively related to the reported experience of childhood maltreatment decades earlier. For a relationship between a molecular measure and reported historical exposure, the effects size is extraordinarily large. This opens a range of new possibilities: given the large effect size and consistency of this association, measurement of the GR promoter methylation may effectively become a blood test measuring the physiological traces left on the genome by early experiences. Although this blood test cannot replace current methods of diagnosis, this unique and addition information adds to our knowledge of how disease may arise and be manifested throughout life. Near-future research will examine whether this measure adds value over and above simple reporting of early adversities when it comes to predicting important outcomes, such as response to treatment or suicide.

3.7.8 Child nutrition and the epigenome



Figure **3.7.5**: Whether or not your parents knew the science behind it, telling you to eat your veggies as a kid really does make you healthier and stronger—at least your DNA, that is. [Image: U.S. Department of Agriculture, https://goo.gl/tpyYzA, CC BY 2.0, https://goo.gl/BRvSA7]

The old adage “you are what you eat” might be true on more than just a physical level: The food you choose (and even what your parents and grandparents chose) is reflected in your own personal development and risk for disease in adult life (Wells, 2003). Nutrients can reverse or change DNA methylation and histone modifications, thereby modifying the expression of critical genes associated with physiologic and pathologic processes, including embryonic development, aging, and carcinogenesis. It appears that nutrients can influence the epigenome either by directly inhibiting enzymes that catalyze DNA methylation or histone modifications, or by altering the availability of substrates necessary for those enzymatic reactions. For example, rat mothers fed a diet low in methyl group donors during pregnancy produce offspring with reduced DNMT-1 expression, decreased DNA

methylation, and increased histone acetylation at promoter regions of specific genes, including the glucocorticoid receptor, and increased gene expression in the liver of juvenile offspring (Lillycrop, Phillips, Jackson, Hanson, & Burdge, 2005) and adult offspring (Lillycrop et al., 2007). These data suggest that early life nutrition has the potential to influence epigenetic programming in the brain not only during early development but also in adult life, thereby modulating health throughout life. In this regard, nutritional epigenetics has been viewed as an attractive tool to prevent pediatric developmental diseases and cancer, as well as to delay aging-associated processes.

The best evidence relating to the impact of adverse environmental conditions development and health comes from studies of the children of women who were pregnant during two civilian famines of World War II: the Siege of Leningrad (1941–44) (Bateson, 2001) and the Dutch Hunger Winter (1944–1945) (Stanner et al., 1997). In the Netherlands famine, women who were previously well nourished were subjected to low caloric intake and associated environmental stressors. Women who endured the famine in the late stages of pregnancy gave birth to smaller babies (Lumey & Stein, 1997) and these children had an increased risk of insulin resistance later in life (Painter, Roseboom, & Bleker, 2005). In addition, offspring who were starved prenatally later experienced impaired glucose tolerance in adulthood, even when food was more abundant (Stanner et al., 1997). Famine exposure at various stages of gestation was associated with a wide range of risks such as increased obesity, higher rates of coronary heart disease, and lower birth weight (Lumey & Stein, 1997). Interestingly, when examined 60 years later, people exposed to famine prenatally showed reduced DNA methylation compared with their unexposed same-sex siblings (Heijmans et al., 2008).

3.7.9 Epigenetic regulation of learning and memory

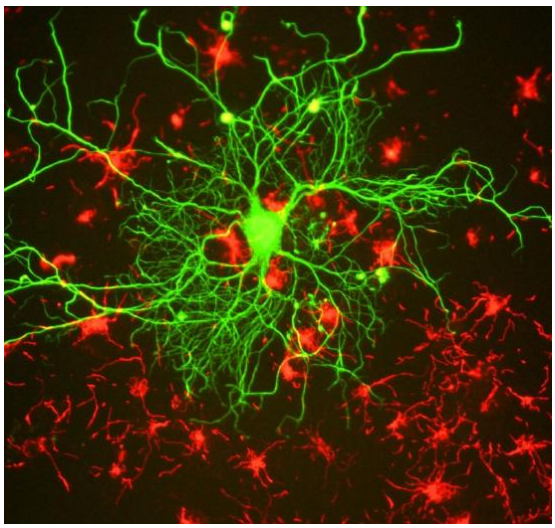


Figure 3.7.6: Neural plasticity is the change of neural pathways and synapses which allows for our ability to learn new things and remember them. [Image: Gerry Shaw, https://goo.gl/JBqlY7, CC BY-SA 3.0, https://goo.gl/eLCn2O]

Memories are recollections of actual events stored within our brains. But how is our brain able to form and store these memories? Epigenetic mechanisms influence genomic activities in the brain to produce long-term changes in synaptic signaling, organization, and morphology, which in turn support learning and memory (Day & Sweatt, 2011).

Neuronal activity in the hippocampus of mice is associated with changes in DNA methylation (Guo et al., 2011), and disruption to genes encoding the DNA methylation machinery cause learning and memory impairments (Feng et al., 2010). DNA methylation has also been implicated in the maintenance of long-term memories, as pharmacological inhibition of DNA methylation and impaired memory (Day & Sweatt, 2011 ; Miller et al., 2010). These findings indicate the importance of DNA methylation in mediating synaptic plasticity and cognitive functions, both of which are disturbed in psychological illness.

Changes in histone modifications can also influence long-term memory formation by altering chromatin accessibility and the expression of genes relevant to learning and memory. Memory formation and the associated enhancements in synaptic transmission are accompanied by increases in histone acetylation (Guan et al., 2002) and alterations in histone methylation (Schaefer et al., 2009), which promote gene expression. Conversely, a neuronal increase in histone deacetylase activity, which promotes gene silencing, results in reduced synaptic plasticity and impairs memory (Guan et al., 2009). Pharmacological inhibition of histone

deacetylases augments memory formation (Guan et al., 2009 ; Levenson et al., 2004), further suggesting that histone (de)acetylation regulates this process.

In humans genetic defects in genes encoding the DNA methylation and chromatin machinery exhibit profound effects on cognitive function and mental health (Jiang, Bressler, & Beaudet, 2004). The two best-characterized examples are Rett syndrome (Amir et al., 1999) and Rubinstein-Taybi syndrome (RTS) (Alarcon et al., 2004), which are profound intellectual disability disorders. Both MECP2 and CBP are highly expressed in neurons and are involved in regulating neural gene expression (Chen et al., 2003 ; Martinowich et al., 2003).

Rett syndrome patients have a mutation in their DNA sequence in a gene called MECP2. MECP2 plays many important roles within the cell: One of these roles is to read the DNA sequence, checking for DNA methylation, and to bind to areas that contain methylation, thereby preventing the wrong proteins from being present. Other roles for MECP2 include promoting the presence of particular, necessary, proteins, ensuring that DNA is packaged properly within the cell and assisting with the production of proteins. MECP2 function also influences gene expression that supports dendritic and synaptic development and hippocampus-dependent memory (Li, Zhong, Chau, Williams, & Chang, 2011 ; Skene et al., 2010). Mice with altered MECP2 expression exhibit genome-wide increases in histone acetylation, neuron cell death, increased anxiety, cognitive deficits, and social withdrawal (Shahbazian et al., 2002). These findings support a model in which DNA methylation and MECP2 constitute a cell-specific epigenetic mechanism for regulation of histone modification and gene expression, which may be disrupted in Rett syndrome.

RTS patients have a mutation in their DNA sequence in a gene called CBP. One of these roles of CBP is to bind to specific histones and promote histone acetylation, thereby promoting gene expression. Consistent with this function, RTS patients exhibit a genome-wide decrease in histone acetylation and cognitive dysfunction in adulthood (Kalkhoven et al., 2003). The learning and memory deficits are attributed to disrupted neural plasticity (Korzus, Rosenfeld, & Mayford, 2004). Similar to RTS in humans, mice with a mutation of CBP perform poorly in cognitive tasks and show decreased genome-wide histone acetylation (for review, see Josselyn, 2005). In the mouse brain CBP was found to act as an epigenetic switch to promote the birth of new neurons in the brain. Interestingly, this epigenetic mechanism is disrupted in the fetal brains of mice with a mutation of CBP, which, as pups, exhibit early behavioral deficits following removal and separation from their mother (Wang et al., 2010). These findings provide a novel mechanism whereby environmental cues, acting through histone modifying enzymes, can regulate epigenetic status and thereby directly promote neurogenesis, which regulates neurobehavioral development.

Together, these studies demonstrate that misregulation of epigenetic modifications and their regulatory enzymes is capable of orchestrating prominent deficits in neuronal plasticity and cognitive function. Knowledge from these studies may provide greater insight into other mental disorders such as depression and suicidal behaviors.

3.7.10 Epigenetic mechanisms in psychological disorders



Figure **3.7.7**: Pictured above is a chromatin, the spiral-looking macromolecule involved in depression. [Image: Zephyris, https://goo.gl/6DBQ1g, CC BY-SA 3.0, https://goo.gl/eLCn2O]

Epigenome-wide studies have identified several dozen sites with DNA methylation alterations in genes involved in brain development and neurotransmitter pathways, which had previously been associated with mental illness (Mill et al., 2008). These disorders are complex and typically start at a young age and cause lifelong disability. Often, limited benefits from treatment make these diseases some of the most burdensome disorders for individuals, families, and society. It has become evident that the efforts

to identify the primary causes of complex psychiatric disorders may significantly benefit from studies linking environmental effects with changes observed within the individual cells.

Epigenetic events that alter chromatin structure to regulate programs of gene expression have been associated with depression-related behavior and action of antidepressant medications, with increasing evidence for similar mechanisms occurring in post-mortem brains of depressed individuals. In mice, social avoidance resulted in decreased expression of hippocampal genes important in mediating depressive responses (Tsankova et al., 2006). Similarly, chronic social defeat stress was found to decrease expression of genes implicated in normal emotion processing (Lutter et al., 2008). Consistent with these findings, levels of histone markers of increased gene expression were down regulated (reduction in gene expression) in human post-mortem brain samples from individuals with a history of clinical depression (Covington et al., 2009).

Administration of antidepressants increased histone markers of increased gene expression and reversed the gene repression induced by defeat stress (Lee, Wynder, Schmidt, McCafferty, & Shiekhattar, 2006 ; Tsankova et al., 2006 ; Wilkinson et al., 2009). These results provide support for the use of HDAC inhibitors against depression. Accordingly, several HDAC inhibitors have been found to exert antidepressant effects by each modifying distinct cellular targets (Cassel et al., 2006 ; Schroeder, Lin, Crusio, & Akbarian, 2007).

There is also increasing evidence that aberrant gene expression resulting from altered epigenetic regulation is associated with the pathophysiology of suicide (McGowan et al., 2008 ; Poulter et al., 2008). Thus, it is tempting to speculate that there is an epigenetically determined reduced capacity for gene expression, which is required for learning and memory, in the brains of suicide victims.

3.7.11 Epigenetic strategy to understanding gene-environment interactions



Figure 3.7.8: Although there is some evidence that a dysfunctional upbringing can increase one's likelihood for schizophrenia (an epigenetically inherited disease), some people who have both the predisposition and the stressful environment never develop the mental illness. [Image: Steve White, CC0 Public Domain, https://goo.gl/m25gce]

While the cellular and molecular mechanisms that influence on physical and mental health have long been a central focus of neuroscience, only in recent years has attention turned to the epigenetic mechanisms behind the dynamic changes in gene expression responsible for normal cognitive function and increased risk for mental illness. The links between early environment and epigenetic modifications suggest a mechanism underlying gene-environment interactions. Early environmental adversity alone is not a sufficient cause of mental illness, because many individuals with a history of severe childhood maltreatment or trauma remain healthy. It is increasingly becoming evident that inherited differences in the segments of specific genes may moderate the effects of adversity and determine who is sensitive and who is resilient through a gene-environment interplay. Genes such as the glucocorticoid receptor appear to moderate the effects of childhood adversity on mental illness. Remarkably, epigenetic DNA modifications have been identified that may underlie the long-lasting effects of environment on biological functions. This new epigenetic research is pointing to a new strategy to understanding gene-environment interactions.

The next decade of research will show if this potential can be exploited in the development of new therapeutic options that may alter the traces that early environment leaves on the genome. However, as discussed in this module, the epigenome is not static and

can be molded by developmental signals, environmental perturbations, and disease states, which present an experimental challenge in the search for epigenetic risk factors in psychological disorders (Rakyan, Down, Balding, & Beck, 2011). The sample size and epigenomic assay required is dependent on the number of tissues affected, as well as the type and distribution of epigenetic modifications. The combination of genetic association maps studies with epigenome-wide developmental studies may help identify novel molecular mechanisms to explain features of inheritance of personality traits and transform our understanding of the biological basis of psychology. Importantly, these epigenetic studies may lead to identification of novel therapeutic targets and enable the development of improved strategies for early diagnosis, prevention, and better treatment of psychological and behavioral disorders.

3.7.12 Outside Resources

Reference: The “Encyclopedia of DNA Elements” (ENCODE) project

http://encodeproject.org/ENCODE/

Reference: THREADS - A new way to explore the ENCODE Project

http://www.nature.com/encode/#/threads

Web: Explore, view, and download genome-wide maps of DNA and histone modifications from the NCBI Epigenomics Portal

http://www.ncbi.nlm.nih.gov/epigenomics

Web: NOVA ScienceNOW - Introduction to Epigenetics

http://www.pbs.org/wgbh/nova/genes

Web: The University of Utah's Genetic Science Learning Center

http://learn.genetics.utah.edu/content/epigenetics/

3.7.13 Discussion Questions

1. Describe the physical state of the genome when genes are active and inactive.
2. Often, the physical characteristics of genetically identical twins become increasingly different as they age, even at the molecular level. Explain why this is so (use the terms “environment” and “epigenome”).
3. Name 3–4 environmental factors that influence the epigenome and describe their effects.
4. The rat nurturing example shows us how parental behavior can shape the behavior of offspring on a biochemical level. Discuss how this relates to humans and include the personal and social implications.
5. Explain how the food we eat affects gene expression.
6. Can the diets of parents affect their offspring’s epigenome?
7. Why is converging evidence the best kind of evidence in the study of brain function?
8. If you were interested in whether a particular brain area was involved in a specific behavior, what neuroscience methods could you use?
9. If you were interested in the precise time in which a particular brain process occurred, which neuroscience methods could you use?

3.7.14 Vocabulary (click on each word to reveal the definition)

DNA methylation

Covalent modifications of mammalian DNA occurring via the methylation of cytosine, typically in the context of the CpG dinucleotide.

DNA methyltransferases (DNMTs)

Enzymes that establish and maintain DNA methylation using methyl-group donor compounds or cofactors. The main mammalian DNMTs are DNMT1, which maintains methylation state across DNA replication, and DNMT3a and DNMT3b, which perform de novo methylation.

Epigenetics

The study of heritable changes in gene expression or cellular phenotype caused by mechanisms other than changes in the underlying DNA sequence. Epigenetic marks include covalent DNA modifications and posttranslational histone modifications.

Epigenome

The genome-wide distribution of epigenetic marks.

Gene

A specific deoxyribonucleic acid (DNA) sequence that codes for a specific polypeptide or protein or an observable inherited trait.

Genome-wide association study (GWAS)

A study that maps DNA polymorphisms in affected individuals and controls matched for age, sex, and ethnic background with the aim of identifying causal genetic variants.

Genotype

The DNA content of a cell's nucleus, whether a trait is externally observable or not.

Histone acetyltransferases (HATs) and histone deacetylases (HDACs)

HATs are enzymes that transfer acetyl groups to specific positions on histone tails, promoting an "open" chromatin state and transcriptional activation. HDACs remove these acetyl groups, resulting in a "closed" chromatin state and transcriptional repression.

Histone modifications

Posttranslational modifications of the N-terminal "tails" of histone proteins that serve as a major mode of epigenetic regulation. These modifications include acetylation, phosphorylation, methylation, sumoylation, ubiquitination, and ADP-ribosylation.

Identical twins

Two individual organisms that originated from the same zygote and therefore are genetically identical or very similar. The epigenetic profiling of identical twins discordant for disease is a unique experimental design as it eliminates the DNA sequence-, age-, and sex-differences from consideration.

Phenotype

The pattern of expression of the genotype or the magnitude or extent to which it is observably expressed—an observable characteristic or trait of an organism, such as its morphology, development, biochemical or physiological properties, or behavior.

3.7.15 References

Adams, R. L., McKay, E. L., Craig, L. M., & Burdon, R. H. (1979). Mouse DNA methylase: methylation of native DNA. *Biochimica et Biophysica Acta*, *561*(2), 345–357.

Alarcon, J. M., Malleret, G., Touzani, K., Vronskaya, S., Ishii, S., Kandel, E. R., & Barco, A. (2004). Chromatin acetylation, memory, and LTP are impaired in CBP+/- mice: a model for the cognitive deficit in Rubinstein-Taybi syndrome and its amelioration. *Neuron*, *42*(6), 947–959. doi: 10.1016/j.neuron.2004.05.021, S0896627304003022 [pii]

Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., & Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genetics*, *23*(2), 185–188.

Bateson, P. (2001). Fetal experience and good adult design. *International Journal of Epidemiology*, *30*(5), 928–934.

Bradshaw, A.D. (1965). Evolutionary significance of phenotypic plasticity in plants. *Advances in Genetics*, *13*, 115–155.

Caldji, C., Tannenbaum, B., Sharma, S., Francis, D., Plotsky, P. M., & Meaney, M. J. (1998). Maternal care during infancy regulates the development of neural systems mediating the expression of fearfulness in the rat. *Proceedings of the National Academy of Sciences U S A*, *95*(9), 5335–5340.

Cassel, S., Carouge, D., Gensburger, C., Anglard, P., Burgun, C., Dietrich, J. B., . . . Zwiller, J. (2006). Fluoxetine and cocaine induce the epigenetic factors MeCP2 and MBD1 in adult rat brain. *Molecular Pharmacology*, *70*(2), 487–492. doi: 10.1124/mol.106.022301

Chen, W. G., Chang, Q., Lin, Y., Meissner, A., West, A. E., Griffith, E. C., . . . Greenberg, M. E. (2003). Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. *Science*, *302*(5646), 885–889.

Covington, H. E., 3rd, Maze, I., LaPlant, Q. C., Vialou, V. F., Ohnishi, Y. N., Berton, O., . . . Nestler, E. J. (2009). Antidepressant actions of histone deacetylase inhibitors. *Journal of Neuroscience, 29*(37), 11451–11460. doi: 10.1523/JNEUROSCI.1758-09.2009

Davie, J. R., & Chadee, D. N. (1998). Regulation and regulatory parameters of histone modifications. *Journal of Cellular Biochemistry Suppl, 30–31 *, 203–213.

Day, J. J., & Sweatt, J. D. (2011). Epigenetic mechanisms in cognition. *Neuron, 70*(5), 813–829. doi: 10.1016/j.neuron.2011.05.019

Dick, D. M., Riley, B., & Kendler, K. S. (2010). Nature and nurture in neuropsychiatric genetics: where do we stand? *Dialogues in Clinical Neuroscience, 12*(1), 7–23.

Dolinoy, D. C., Weidman, J. R., & Jirtle, R. L. (2007). Epigenetic gene regulation: linking early developmental environment to adult disease. *Reproductive Toxicology, 23*(3), 297–307. doi: S0890-6238(06)00197-3 [pii], 10.1016/j.reprotox.2006.08.012

Feng, J., Chang, H., Li, E., & Fan, G. (2005). Dynamic expression of de novo DNA methyltransferases Dnmt3a and Dnmt3b in the central nervous system. *Journal of Neuroscience Research, 79*(6), 734–746. doi: 10.1002/jnr.20404

Feng, J., Zhou, Y., Campbell, S. L., Le, T., Li, E., Sweatt, J. D., . . . Fan, G. (2010). Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nature Neuroscience, 13*(4), 423–430. doi: 10.1038/nn.2514

Francis, D., Diorio, J., Liu, D., & Meaney, M. J. (1999). Nongenomic transmission across generations of maternal behavior and stress responses in the rat. *Science, 286*(5442), 1155–1158.

Gershon, E. S., Alliey-Rodriguez, N., & Liu, C. (2011). After GWAS: searching for genetic risk for schizophrenia and bipolar disorder. *American Journal of Psychiatry, 168*(3), 253–256. doi: 10.1176/appi.ajp.2010.10091340

Goll, M. G., & Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annual Review of Biochemistry, 74*, 481–514. doi: 10.1146/annurev.biochem.74.010904.153721

Goto, K., Numata, M., Komura, J. I., Ono, T., Bestor, T. H., & Kondo, H. (1994). Expression of DNA methyltransferase gene in mature and immature neurons as well as proliferating cells in mice. *Differentiation, 56*(1–2), 39–44.

Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G. P., Haig, D., & Dulac, C. (2010). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science, 329*(5992), 643–648. doi: 10.1126/science.1190830

Guan, J. S., Haggarty, S. J., Giacometti, E., Dannenberg, J. H., Joseph, N., Gao, J., . . . Tsai, L. H. (2009). HDAC2 negatively regulates memory formation and synaptic plasticity. *Nature, 459*(7243), 55-60. doi: 10.1038/nature07925

Guan, Z., Giustetto, M., Lomvardas, S., Kim, J. H., Miniaci, M. C., Schwartz, J. H., . . . Kandel, E. R. (2002). Integration of long-term-memory-related synaptic plasticity involves bidirectional regulation of gene expression and chromatin structure. *Cell, 111*(4), 483–493.

Guo, J. U., Ma, D. K., Mo, H., Ball, M. P., Jang, M. H., Bonaguidi, M. A., . . . Song, H. (2011). Neuronal activity modifies the DNA methylation landscape in the adult brain. *Nature Neuroscience, 14*(10), 1345–1351. doi: 10.1038/nn.2900

Heijmans, B. T., Tobi, E. W., Stein, A. D., Putter, H., Blauw, G. J., Susser, E. S., . . . Lumey, L. H. (2008). Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proceedings of the National Academy of Sciences U S A, 105*(44), 17046–17049. doi: 0806560105 [pii]10.1073/pnas.0806560105

Hong, L., Schroth, G. P., Matthews, H. R., Yau, P., & Bradbury, E. M. (1993). Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA. *Journal of Biological Chemistry, 268*(1), 305–314.

Jenuwein, T., & Allis, C. D. (2001). Translating the histone code. *Science, 293*(5532), 1074–1080. doi: 10.1126/Science.1063127293/5532/1074 [pii]

Jiang, Y. H., Bressler, J., & Beaudet, A. L. (2004). Epigenetics and human disease. *Annual Review of Genomics and Human Genetics, 5*, 479–510. doi: 10.1146/annurev.genom.5.061903.180014

Josselyn, S. A. (2005). What's right with my mouse model? New insights into the molecular and cellular basis of cognition from mouse models of Rubinstein-Taybi Syndrome. *Learning & Memory, 12*(2), 80–83. doi: 12/2/80 [pii]10.1101/lm.93505

Kadonaga, J. T. (1998). Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. *Cell*, *92*(3), 307–313.

Kalkhoven, E., Roelfsema, J. H., Teunissen, H., den Boer, A., Ariyurek, Y., Zantema, A., . . . Peters, D. J. (2003). Loss of CBP acetyltransferase activity by PHD finger mutations in Rubinstein-Taybi syndrome. *Human Molecular Genetics*, *12*(4), 441–450.

Korzus, E., Rosenfeld, M. G., & Mayford, M. (2004). CBP histone acetyltransferase activity is a critical component of memory consolidation. *Neuron*, *42*(6), 961–972. doi: 10.1016/j.neuron.2004.06.002S0896627304003526 [pii]

Kuo, M. H., & Allis, C. D. (1998). *Roles of histone acetyltransferases and deacetylases in gene regulation*. Bioessays, 20(8), 615–626. doi: 10.1002/(SICI)1521–1878(199808)20:8<615::AID-BIES4>3.0.CO;2-H [pii] 10.1002/(SICI)1521-1878(199808)20:8<615::AID-BIES4>3.0.CO;2-H

Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, *11*(3), 204–220. doi: nrg2719 [pii]10.1038/nrg2719

Lee, M. G., Wynder, C., Schmidt, D. M., McCafferty, D. G., & Shiekhattar, R. (2006). Histone H3 lysine 4 demethylation is a target of nonselective antidepressive medications. *Chemistry & Biology*, *13*(6), 563–567. doi: 10.1016/j.chembiol.2006.05.004

Levenson, J. M., O'Riordan, K. J., Brown, K. D., Trinh, M. A., Molfese, D. L., & Sweatt, J. D. (2004). Regulation of histone acetylation during memory formation in the hippocampus. *Journal of Biological Chemistry*, *279*(39), 40545–40559.

Li, H., Zhong, X., Chau, K. F., Williams, E. C., & Chang, Q. (2011). Loss of activity-induced phosphorylation of MeCP2 enhances synaptogenesis, LTP and spatial memory. *Nature Neuroscience*, *14*(8), 1001–1008. doi: 10.1038/nn.2866

Lillycrop, K. A., Phillips, E. S., Jackson, A. A., Hanson, M. A., & Burdge, G. C. (2005). Dietary protein restriction of pregnant rats induces and folic acid supplementation prevents epigenetic modification of hepatic gene expression in the offspring. *Journal of Nutrition*, *135*(6), 1382–1386. doi: 135/6/1382 [pii]

Lillycrop, K. A., Slater-Jefferies, J. L., Hanson, M. A., Godfrey, K. M., Jackson, A. A., & Burdge, G. C. (2007). Induction of altered epigenetic regulation of the hepatic glucocorticoid receptor in the offspring of rats fed a protein-restricted diet during pregnancy suggests that reduced DNA methyltransferase-1 expression is involved in impaired DNA methylation and changes in histone modifications. *British Journal of Nutrition*, *97*(6), 1064–1073. doi: S000711450769196X [pii]10.1017/S000711450769196X

Liu, D., Diorio, J., Tannenbaum, B., Caldji, C., Francis, D., Freedman, A., . . . Meaney, M. J. (1997). Maternal care, hippocampal glucocorticoid receptors, and hypothalamic- pituitary-adrenal responses to stress [see comments]. *Science*, *277*(5332), 1659–1662.

Lumey, L. H., & Stein, A. D. (1997). Offspring birth weights after maternal intrauterine undernutrition: a comparison within sibships. *American Journal of Epidemiology*, *146*(10), 810–819.

Lutter, M., Krishnan, V., Russo, S. J., Jung, S., McClung, C. A., & Nestler, E. J. (2008). Orexin signaling mediates the antidepressant-like effect of calorie restriction. *Journal of Neuroscience*, *28*(12), 3071–3075. doi: 10.1523/JNEUROSCI.5584-07.2008

Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., . . . Sun, Y. E. (2003). DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science*, *302*(5646), 890–893.

McGowan, P. O., Sasaki, A., D'Alessio, A. C., Dymov, S., Labonte, B., Szyf, M., . . . Meaney, M. J. (2009). Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nature Neuroscience*, *12*(3), 342–348. doi: nn.2270 [pii]10.1038/nn.2270

McGowan, P. O., Sasaki, A., Huang, T. C., Unterberger, A., Suderman, M., Ernst, C., . . . Szyf, M. (2008). Promoter-wide hypermethylation of the ribosomal RNA gene promoter in the suicide brain. *PLoS ONE*, *3*(5), e2085. doi: 10.1371/journal.pone.0002085

Mill, J., Tang, T., Kaminsky, Z., Khare, T., Yazdanpanah, S., Bouchard, L., . . . Petronis, A. (2008). Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. *American Journal of Human Genetics*, *82*(3), 696–711. doi: 10.1016/j.ajhg.2008.01.008

Miller, C. A., Gavin, C. F., White, J. A., Parrish, R. R., Honasoge, A., Yancey, C. R., . . . Sweatt, J. D. (2010). Cortical DNA methylation maintains remote memory. *Nature Neuroscience*, *13*(6), 664–666. doi: 10.1038/nn.2560

Myers, M. M., Brunelli, S. A., Shair, H. N., Squire, J. M., & Hofer, M. A. (1989). Relationships between maternal behavior of SHR and WKY dams and adult blood pressures of cross-fostered F1 pups. *Developmental Psychobiology*, *22*(1), 55–67.

Oberlander, T. F., Weinberg, J., Papsdorf, M., Grunau, R., Misri, S., & Devlin, A. M. (2008). Prenatal exposure to maternal depression, neonatal methylation of human glucocorticoid receptor gene (NR3C1) and infant cortisol stress responses. *Epigenetics*, *3*(2), 97–106. doi: 6034 [pii]

Ooi, S. K., O'Donnell, A. H., & Bestor, T. H. (2009). Mammalian cytosine methylation at a glance. *Journal of Cell Science*, *122*(Pt 16), 2787–2791. doi: 122/16/2787 [pii]10.1242/jcs.015123

Painter, R. C., Roseboom, T. J., & Bleker, O. P. (2005). Prenatal exposure to the Dutch famine and disease in later life: an overview. *Reproductive Toxicology*, *20*(3), 345–352. doi: S0890-6238(05)00088-2 [pii]10.1016/j. Reproductive Toxicology.2005.04.005

Petronis, A. (2010). Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature*, *465*(7299), 721–727. doi: 10.1038/nature09230

Poulter, M. O., Du, L., Weaver, I. C., Palkovits, M., Faludi, G., Merali, Z., . . . Anisman, H. (2008). GABAA receptor promoter hypermethylation in suicide brain: implications for the involvement of epigenetic processes. *Biological Psychiatry*, *64*(8), 645–652. doi: 10.1016/j.biopsych.2008.05.028

Rakyan, V. K., Down, T. A., Balding, D. J., & Beck, S. (2011). Epigenome-wide association studies for common human diseases. *Nature Reviews Genetics*, *12*(8), 529–541. doi: 10.1038/nrg3000

Razin, A. (1998). CpG methylation, chromatin structure and gene silencing-a three-way connection. *European Molecular Biology Organization*, *17*(17), 4905–4908.

Schaefer, A., Sampath, S. C., Intrator, A., Min, A., Gertler, T. S., Surmeier, D. J., . . . Greengard, P. (2009). Control of cognition and adaptive behavior by the GLP/G9a epigenetic suppressor complex. *Neuron*, *64*(5), 678–691. doi: 10.1016/j.neuron.2009.11.019

Schroeder, F. A., Lin, C. L., Crusio, W. E., & Akbarian, S. (2007). Antidepressant-like effects of the histone deacetylase inhibitor, sodium butyrate, in the mouse. *Biological Psychiatry*, *62*(1), 55-64. doi: 10.1016/j.biopsych.2006.06.036

Sealy, L., & Chalkley, R. (1978). DNA associated with hyperacetylated histone is preferentially digested by DNase I. *Nucleic Acids Research*, *5*(6), 1863–1876.

Shahbazian, M., Young, J., Yuva-Paylor, L., Spencer, C., Antalffy, B., Noebels, J., . . . Zoghbi, H. (2002). Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. *Neuron*, *35*(2), 243–254.

Skene, P. J., Illingworth, R. S., Webb, S., Kerr, A. R., James, K. D., Turner, D. J., . . . Bird, A. P. (2010). Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Molecular Cell*, *37*(4), 457–468. doi: 10.1016/j.molcel.2010.01.030

Stanner, S. A., Bulmer, K., Andres, C., Lantseva, O. E., Borodina, V., Poteen, V. V., & Yudkin, J. S. (1997). Does malnutrition in utero determine diabetes and coronary heart disease in adulthood? Results from the Leningrad siege study, a cross sectional study. *British Medical Journal*, *315*(7119), 1342–1348.

Stern, J. M. (1997). Offspring-induced nurturance: animal-human parallels. *Developmental Psychobiololgy*, *31*(1), 19–37.

Sutter, D., Doerfler, W., 1980. Methylation of integrated adenovirus type 12 DNA sequences in transformed cells is inversely correlated with viral gene expression. *Proceedings of the National Academy of Sciences U S A. 77*, 253–256.

Suzuki, M. M., & Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics*, *9*(6), 465–476. doi: nrg2341 [pii]10.1038/nrg2341

Tsankova, N. M., Berton, O., Renthal, W., Kumar, A., Neve, R. L., & Nestler, E. J. (2006). Sustained hippocampal chromatin regulation in a mouse model of depression and antidepressant action. *Nature Neuroscience. 9*(4): 519–525. doi:10.1038/nn1659

Turner, J. D., Pelascini, L. P., Macedo, J. A., & Muller, C. P. (2008). Highly individual methylation patterns of alternative glucocorticoid receptor promoters suggest individualized epigenetic regulatory mechanisms. *Nucleic Acids Research*, *36*(22),

7207–7218. doi: gkn897 [pii] 10.1093/nar/gkn897

Vardimon, L., Kressmann, A., Cedar, H., Maechler, M., Doerfler, W., 1982. Expression of a cloned adenovirus gene is inhibited by in vitro methylation. *Proceedings of the National Academy of Sciences U S A. 79*, 1073–1077.

Waddington, C. H. (1942). Epigenotype. *Endeavour*(1), 18–21.

Wade, P. A., Pruss, D., & Wolffe, A. P. (1997). Histone acetylation: chromatin in action. *Trends in Biochemical Sciences*, *22*(4), 128–132. doi: S0968000497010165 [pii]

Wang, J., Weaver, I. C., Gauthier-Fisher, A., Wang, H., He, L., Yeomans, J., . . . Miller, F. D. (2010). CBP histone acetyltransferase activity regulates embryonic neural differentiation in the normal and Rubinstein-Taybi syndrome brain. *Developmental Cell*, *18*(1), 114–125. doi: 10.1016/j.devcel.2009.10.023

Weaver, I. C., Cervoni, N., Champagne, F. A., D'Alessio, A. C., Sharma, S., Seckl, J. R., . . . Meaney, M. J. (2004). Epigenetic programming by maternal behavior. *Nature Neuroscience*, *7*(8), 847–854. doi: 10.1038/nn1276

Weaver, I. C., Champagne, F. A., Brown, S. E., Dymov, S., Sharma, S., Meaney, M. J., & Szyf, M. (2005). Reversal of maternal programming of stress responses in adult offspring through methyl supplementation: altering epigenetic marking later in life. *Journal of Neuroscience*, *25*(47), 11045–11054. doi: 10.1523/JNEUROSCI.3652-05.2005

Weaver, I. C., Meaney, M. J., & Szyf, M. (2006). Maternal care effects on the hippocampal transcriptome and anxiety-mediated behaviors in the offspring that are reversible in adulthood. *Proceedings of the National Academy of Sciences U S A*, *103*(9), 3480–3485. doi: 10.1073/pnas.0507526103

Wells, J. C. (2003). The thrifty phenotype hypothesis: thrifty offspring or thrifty mother? *Journal of Theoretical Biology*, *221*(1), 143–161.

Wilkinson, M. B., Xiao, G., Kumar, A., LaPlant, Q., Renthal, W., Sikder, D., . . . Nestler, E. J. (2009). Imipramine treatment and resiliency exhibit similar chromatin regulation in the mouse nucleus accumbens in depression models. *Journal of Neuroscience*, *29*(24), 7820–7832. doi: 10.1523/JNEUROSCI.0932-09.2009

Wolffe, A. P., & Matzke, M. A. (1999). Epigenetics: regulation through repression. *Science*, *286*(5439), 481–486.

3.7.16 Attributions

Adapted by Kenneth A. Koenigshofer, Ph.D, Chaffey College, from Epigenetics in Psychology by Ian Weaver licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

This page titled 3.7: Epigenetics in Psychology is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

3.8: Human Genetics

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Explain the basic principles of the theory of evolution by natural selection.
2. Discuss evolutionary psychology and behavior genetics.
3. Describe the difference between genotype and phenotype.
4. Explain the structure and function of codons in protein synthesis.
5. Describe transcription and translation and the roles of DNA, mRNA, and tRNA in protein synthesis.
6. Describe types of inheritance.
7. Discuss examples of genetic diseases and their patterns of inheritance.
8. Describe non-heritable genetic disorders and give examples.

3.8.0.1 Overview

Psychological researchers study genetics in order to better understand the biological basis that contributes to certain behaviors. While all humans share certain biological mechanisms, we are each unique. And while our bodies have many of the same parts—brains and hormones and cells with genetic codes—these are expressed in a wide variety of behaviors, thoughts, and reactions.

Why do two people infected by the same disease have different outcomes: one surviving and one succumbing to the ailment? How are genetic diseases passed through family lines? Are there genetic components to psychological disorders, such as depression or schizophrenia? To what extent might there be a psychological basis to health conditions such as childhood obesity?

To explore these questions, let's start by focusing on a specific disease, sickle-cell anemia, and how it might affect two infected sisters. Sickle-cell anemia is a genetic condition in which red blood cells, which are normally round, take on a crescent-like shape. The changed shape of these cells affects how they function: sickle-shaped cells can clog blood vessels and block blood flow, leading to high fever, severe pain, swelling, and tissue damage.

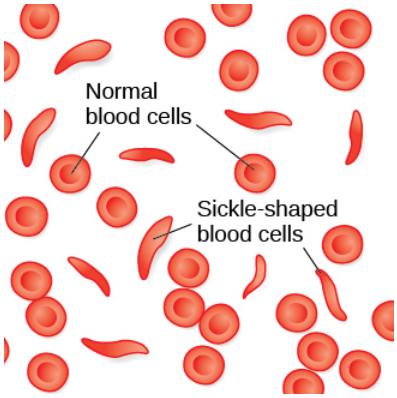


Figure 3.8.1: Normal blood cells travel freely through the blood vessels, while sickle-shaped cells form blockages preventing blood flow.

Many people with sickle-cell anemia—and the particular genetic mutation that causes it—die at an early age. While the notion of "survival of the fittest" may suggest that people suffering from this disease have a low survival rate and therefore the disease will become less common, this is not the case. Despite the negative evolutionary effects associated with this genetic mutation, the sickle-cell gene remains relatively common among people of African descent. Why is this? The explanation is illustrated with the following scenario.

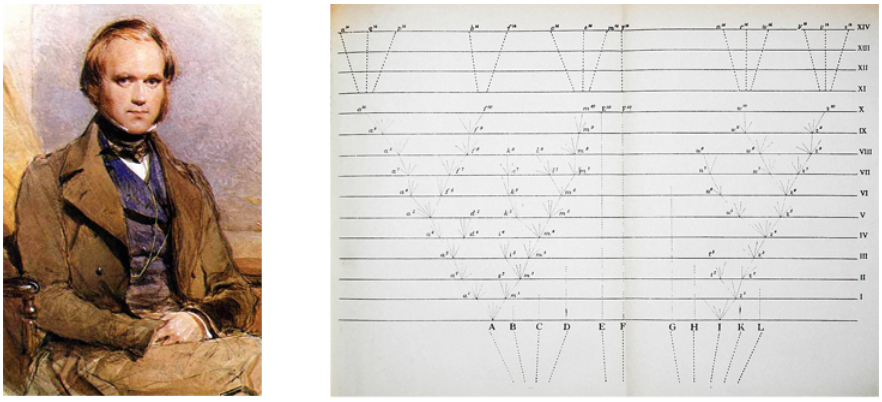
Imagine two young women—Luwi and Sena—sisters in rural Zambia, Africa. Luwi carries the gene for sickle-cell anemia; Sena does not carry the gene. Sickle-cell carriers have one copy of the sickle-cell gene but do not have full-blown sickle-cell anemia.

They experience symptoms only if they are severely dehydrated or are deprived of oxygen (as in mountain climbing). Carriers are thought to be immune from malaria (an often deadly disease that is widespread in tropical climates) because changes in their blood chemistry and immune functioning prevent the malaria parasite from having its effects (Gong, Parikh, Rosenthal, & Greenhouse, 2013). However, full-blown sickle-cell anemia, with two copies of the sickle-cell gene, does not provide immunity to malaria.

While walking home from school, both sisters are bitten by mosquitos carrying the malaria parasite. Luwi does not get malaria because she carries the sickle-cell mutation. Sena, on the other hand, develops malaria and dies just two weeks later. Luwi survives and eventually has children, to whom she may pass on the sickle-cell mutation.

Malaria is rare in the United States, so the sickle-cell gene benefits nobody: the gene manifests primarily in health problems—minor in carriers, severe in the full-blown disease—with no health benefits for carriers. However, the situation is quite different in other parts of the world. In parts of Africa where malaria is prevalent, having the sickle-cell mutation does provide health benefits for carriers (protection from malaria).

This is precisely the situation that Charles Darwin describes in the theory of evolution by natural selection. In simple terms, the theory states that organisms that are better suited for their environment will survive and reproduce, while those that are poorly suited for their environment will die off. In our example, we can see that as a carrier, Luwi's mutation is highly adaptive in her African homeland; however, if she resided in the United States (where malaria is much less common), her mutation could prove costly—with a high probability of the disease in her descendants and minor health problems of her own.



(a) (b)

Figure **3.8.2**: (a) In 1859, Charles Darwin proposed his theory of evolution by natural selection in his book, *On the Origin of Species*. (b) The book contains just one illustration: this diagram that shows how species evolve over time through natural selection.

3.8.0.1 Two Perspectives on Genetics and Behavior

It's easy to get confused about two fields that study the interaction of genes and the environment, such as the fields of **evolutionary psychology** and **behavioral genetics**. How can we tell them apart?

In both fields, it is understood that genes not only code for particular traits, but also contribute to certain patterns of cognition and behavior. Evolutionary psychology focuses on how universal patterns of behavior and cognitive processes have evolved over time. Therefore, variations in cognition and behavior would make individuals more or less successful in reproducing and passing those genes to their offspring. Evolutionary psychologists study a variety of psychological phenomena that may have evolved as adaptations, including fear response, food preferences, mate selection, and cooperative behaviors (Confer et al., 2010).

Whereas evolutionary psychologists focus on universal patterns that evolved over millions of years, behavioral geneticists study how individual differences arise, in the present, through the interaction of genes and the environment. When studying human behavior, behavioral geneticists often employ twin and adoption studies to research questions of interest. Twin studies compare the rates that a given behavioral trait is shared among identical and fraternal twins; adoption studies compare those rates among biologically related relatives and adopted relatives. Both approaches provide some insight into the relative importance of genes and environment for the expression of a given trait.

Watch this interview with renowned **evolutionary psychologist** Davis Buss for an explanation of how a psychologist approaches evolution and how this approach fits within the field of social science.

3.8.0.1 Chromosomes, Genes, and Genetic Variation

Genetics is the science of the way traits are passed from parent to offspring. For all forms of life, continuity of the species depends upon the genetic code being passed from parent to offspring. Evolution by natural selection is dependent on traits being heritable. Genetics is very important in human physiology because all attributes of the human body are affected by a person's genetic code. It can be as simple as eye color, height, or hair color. Or it can be as complex as how well your liver processes toxins, whether you will be prone to heart disease or breast cancer, and whether you will be color blind. Defects in the genetic code can be tragic. For example: Down Syndrome, Turner Syndrome, and Klinefelter's Syndrome are diseases caused by chromosomal abnormalities. Cystic fibrosis is caused by a single change in the genetic sequence.

Genetic inheritance begins at the time of conception. You inherited 23 chromosomes from your mother and 23 from your father. Together they form 22 pairs of autosomal chromosomes and a pair of sex chromosomes (either XX if you are female, or XY if you are male). Homologous chromosomes have the same genes in the same positions, but may have different alleles (varieties) of those genes. There can be many alleles of a gene within a population, but an individual within that population only has two copies, and can be homozygous (both copies the same) or heterozygous (the two copies are different) for any given gene. The sequence of the human genome (approximately 3 billion base pairs in a human haploid genome with an estimated 20,000-25,000 protein-coding genes) was completed in 2003, but we are far from understanding the functions and regulations of all the genes.

Deoxyribonucleic acid (DNA) is the macromolecule that stores the information necessary to build structural and functional cellular components. It has a double-helix structure (see figure below) in which two strands wrap around one another. Stretched end-to-end, the DNA molecules in a single human cell would come to a **length of about 2 meters**. Thus, the DNA for a cell must be packaged in a very ordered way to fit and function within the cell. To fit their DNA inside the nucleus, DNA is wrapped around proteins known as histones.

DNA has three types of chemical component: phosphate, a sugar called deoxyribose, and four bases—**adenine, guanine, cytosine, and thymine**. Groups of three bases, known as **base triplets**, or **codons**, are the basic coding unit. **Each base triplet** (also known as a **codon**) **codes for a specific amino acid**. **Proteins** are composed of **strings of amino acids.**

DNA provides the basis for inheritance when DNA is passed from parent to offspring. A **gene** is a segment of DNA that codes for the synthesis of a protein and acts as a unit of inheritance that can be transmitted from generation to generation. The external appearance (***phenotype***) of an organism is determined to a large extent by the genes it inherits (***genotype***). Thus, one can begin to see how variation at the DNA level can cause variation at the level of the entire organism. These concepts form the basis of **genetics** and **evolutionary theory**. Genetic variation in a species provides the raw material, the genetic variants, for natural selection to operate upon thereby creating evolutionary change.

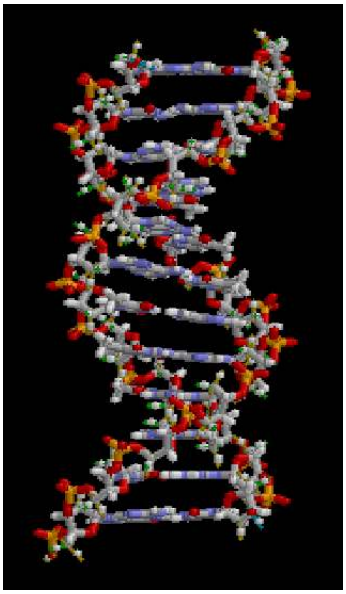
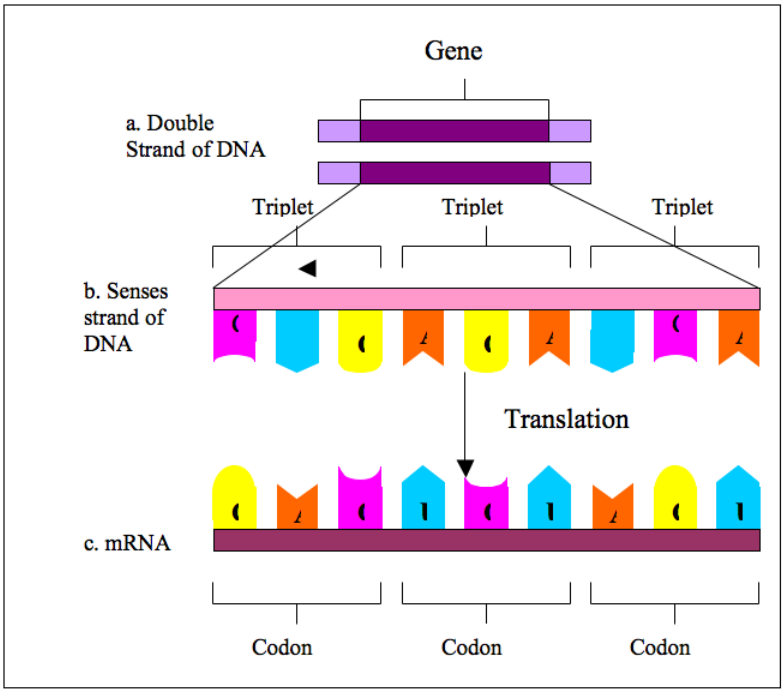


Figure **3.8.3**: Rotating animation of a DNA molecule, showing its double-helix structure in which two strands of nucleotides wind around each other in a spiral shape (Image from Wikimedia Commons; File:DNA animation.gif;

https://commons.wikimedia.org/wiki/F..._animation.gif; by brian0918&#153. This work has been released into the **public domain** by its author, **brian0918**. This applies worldwide. Caption by Kenneth A. Koenigshofer, Ph.D., Chaffey College).

A **gene** is made up of short sections of DNA which are contained on a chromosome within the nucleus of a cell. Genes control the development and function of all organs and all working systems in the body. A gene has a certain influence on how the cell works; the same gene in many different cells determines a certain physical or biochemical feature of the whole body (e.g. eye color or reproductive functions). All human cells hold approximately 20,000-30,000 different protein-coding genes.



a. genes are portions of DNA that code for a particular protein.
b. Triplets, the three-base sequences code for amino acid sequences.
c. Transcribed mRNA codons, which are complementary to the code in DNA triplets.

Figure **3.8.4**: Genes, codons, and transcription (the process of making RNA) and translation (the synthesis of the protein on the ribosome as the mRNA moves across the ribosome). Also see Figures 3.13.5 and 3.13.6 and text for additional details. (Image from Wikibooks; Human Physiology/Genetics and inheritance; https://en.wikibooks.org/wiki/Human_...nheritance#DNA; under the Creative Commons Attribution-ShareAlike License).

Even though each cell has identical copies of all of the same genes, different cells express or repress different genes. This is what accounts for the differences between, let's say, a liver cell and a brain cell. **Genotype** is the actual pair of genes that a person has for a trait of interest. For example, a woman could be a carrier for hemophilia by having one normal copy of the gene for a particular clotting protein and one defective copy. A **Phenotype** is the organism's physical appearance or functioning as it relates to a certain trait. In the case of the woman carrier, her phenotype is normal (because the normal copy of the gene is dominant to the defective copy). The phenotype can be for any measurable trait, such as eye color, finger length, height, physiological traits like the ability to pump calcium ions from mucosal cells, behavioral traits like smiles, and biochemical traits like blood types and cholesterol levels. Genotype cannot always be predicted by phenotype (we would not know the woman was a carrier of hemophilia just based on her appearance), but can be determined through pedigree charts or direct genetic testing. Even though genotype is a strong predictor of phenotype, environmental factors can also play a strong role in determining phenotype. Identical twins, for example, are genetic clones resulting from the early splitting of an embryo, but they can be quite different in personality, body mass, and even fingerprints.

Genes encode the information necessary for synthesizing the amino-acid sequences in proteins, which in turn play a large role in determining the final phenotype, or physical appearance and functioning of the organism. In **diploid** organisms (organisms that

have paired chromosomes, one from each parent), a dominant allele on one chromosome will mask the expression of a recessive allele on the other. While most genes are dominant/recessive, others may be codominant or show different patterns of expression. The phrase "to code for" is often used to mean a gene contains the instructions about a particular protein, (as in the gene codes for the protein). The "one gene, one protein" concept is now known to be the simplistic. For example, a single gene may produce multiple products, depending on how its transcription is regulated. Genes code for the nucleotide sequence in **messenger RNA** (mRNA) and **transfer RNA** (rRNA), required for protein synthesis (see Figure 3.13.5 below).

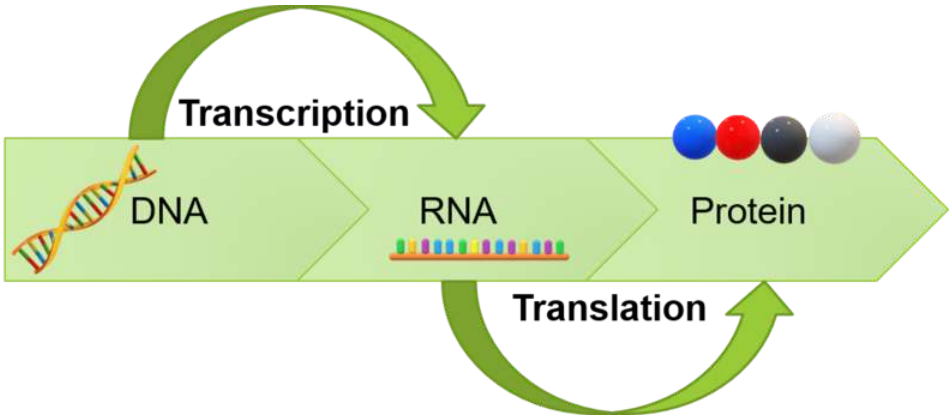


Figure 3.8.5: Transcription (information transcribed from DNA to RNA) and Translation (messenger RNA to protein synthesis on the ribosome). (Image from Wikimedia Commons; File:Transcription and Translation.png; https://commons.wikimedia.org/wiki/F...ranslation.png; by Christinelmiller; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

DNA must be “read” to produce the molecules, such as proteins, to carry out the functions of the cell. This "reading" of information in DNA involves two related processes--transcription and translation. **Transcription** is the process of making RNA (ribonucleic acid) directed by information in DNA. RNA is in all cells and like DNA is composed of nucleotides. RNA nucleotides contain the bases adenine, cytosine, and guanine. However, they do **not contain thymine**, which is instead **replaced by uracil**, symbolized by a “U.” RNA exists as a single-stranded molecule rather than a double-stranded helix like that of DNA. There are several kinds of RNA, named on the basis of their function. As mentioned above, these include messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA)—molecules that are involved in the production of proteins from the DNA code. **Translation** is the synthesis of the protein on the **ribosome** as the messenger RNA (mRNA) moves across the ribosome. Ribosomes are macromolecular machines, found within all living cells, that perform biological protein synthesis.

As noted, the gene is a segment of DNA which contains the information for making a protein. In response to an enzyme RNA polymerase breaks the hydrogen bonds of the gene. As it breaks the hydrogen bonds it begins to move down the gene. Next the RNA polymerase will line up the nucleotides so they are complementary. **Transcription occurs in** the **nucleus**, and once transcription is completed the **messenger RNA (mRNA) will leave** the **nucleus, and go into the cytoplasm where** the **mRNA will bind to** a free floating **ribosome. mRNA carries the protein blueprint from a cell's DNA to its ribosomes**, which are **the "machines" that drive protein synthesis. Transfer RNA (tRNA) carries** the appropriate **amino acids into** the **ribosome for inclusion in the new protein. Ribosomes link amino acids together in** the **order specified by the codons (base triplets) of messenger RNA** molecules to form polypeptide chains. **The mRNA base sequence determines the order of assembling of the amino acids to form specific proteins.**

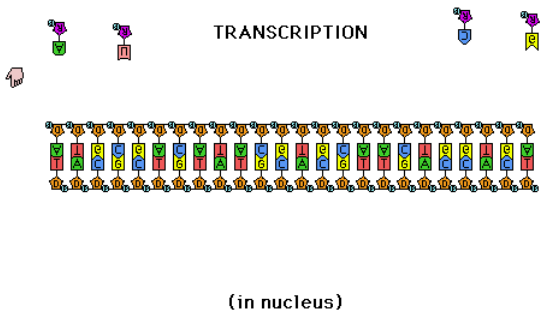


Figure 3.8.6: This file represents the transcription of a gene into messenger RNA, followed by the translation of mRNA into a polypeptide. Inherited information from DNA is transcribed to mRNA in the nucleus, then mRNA moves to the cytoplasm, tRNA brings amino acids, small ribosomal unit, including rRNA, attaches to mRNA which assembles polypeptides in synthesis of specific protein; finally, ribosome detaches from mRNA. (Image from Wikimedia Commons; File:DNA Transcription and Translation.gif; https://commons.wikimedia.org/wiki/F...ranslation.gif; by Steven Kuensting; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

Inheritance

A person's cells hold the exact genes that originated from the sperm and egg of his parents at the time of conception. The genes of a cell are formed into long strands of DNA. Most of the genes that control characteristic are in pairs, one gene from mom and one gene from dad. Everybody has 22 pairs of chromosomes (*autosomes*) and two more genes called sex-linked chromosomes. Females have two X (XX) chromosomes and males have an X and a Y (XY) chromosome. Inherited traits and disorders can be divided into three categories: **unifactorial inheritance, sex-linked inheritance, and multifactor inheritance.**

3.8.0.0.1 Unifactorial Inheritance

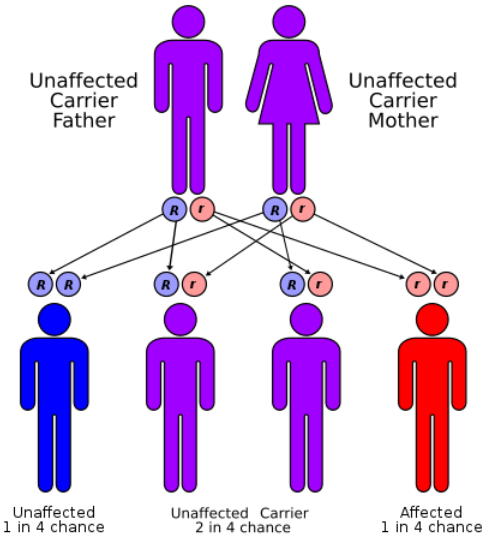


Figure 3.8.7: Chart showing the possibilities of contracting a recessive defect, from two carrier parents.

Traits such as blood type, eye color, hair color, and taste are each thought to be controlled by a single pair of genes. The Austrian monk Gregor Mendel was the first to discover this phenomenon, and it is now referred to as *the laws of Mendelian inheritance*. The genes deciding a single trait may have several forms (*alleles*). For example, the gene responsible for hair color has two main alleles: red and brown. The four possibilities are thus

Brown/red, which would result in brown hair,
Red/red, resulting in red hair,
Brown/brown, resulting in brown hair, or
Red/brown, resulting in red hair.

The genetic codes for red and brown can be either dominant or recessive. In any case, the dominant gene overrides the recessive.

When two people create a child, they each supply their own set of genes. In simplistic cases, such as the red/brown hair, each parent supplies one "code", contributing to the child's hair color. For example, if dad has brown/red he has a 50% chance of passing brown hair to his child and a 50% of passing red hair. When combined with a mom who has brown/brown (who would supply 100% brown), the child has a 75% chance of having brown hair and a 25% chance of having red hair. Similar rules apply to different traits and characteristics, though they are usually far more complex.

3.8.0.0.2 Multifactorial inheritance

Some traits are found to be determined by genes and environmental effects. Height for example seems to be controlled by multiple genes, some are "tall" genes and some are "short" genes. A child may inherit all the "tall" genes from both parents and will end up taller than both parents. Or the child my inherit all the "short" genes and be the shortest in the family. More often than not the child inherits both "tall" and "short" genes and ends up about the same height as the rest of the family. Good diet and exercise can help a person with "short" genes end up attaining an average height. Babies born with drug addiction or alcohol addiction are a sad example of environmental inheritance. When mom is doing drugs or drinking, everything that she takes the baby takes. These babies often have developmental problems and learning disabilities. A baby born with *Fetal alcohol syndrome* is usually abnormally short, has small eyes and a small jaw, may have heart defects, a cleft lip and palate, may suck poorly, sleep poorly, and be irritable. About one fifth of the babies born with fetal alcohol syndrome die within the first weeks of life, those that live are often mentally and physically handicapped.

3.8.0.0.3 Sex-linked Inheritance

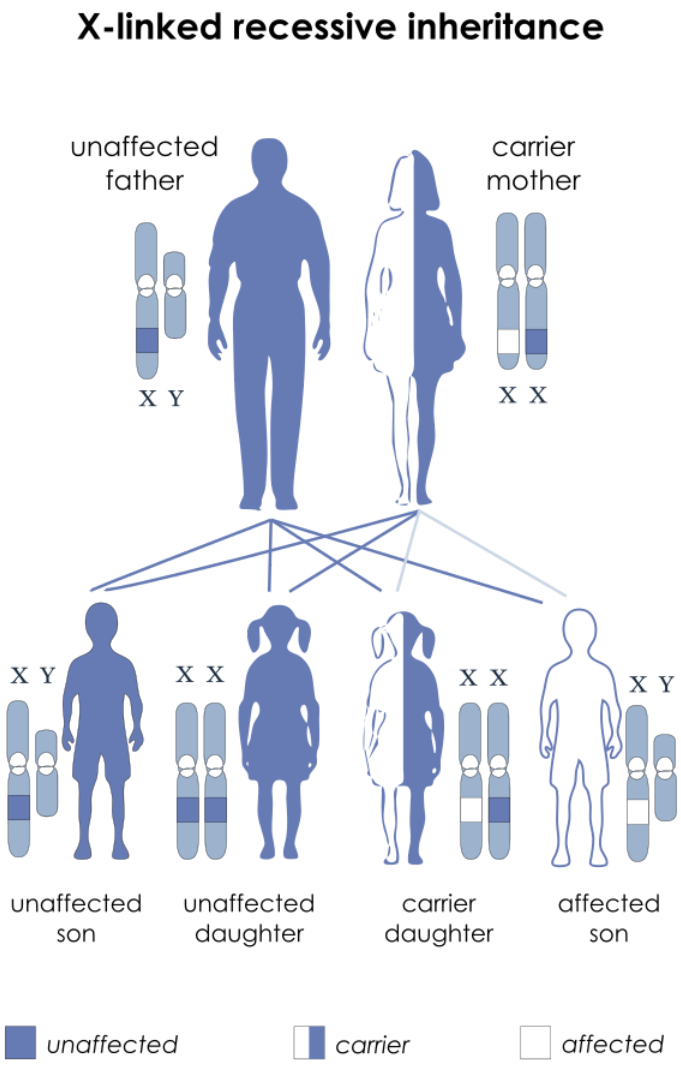


Figure 3.8.8: X-linked recessive inheritance.

Sex-linked inheritance is quite obvious, it determines your gender. Male gender is caused by the Y chromosome which is only found in males and is inherited from their fathers. The genes on the Y chromosomes direct the development of the male sex organs. The x chromosome is not as closely related to the female sex because it is contained in both males and females. Males have a single X and females have double XX. The X chromosome is to regulate regular development and it seems that the Y is added just for the male genitalia. When there is a default with the X chromosomes in males it is almost always persistent because there is not the extra X chromosome that females have to counteract the problem. Certain traits like colorblindness and hemophilia are on alleles carried on the X chromosome. For example if a woman is colorblind all of her sons will be colorblind. Whereas all of her daughters will be carriers for colorblindness.

3.8.0.0.4 Exceptions to simple inheritance

Our knowledge of the mechanisms of genetic inheritance has grown a lot since Mendel's time. It is now understood, that if you inherit one allele, it can sometimes increase the chance of inheriting another and can affect when or how a trait is expressed in an individuals phenotype. There are levels of dominance and recessiveness with some traits. Mendel's simple rules of inheritance does not always apply in these exceptions.

3.8.0.0.5 Polygenic Traits

Polygenic traits are traits determined by the combined effect of more than one pair of genes. Human stature is an example of this trait. The size of all body parts from head to foot combined determines height. The size of each individual body part are determined by numerous genes. Human skin, eyes, and hair are also polygenic genes because they are determined by more than one allele at a different location.

3.8.0.0.6 Intermediate Expressions

When there is incomplete dominance, blending can occur resulting in heterozygous individuals. An example of intermediate expression is the pitch of a human male voice. Homozygous men have the lowest and highest voice for this trait (AA and aa). Tay-Sachs, causing death in childhood, is also characterized by incomplete dominance.

3.8.0.0.7 Co-dominance

For some traits, two alleles can be co-dominant. Were both alleles are expressed in heterozygous individuals. An example of that would be a person with AB blood. These people have the characteristics of both A and B blood types when tested.

3.8.0.0.8 Multiple-Allele Series

There are some traits that are controlled by far more alleles. For example, the human HLA system, which is responsible for accepting or rejecting foreign tissue in our bodies, can have as many as 30,000,000 different genotypes! The HLA system is what causes the rejection of organ transplants. The multiple allele series is very common, as geneticists learn more about genetics, they realize that it is more common than the simple two allele ones.

3.8.0.0.9 Modifying and Regulator Genes

Modifying and regulator genes are the two classes of genes that may have an effect on how the other genes function. *Modifying Genes* alter how other genes are expressed in the phenotype. For example, a dominant cataracts gene may impair vision at various degrees, depending on the presence of a specific allele for a companion modifying gene. However, cataracts can also come from excessive exposure to ultraviolet rays and diabetes. *Regulator Genes* also known as homoerotic genes, can either initiate or block the expression of other genes. They also control a variety of chemicals in plants and animals. For example, Regulator genes control the time of production of certain proteins that will be new structural parts of our bodies. Regulator genes also work as a master switch starting the development of our body parts right after conception and are also responsible for the changes in our bodies as we get older. They control the aging processes and maturation.

3.8.0.0.10 Incomplete penetrates

Some genes are incomplete penetrate, which means, unless some environmental factors are present, the effect does not occur. For example, you can inherit the gene for diabetes, but never get the disease, unless you were greatly stressed, extremely overweight, or didn't get enough sleep at night. These interactions between genotype and environment fall under the category of epigenetics to be discussed in more detail later in this chapter.

3.8.0.1 Genetic Diseases and Patterns of Inheritance

Inheritance pattern	Description	Examples
Autosomal dominant	Only one mutated copy of the gene is needed for a person to be affected by an autosomal dominant disorder. Each affected person usually has one affected parent. There is a 50% chance that a child will inherit the mutated gene. Many disease conditions that are autosomal dominant have low penetrance, which means that although only one mutated copy is needed, a relatively small proportion of those who inherit that mutation go on to develop the disease, often later in life.	Huntingtons disease, Neurofibromatosis 1, HBOC syndrome, Hereditary nonpolyposis colorectal cancer
Autosomal recessive	Two copies of the gene must be mutated for a person to be affected by an autosomal recessive disorder. An affected person usually has unaffected parents who each carry a single copy of the mutated gene (and are referred to as carriers). Two unaffected people who each carry one copy of the mutated gene have a 25% chance with each pregnancy of having a child affected by the disorder.	Cystic fibrosis, Sickle cell anemia, Tay-Sachs disease, Spinal muscular atrophy, Muscular dystrophy
X-linked dominant	X-linked dominant disorders are caused by mutations in genes on the X chromosome. Only a few disorders have this inheritance pattern. Females are more frequently affected than males, and the chance of passing on an X-linked dominant disorder differs between men and women. The sons of a man with an X-linked dominant disorder will not be affected, and his daughters will all inherit the condition. A woman with an X-linked dominant disorder has a 50% chance of having an affected daughter or son with each pregnancy. Some X-linked dominant conditions, such as Aicardi Syndrome, are fatal to boys, therefore only girls have them (and boys with Klinefelter Syndrome).	Hypophosphatemia, Aicardi Syndrome
X-linked recessive	X-linked recessive disorders are also caused by mutations in genes on the X chromosome. Males are more frequently affected than females, and the chance of passing on the disorder differs between men and women. The sons of a man with an X-linked recessive disorder will not be affected, and his daughters will carry one copy of the mutated gene. With each	Hemophilia A, Duchenne muscular dystrophy, Color blindness, Turner Syndrome

	pregnancy, a woman who carries an X-linked recessive disorder has a 50% chance of having sons who are affected and a 50% chance of having daughters who carry one copy of the mutated gene.	
Y-linked	Y-linked disorders are caused by mutations on the Y chromosome. Only males can get them, and all of the sons of an affected father are affected. Since the Y chromosome is very small, Y-linked disorders only cause infertility, and may be circumvented with the help of some fertility treatments.	Male Infertility
Mitochondrial	This type of inheritance, also known as maternal inheritance, applies to genes in mitochondrial DNA. Because only egg cells contribute mitochondria to the developing embryo, only females can pass on mitochondrial conditions to their children.	Leber's Hereditary Optic Neuropathy (LHON)

Table 3.13.1. Genetic Diseases (above).

3.8.0.1 Non-heritable Genetic Disorders

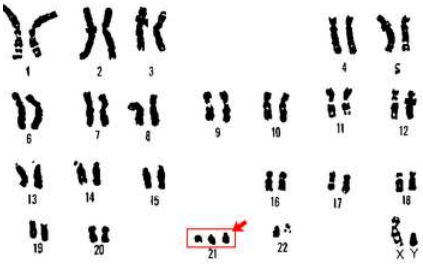


Figure **3.8.9**: Karyotype of 21 trisomy-Down syndrome.

Any disorder caused totally or in part by a fault (or faults) of the genetic material passed from parent to child is considered a genetic disorder. The genes for many of these disorders are passed from one generation to the next, and children born with a heritable genetic disorder often have one or more extended family members with the same disorder. There are also genetic disorders that appear due to spontaneous faults in the genetic material, in which case a child is born with a disorder with no apparent family history.

Down Syndrome, also known as Trisomy 21, is a chromosome abnormality that effects one out of every 800-1000 newborn babies. During anaphase II of meiosis the sister chromatids of chromosome 21 fail to separate, resulting in an egg with an extra chromosome, and a fetus with three copies (trisomy) of this chromosome. At birth this defect is recognizable because of the physical features such as almond shaped eyes, a flattened face, and less muscle tone than a normal newborn baby. During pregnancy, it is possible to detect the Down Syndrome defect by doing amniocentesis testing. There is a risk to the unborn baby and it is not recommended unless the pregnant mother is over the age of thirty-five. Other non-lethal chromosomal abnormalities include additional osex chromosome abnormalities which is when a baby girl (about 1 in 2,500)is born with one x instead of two (xx) this can cause physical abnormalities and defective reproduction systems. Boys can also be born with extra X's (XXY or XXXY) which will cause reproductive problems and sometimes mental retardation.

Chromosomal Abnormalities In most cases with a chromosomal abnormality all the cells are affected. Defects can have anywhere from little effect to a lethal effect depending on the type of abnormality. Of the 1 in 200 babies born having some sort of

chromosomal abnormality, about 1/3 of these results in spontaneous abortion. Abnormalities usually form shortly after fertilization and mom or dad usually has the same abnormality. There is no cure for these abnormalities. Tests are possible early in pregnancy and if a problem is detected the parents can choose to abort the fetus.

Genetics (from the Greek *genno* = give birth) is the science of genes, heredity, and the variation of organisms.

Genetic variation, the genetic difference between individuals, is what contributes to a species' adaptation to its environment by providing genetic alternatives that natural selection can "pick" from to achieve improved adaptation over generations by evolution. In humans, genetic variation begins with an egg, about 100 million sperm, and fertilization. Fertile women ovulate roughly once per month, releasing an egg from follicles in the ovary. The egg travels, via the fallopian tube, from the ovary to the uterus, where it may be fertilized by a sperm.

The egg and the sperm each contain 23 chromosomes. **Chromosomes** are long strings of genetic material, deoxyribonucleic acid (DNA). DNA is a helix-shaped molecule made up of nucleotide base pairs. In each chromosome, sequences of DNA make up **genes** that control or partially control a number of visible characteristics, known as traits, such as eye color, hair color, and so on. A single gene may have multiple possible variations, or alleles. An allele is a specific version of a gene. So, a given gene may code for the trait of hair color, and the different alleles of that gene affect which hair color an individual has.

When a sperm and egg fuse, their 23 chromosomes pair up and create a zygote with 23 pairs of chromosomes. Therefore, each parent contributes half the genetic information carried by the offspring; the resulting physical characteristics of the offspring (called the phenotype) are determined by the interaction of genetic material supplied by the parents (called the genotype). A person's genotype is the genetic makeup of that individual. **Phenotype**, on the other hand, refers to the individual's inherited physical characteristics.

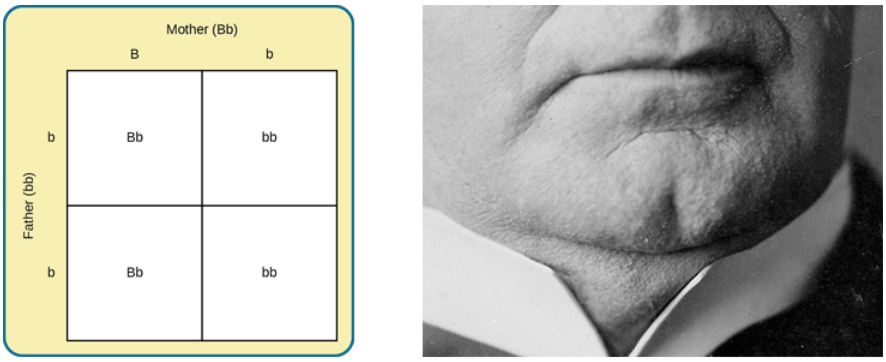


(a) (b)

Figure **3.8.10**: (a) Genotype refers to the genetic makeup of an individual based on the genetic material (DNA) inherited from one's parents. (b) Phenotype describes an individual's observable characteristics, such as hair color, skin color, height, and build. (credit a: modification of work by Caroline Davis; credit b: modification of work by Cory Zanker)

Most traits are controlled by multiple genes, but some traits are controlled by one gene. A characteristic like cleft chin, for example, is influenced by a single gene from each parent. In this example, we will call the gene for cleft chin "B," and the gene for smooth chin "b." Cleft chin is a dominant trait, which means that having the dominant allele either from one parent (Bb) or both parents (BB) will always result in the phenotype associated with the dominant allele. When someone has two copies of the same allele, they are said to be homozygous for that allele. When someone has a combination of alleles for a given gene, they are said to be heterozygous. For example, smooth chin is a recessive trait, which means that an individual will only display the smooth chin phenotype if they are homozygous for that recessive allele (bb).

Imagine that a woman with a cleft chin mates with a man with a smooth chin. What type of chin will their child have? The answer to that depends on which alleles each parent carries. If the woman is homozygous for cleft chin (BB), her offspring will always have cleft chin. It gets a little more complicated, however, if the mother is heterozygous for this gene (Bb). Since the father has a smooth chin—therefore homozygous for the recessive allele (bb)—we can expect the offspring to have a 50% chance of having a cleft chin and a 50% chance of having a smooth chin.



(a) (b)

Figure **3.8.11**: (a) A **Punnett square** is a tool used to predict how genes will interact in the production of offspring. The capital B represents the dominant allele, and the lowercase b represents the recessive allele. In the example of the cleft chin, where B is cleft chin (dominant allele), wherever a pair contains the dominant allele, B, you can expect a cleft chin phenotype. You can expect a smooth chin phenotype only when there are two copies of the recessive allele, bb. (b) A cleft chin, shown here, is an inherited trait.

Sickle-cell anemia is just one of many genetic disorders caused by the pairing of two recessive genes. For example, **phenylketonuria** (PKU) is a condition in which individuals lack an enzyme that normally converts harmful amino acids into harmless byproducts. If someone with this condition goes untreated, he or she will experience significant deficits in cognitive function, seizures, and increased risk of various psychiatric disorders. Because PKU is a recessive trait, each parent must have at least one copy of the recessive allele in order to produce a child with the condition (Figure (Links to an external site.)).

So far, we have discussed traits that involve just one gene, but few human characteristics are controlled by a single gene. Most traits are polygenic: controlled by more than one gene. Height is one example of a polygenic trait, as are skin color and weight.

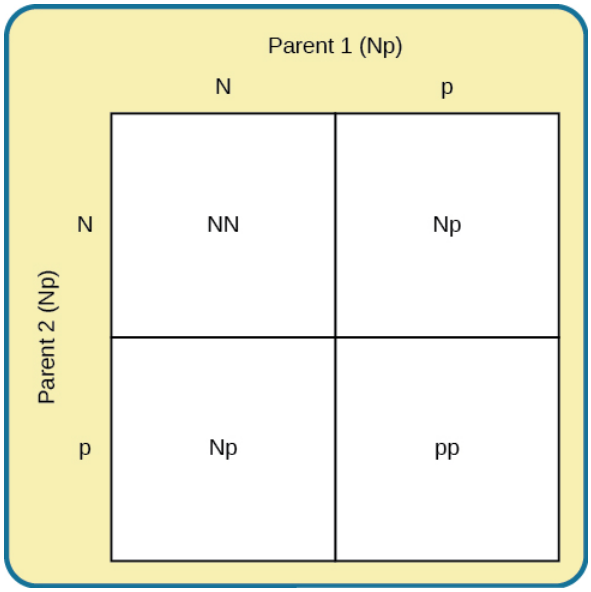


Figure **3.8.12**: In this **Punnett square**, N represents the normal allele, and p represents the recessive allele that is associated with PKU. If two individuals mate who are both heterozygous for the allele associated with PKU, their offspring have a 25% chance of expressing the PKU phenotype.

3.8.0.1 Mutant Genes

Mutation is a permanent change in a segment of DNA.

Mutations are changes in the genetic material of the cell. Substances that can cause genetic mutations are called mutagen agents. Mutagen agents can be anything from radiation from x-rays, the sun, toxins in the earth, air, and water viruses. Many gene mutations are completely harmless since they do not change the amino acid sequence of the protein the gene codes for.

Mutations can be good, bad, or indifferent. They can be good for you because their mutation can be better and stronger than the original. They can be bad because it might take away the survival of the organism. However, most of the time, they are indifferent

because the mutation is no different than the original.

The not so harmless ones can lead to cancer, birth defects, and inherited diseases. Mutations usually happen at the time of cell division. When the cell divides, one cell contracts a defect, which is then passed down to each cell as they continue to divide.

Teratogens refers to any environmental agent that causes damage during the prenatal period. Examples of Common Teratogens:

- drugs: prescription, non-prescription, and illegal drugs
- tobacco, alcohol,
- radiation,
- environmental pollution,
- infectious disease,
- STD's,
- Aids,
- Parasites

The sensitive period to teratogen exposure, in the embryonic period, is most vital and most harmful. By contrast, damage from exposure during the fetal stage is typically minor.

Gene mutations provide one source of harmful genes. As noted above, a mutation is a sudden, permanent change in a gene. While many mutations can be harmful or lethal, once in a while, a mutation benefits an individual by giving that person an advantage over those who do not have the mutation. Recall that the theory of evolution asserts that individuals best adapted to their particular environments are more likely to reproduce and pass on their genes to future generations. In order for this process to occur, there must be competition—more technically, there must be variability in genes (and resultant traits) that allow for variation in adaptability to the environment. If a population consisted of identical individuals, then any dramatic changes in the environment would affect everyone in the same way, and there would be no variation in selection, making evolution impossible. In contrast, diversity in genes and associated traits allows some individuals to perform slightly better than others when faced with environmental change. This creates a distinct advantage for individuals best suited for their environments in terms of successful reproduction and genetic transmission, leading to natural selection and evolutionary change.

3.8.1 Summary

Genes are sequences of DNA that code for a particular trait. Different versions of a gene are called alleles—sometimes alleles can be classified as dominant or recessive. A dominant allele always results in the dominant phenotype. In order to exhibit a recessive phenotype, an individual must be homozygous for the recessive allele. Genes affect both physical and psychological characteristics. Ultimately, how and when a gene is expressed, and what the outcome will be—in terms of both physical and psychological characteristics—is a function of the interaction between our genes and our environments.

3.8.2 Review Questions

A(n) _____ is a sudden, permanent change in a sequence of DNA.

1. allele
2. chromosome
3. epigenetic
4. mutation

_____ refers to a person’s genetic makeup, while _____ refers to a person’s physical characteristics.

1. Phenotype; genotype
2. Genotype; phenotype
3. DNA; gene
4. Gene; DNA

_____ is the field of study that focuses on genes and their expression.

1. Social psychology
2. Evolutionary psychology

3. Epigenetics
4. Behavioral neuroscience

Humans have _____ pairs of chromosomes.

1. 15
2. 23
3. 46
4. 78

3.8.3 Critical Thinking Questions

The theory of evolution by natural selection requires variability of a given trait. Why is variability necessary and where does it come from?

3.8.4 Personal Application Questions

You share half of your genetic makeup with each of your parents, but you are no doubt very different from both of them. Spend a few minutes jotting down the similarities and differences between you and your parents. How do you think your unique environment and experiences have contributed to some of the differences you see?

3.8.4.1 Attributions

Adapted by Kenneth A. Koenigshofer, PhD, from: Wikibooks, Human Physiology/Genetics and inheritance, https://en.wikibooks.org/wiki/Human_...nheritance#DNA; Concepts of Biology, First Canadian Edition, by Charles Molnar and Jane Gair is licensed under a Creative Commons Attribution 4.0 International License, except where otherwise noted. Chapter 9, Introduction to Molecular Biology, 9.1 The Structure of DNA, https://opentextbc.ca/biology/chapte...ucture-of-dna/; and Openstax Psychology 2e Human Genetics (Links to an external site.)

This page titled 3.8: Human Genetics is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

3.9: Genes and Behavior

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe the field of behavioral genetics.
2. Define epigenetics.
3. Describe the major ideas that comprise an evolutionary and computational view of the mind.
4. Explain the steps involved in how genes make proteins including how gene expression is regulated.
5. Describe the types of variation in the genetic code.
6. Discuss the ways in which behavior can influence genes.
7. Describe gene-environment interactions.

3.9.0.1 Overview

Genes, in interaction with the environment, have an enormous influence on the organization of minds and behavior. Each species, with its own unique genetic makeup, has an evolved brain structure and functional organization that determines its innate psychological nature. Human nature is one example, but other species possess their own psychological natures as well, determined by their own brain evolution. Wolves have an innate wolf nature, lions have an innate lion nature, hawks have an innate hawk nature, and so on (Koenigshofer, 2010, 2016). The innate psychological nature of each species is the result of the genes affecting brain organization and function that each species carries in its cells, and the complement of genes that each species possesses is a consequence of the evolutionary history of that species. Behavior genetics studies the effects of genes on behavior and on the information processing operations of the brain, as expressed in behavior. It also investigates the influence of environment and behavior on the expression of genes in the organism--a field called epigenetics (see Module 3.14).

3.9.1 The Influence of Genes on Mind and Behavior

Genetic makeup has a large role in determining human behavior. The influence of genes on behavior has been well established in the scientific community. To a large extent, who we are and how we behave is a result of our genetic makeup. While genes do not strictly determine behavior or our mental processes, they play a huge role in what we do and why we do it.

3.9.1.1 An Evolutionary and Computational View of the Mind

Since we will be talking not only about behavior, but also about the mind, it is best to talk a little bit about what the mind is. Everyone knows something about what the mind is because we each have a mind and we experience our own mental activities such as conscious awareness of our surroundings, feelings, thoughts, and memories each and every moment of our waking lives. You are using some parts of your mind right now as you read this. So, we know the mind as the collection of our own mental processes and experiences.

But there is another way to think about the mind--as a form of biological adaptation, as a collection of information processing solutions to problems of survival and reproduction that have persisted over much of our evolutionary history as a species. On this evolutionary view, the mind is a collection or bundle of information processing "organs" or "modules" each of which has evolved, over our evolutionary history as a species, to process particular kinds of information from the environment in quite specific ways to help us to survive and reproduce our genes. This is a Darwinian or evolutionary model of what the mind is.

This evolutionary model of the mind makes several assumptions (Cosmides & Tooby, 1997): 1) the mind is brain activity (i.e. the mind is what the brain does), 2) the mind/brain of each species, including the human species, has been constructed to take its present day form as a result of evolution by natural selection in that species (therefore, minds just like bodies differ in some ways from species to species, but also like bodies are also similar across species in some ways), 3) this means that the human mind is "hard wired" in many ways by genetic evolution and thus each individual brain comes "hard wired" by genetic information (genes, DNA) which direct brain development from conception onward, 4) this means that each species has its own genetically evolved

psychological nature--applied to humans this means that humans are born with an innate human nature, 5) the mind has many different parts that do many different information processing tasks important to survival and reproduction, especially in the evolutionary past of each species, 6) different parts of our minds are localized in different parts of our brains or in different circuits in our brains (localization of function), 7) the human brain is not a general purpose learning machine but a collection of specialized information processing organs or modules that collectively create in our heads a workable model of reality that guides our behavior successfully toward effective adaptation to the environment, 8) although learning is important in shaping our minds and behavior, its role is secondary and supplementary compared to the much larger role played by our genes in shaping our innate human nature.

The following quote from Steven Pinker (1997) expresses these ideas eloquently:

"The mind is a system of organs of computation, designed by natural selection to solve the kinds of problems our ancestors faced in their foraging way of life, in particular, understanding and outmaneuvering objects, animals, plants, and other people. [This view] can be unpacked into several claims. The mind is what the brain does; specifically, the brain processes information, and thinking is a kind of computation. The mind is organized into modules or mental organs, each with a specialized design that makes it an expert in one arena of interaction with the world. The modules' basic logic is specified by our genetic program. Their operation was shaped by natural selection to solve the problems of the hunting and gathering life led by our ancestors in most of our evolutionary history. The various problems for our ancestors were sub-tasks of one big problem for their genes, maximizing the number of copies that made it into the next generation."

3.9.2 How Genes Affect Body and Brain

Genes are responsible for producing the proteins that run everything in our bodies. Some proteins are visible, such as the ones that compose our hair and skin. Others work out of sight, coordinating our basic biological functions.

For the most part, every cell in our body contains exactly the same genes, but inside individual cells some genes are active while others are not. When genes are active, they are capable of producing proteins. This process is called gene expression. When genes are inactive, they are silent or inaccessible for protein production.

At least a third of the approximately 20,000 different genes that make up the human genome are active (expressed) primarily in the brain. This is the highest proportion of genes expressed in any part of the body. These genes influence the development and function of the brain, and ultimately control how we move, think, feel, and behave.

3.9.2.1 From DNA

In order to understand how genes work in the brain, we have to understand how genes make proteins. This begins with DNA (deoxyribonucleic acid).

DNA is a long molecule packaged into structures called chromosomes. Humans have 23 pairs of chromosomes, including a single pair of sex chromosomes (XX in females and XY in males). Within each pair, one chromosome comes from an individual’s mother and the other comes from the father. In other words, we inherit half of our DNA from each of our parents.

DNA consists of two strands wound together to form a double helix. Within each strand, chemicals called nucleotides are used as a code for making proteins. DNA contains only four nucleotides – adenine (A), thymine (T), cytosine (C), and guanine (G) – but this simple genetic alphabet is the starting point for making all of the proteins in the human body, estimated to be as many as one million. Coding this many proteins with just 4 "letters" is possible because these "letters" can be combined into long chains in an enormous number of different combinations and sequences.

3.9.2.2 To Gene

A gene is a stretch of DNA that contains the instructions for making or regulating a specific protein.

Genes that make proteins are called protein-coding genes. In order to make a protein, a molecule closely related to DNA called ribonucleic acid (RNA) first copies the code within DNA. Then, protein-manufacturing machinery within the cell scans the RNA, reading the nucleotides in groups of three. These triplets encode 20 distinct amino acids, which are the building blocks for proteins. The largest known human protein is a muscle protein called titin, which consists of about 27,000 amino acids.

Some genes encode small bits of RNA that are not used to make proteins, but are instead used to tell proteins what to do and where to go. These are called non-coding or RNA genes. There are many more RNA genes than protein-coding genes.

3.9.2.3 To Protein

Proteins form the internal machinery within brain cells and the connective tissue between brain cells. They also control the chemical reactions that allow brain cells to communicate with each other.

Some genes make proteins that are important for the early development and growth of the infant brain. For example, the ASPM gene makes a protein that is needed for producing new nerve cells (or neurons) in the developing brain. Alterations in this gene can cause microcephaly, a condition in which the brain fails to grow to its normal size.

Certain genes make proteins that in turn make neurotransmitters, which are chemicals that transmit information from one neuron to the next. Other proteins are important for establishing physical connections that link various neurons together in networks.

Still other genes make proteins that act as housekeepers in the brain, keeping neurons and their networks in good working order. Harmful mutations in these genes can cause some neurological diseases such as amytrophic lateral sclerosis (ALS, Lou Gehrig's Disease).

3.9.3 How Gene Expression Is Regulated

We know which protein a gene will make by looking at its code, also called its DNA sequence. What we cannot predict is the amount of protein that will be made, when it will be made, or what cell will make it.

Each cell turns on only a fraction of its genes, while it silences the rest. For example, genes that are expressed in brain cells may be silenced in liver cells or heart cells. Some genes are only turned on during the early months of human development and then are silenced later.

What determines these unique patterns of gene expression? Like people, cells have a unique lineage, and they tend to inherit traits from their parents. So, a cell’s origins influence the genes it turns on to make proteins. The cell’s environment – its exposure to surrounding cells and to hormones and other signals – also helps determine which proteins the cell makes. These cues from a cell’s past and from its environment act through many regulatory factors inside the cell, some of which are described in the following sections.

DNA Binding Proteins

About 10 percent of the genes in the human genome encode DNA binding proteins. Some of these proteins recognize and attach to specific bits of DNA to activate gene expression. Another type of DNA binding protein, called a histone, acts as a spool that can keep DNA in tight coils and thus suppress gene expression.

sRNA

Scattered throughout the genome are many types of small RNA (sRNA) that actively regulate gene expression. Because of their short length, they are able to target, match, and deactivate small bits of genetic code.

Epigenetic Factors

The word epigenetics comes from the Greek word epi, meaning above or beside. In a broad sense, epigenetics refers to long-lasting changes in gene expression without any changes to the genetic code. Epigenetic factors include chemical marks or tags on DNA or on histones that can affect gene expression. We shall discuss more about epigenetics later in this chapter.

3.9.4 Variations In Genetic Code

A genetic variation is a permanent change in the DNA sequence that makes up a gene. Most variations are harmless or have no effect at all. However, other variations can have harmful effects leading to disease. Still others can be beneficial in the long run, helping a species adapt to change.

Single Nucleotide Polymorphism (SNP)

SNPs are variations that involve a change in just one nucleotide. It is estimated that the human genome contains more than 10 million different SNPs. Because SNPs are such small changes within DNA, most of them have no effect upon gene expression. Some SNPs, however, are responsible for giving us unique traits, such as our hair and eye color. Other SNPs may have subtle effects on our risk of developing common diseases, such as heart disease, diabetes, or stroke.

Copy Number Variation (CNV)

At least 10 percent of the human genome is made up of CNVs, which are large chunks of DNA that are deleted, copied, flipped or

otherwise rearranged in combinations that can be unique for each individual. These chunks of DNA often involve protein-coding genes. This means that CNVs are likely to change how a gene makes its protein.

Since genes usually occur in two copies, one inherited from each parent, a CNV that involves a single missing gene could lower the production of a protein below the amount needed.

Having too many copies of a gene can be harmful, too. Although most cases of Parkinson's disease are sporadic (without a known cause), some cases have been linked to having two or more copies of the SNCA gene, which encodes a protein called alpha-synuclein. The excess alpha-synuclein accumulates in clumps inside brain cells, and appears to jam the cells' machinery. For reasons that are not clear, similar clumps are associated with sporadic Parkinson's disease.

Single Gene Mutation

Some genetic variations are small and affect only a single gene. These single gene mutations can have large consequences, however, because they affect a gene's instructions for making a protein. Single gene mutations are responsible for many rare inherited neurological diseases, such as Huntington's disease.

3.9.5 Behavioral Genetics

Behavioral genetics studies heritability of behavioral traits, and it overlaps with genetics, psychology, and ethology (the scientific study of human and animal behavior). Genetics plays a large role in when and how learning, growing, and development occurs. For example, although environment has an effect on the walking behavior of infants and toddlers, children are unable to walk at all before an age that is predetermined by their genome. However, while the genetic makeup of a child determines the age range for when he or she will begin walking, environmental influences determine how early or late within that range the event will actually occur.



Figure **3.9.1**: Gregor Mendel, the father of genetics: Mendel's work with pea plants demonstrated that certain traits follow particular patterns.

3.9.5.1 Classical Genetics

Classical, or Mendelian, genetics examines how genes are passed from one generation to the next, as well as how the presence or absence of a gene can be determined via sexual reproduction. Gregor Mendel is known as the father of the field of genetics, and his work with plant hybridization (specifically pea plants) demonstrated that certain traits follow particular patterns. This is referred to as the law of Mendelian inheritance.

Genes can be manipulated by selective breeding, which can have an enormous impact on behavior. For example, some dogs are bred specifically to be obedient, like golden retrievers; others are bred to be protective, like German shepards. In another example, Seymour Benzer discovered he could breed certain fruit flies with others to create distinct behavioral characteristics and change their circadian rhythms.

3.9.5.2 The Influence of Behavior on Genes

Behavior can influence genetic expression in humans and animals by activating or deactivating genes. Behavior can have an impact on genetic makeup, even as early as the prenatal period. It is important to understand the implications of behavior on genetic makeup in order to reduce negative environmental and behavioral influences on genes.

EEG and PET scans have the ability to show psychologists how certain behaviors trigger reactions in the brain. This has led to the discovery of specific genes, such as those that influence addictive behaviors. A variety of behaviors have been shown to influence gene expression, including—but not limited to—drug use, exposure to the elements, and dietary habits.

3.9.5.2.1 Drugs and Alcohol

Prenatal exposure to certain substances, particularly drugs and alcohol, has detrimental effects on a growing fetus. The most serious consequences of prenatal drug or alcohol exposure involve newborn addiction and fetal alcohol syndrome (FAS). Fetal alcohol syndrome affects both physical and mental development, damaging neurons within the brain and often leading to cognitive impairment and below-average weight. Exposure to drugs and alcohol can also influence the genes of children and adults. Addiction is thought to have a genetic component, which may or may not be caused by a genetic mutation resulting from drug or alcohol use.

3.9.5.2.2 Temperature

Temperature exposure can affect gene expression. For example, in Himalayan rabbits, the genetic expressions of fur, skin, and eyes are regulated by temperature. In the warm areas of the rabbits' bodies, the fur lacks pigment due to gene inactivity and turns white. On the extremities of the rabbits' bodies (nose, ears and feet) the gene is activated and therefore pigmented (usually black).



Figure **3.9.2**: Himalayan rabbit: Exposure to cold temperatures activates pigment-producing genes in the rabbit's extremities.

3.9.5.2.3 Light

Light exposure also influences genetic expression. Thomas Hunt Morgan performed an experiment in which he exposed some caterpillars to light and kept others in darkness. Those exposed to certain light frequencies had corresponding wing colors when they became butterflies (for example, red produced vibrant wing color, whereas blue led to pale wings). Darkness resulted in the palest wing color, leading him to conclude that light exposure influenced the genes of the butterflies. In this manner a caterpillar's behavior can directly affect gene expression; a caterpillar that actively seeks out light will appear different as a butterfly than one that avoids it.

3.9.5.2.4 Nutrition

Lack of proper nutrition in early childhood is yet another factor that can lead to the alteration of genetic makeup. Human children who lack proper nutrition in the first three years of life tend to have more genetic problems later in life, such as health issues and problems with school performance.

3.9.6 Gene-Environment Interactions

Genes do not exist in a vacuum. Although we are all biological organisms, we also exist in an environment that is incredibly important in determining not only when and how our genes express themselves, but also in what combination. Each of us represents a unique interaction between our genetic makeup and our environment; range of reaction is one way to describe this

interaction. **Range of reaction** asserts that our genes set the boundaries within which we can operate, and our environment interacts with the genes to determine where in that range we will fall. For example, if an individual's genetic makeup predisposes her to high levels of intellectual potential and she is reared in a rich, stimulating environment, then she will be more likely to achieve her full potential than if she were raised under conditions of significant deprivation. According to the concept of range of reaction, genes set definite limits on potential, and environment determines how much of that potential is achieved. Some disagree with this theory and argue that genes do not set a limit on a person's potential.

Another perspective on the interaction between genes and the environment is the concept of genetic environmental correlation. Stated simply, our genes influence our environment, and our environment influences the expression of our genes. Not only do our genes and environment interact, as in range of reaction, but they also influence one another bidirectionally. For example, the child of an NBA player would probably be exposed to basketball from an early age. Such exposure might allow the child to realize his or her full genetic, athletic potential. Thus, the parents' genes, which the child shares, influence the child's environment, and that environment, in turn, is well suited to support the child's genetic potential.



Figure **3.9.3**: Nature and nurture work together like complex pieces of a human puzzle. The interaction of our environment and genes makes us the individuals we are. (credit "puzzle": modification of work by Cory Zanker; credit "houses": modification of work by Ben Salter; credit "DNA": modification of work by NHGRI)

In another approach to gene-environment interactions, the field of epigenetics looks beyond the genotype itself and studies how the same genotype can be expressed in different ways. In other words, researchers study how the same genotype can lead to very different phenotypes. As mentioned earlier, gene expression is often influenced by environmental context in ways that are not entirely obvious. For instance, identical twins share the same genetic information (identical twins develop from a single fertilized egg that split, so the genetic material is exactly the same in each; in contrast, fraternal twins develop from two different eggs fertilized by different sperm, so the genetic material varies as with non-twin siblings). But even with identical genes, there remains an incredible amount of variability in how gene expression can unfold over the course of each twin's life. Sometimes, one twin will develop a disease and the other will not. In one example, Tiffany, an identical twin, died from cancer at age 7, but her twin, now 19 years old, has never had cancer. Although these individuals share an identical genotype, their phenotypes differ as a result of how that genetic information is expressed over time. The epigenetic perspective is very different from range of reaction, because here the genotype is not fixed and limited.

Genes affect more than our physical characteristics. Indeed, scientists have found genetic linkages to a number of behavioral characteristics, ranging from basic personality traits to sexual orientation to spirituality (for examples, see Mustanski et al., 2005; Comings, Gonzales, Saucier, Johnson, & MacMurray, 2000). Genes are also associated with temperament and a number of psychological disorders, such as depression and schizophrenia. So while it is true that genes provide the biological blueprints for our cells, tissues, organs, and body, they also have significant impact on our experiences and our behaviors.

Let's look at the following findings regarding schizophrenia in light of our three views of gene-environment interactions. Which view do you think best explains this evidence?

In a study of people who were given up for adoption, adoptees whose biological mothers had schizophrenia *and* who had been raised in a disturbed family environment were much more likely to develop **schizophrenia** or another psychotic disorder than were any of the other groups in the study:

- Of adoptees whose biological mothers had schizophrenia (high genetic risk) and who were raised in disturbed family environments, 36.8% were likely to develop schizophrenia.
- Of adoptees whose biological mothers had schizophrenia (high genetic risk) and who were raised in healthy family environments, 5.8% were likely to develop schizophrenia.
- Of adoptees with a low genetic risk (whose mothers did not have schizophrenia) and who were raised in disturbed family environments, 5.3% were likely to develop schizophrenia.
- Of adoptees with a low genetic risk (whose mothers did not have schizophrenia) and who were raised in healthy family environments, 4.8% were likely to develop schizophrenia (Tienari et al., 2004).

The study shows that adoptees with high genetic risk were especially likely to develop schizophrenia only if they were raised in disturbed home environments. This research lends credibility to the notion that both genetic vulnerability and environmental stress are necessary for schizophrenia to develop, and that genes alone do not tell the full tale.

3.9.7 Summary

Genes are sequences of DNA that code for a particular trait. Different versions of a gene are called alleles—sometimes alleles can be classified as dominant or recessive. A dominant allele always results in the dominant phenotype. In order to exhibit a recessive phenotype, an individual must be homozygous for the recessive allele. Genes affect both physical and psychological characteristics. The mind has evolved as a bundle of information processing modules "designed" by natural selection to increase survival and reproduction. Like other biological traits, the computational modules of the brain/mind are organized by genetic information during brain development. Ultimately, how and when a gene is expressed, and what the outcome will be—in terms of both physical and psychological characteristics—is a function of the interaction between our genes and our environments.

3.9.7.1 KEY TAKEAWAYS

3.9.7.1.1 Key Points

- Classical, or Mendelian, genetics examines how genes are passed from one generation to the next.
- Behavioral genetics examines the role of genetic and environmental influences on animal (including human) behavior.
- There are many ways to manipulate genetic makeup, such as cross-breeding to achieve certain characteristics.
- It is difficult to ascertain whether genetics (“nature”) or the environment (“nurture”) has a stronger influence on behavior. It is generally believed that human behavior is determined by complex interactions of both nature and nurture.
- Drug use, environmental exposure, and eating habits have all been linked to changes in gene expression. While some such influences are harmless or even beneficial, others can be extremely detrimental. Researchers hope to identify these behaviors and their effects.
- EEG and PET scans show psychologists how certain behaviors trigger reactions in the brain, which can lead to the discovery of certain determinant genes, such as those that influence addictive behaviors.
- Exposure of a fetus to alcohol and drugs can lead to a host of developmental problems after birth, the most serious of which is fetal alcohol syndrome.

3.9.7.1.2 Key Terms

- **behavioral genetics**: The field of study that examines the role of genetics in animal (including human) behavior; often involves the nature-versus-nurture debate.
- **ethology**: The scientific study of human and animal behavior.
- **genetics**: The branch of biology that deals with the transmission and variation of inherited characteristics, particularly chromosomes and DNA.
- **gene**: A unit of heredity; a segment of DNA or RNA that is transmitted from one generation to the next and that carries genetic information such as the sequence of amino acids for a protein.
- **fetal alcohol syndrome**: Any of a spectrum of birth defects resulting from excessive alcohol consumption by the mother during pregnancy.

3.9.8 Review Questions

A(n) _____ is a sudden, permanent change in a sequence of DNA.

1. allele
2. chromosome
3. epigenetic
4. mutation

_____ refers to a person's genetic makeup, while _____ refers to a person's physical characteristics.

1. Phenotype; genotype
2. Genotype; phenotype
3. DNA; gene
4. Gene; DNA

_____ is the field of study that focuses on genes and their expression.

1. Social psychology
2. Evolutionary psychology
3. Epigenetics
4. Behavioral neuroscience

Humans have _____ pairs of chromosomes.

1. 15
2. 23
3. 46
4. 78

3.9.9 Critical Thinking Questions

The theory of evolution by natural selection requires variability of a given trait. Why is variability necessary and where does it come from?

3.9.10 Personal Application Questions

You share half of your genetic makeup with each of your parents, but you are no doubt very different from both of them. Spend a few minutes jotting down the similarities and differences between you and your parents. How do you think your unique environment and experiences have contributed to some of the differences you see?

3.9.11 References

Cosmides, L., & Tooby, J. (1997). *Evolutionary psychology: A primer.* Center for Evolutionary Psychology. University of California, Santa Barbara.

Pinker, S. (1997). *How the Mind Works*. Norton. New York.

3.9.12 Attributions

"An Evolutionary and Computational View of the Mind" is original material written by Kenneth A. Koenigshofer, PhD., Chaffey College.

The remainder of this section is adapted by Kenneth A. Koenigshofer, Ph.D. from:

"How Genes affect Body and Brain" adapted from "Brain Basics: Genes at Work in the Brain," by National Institute of Neurological Disorders and Stroke (NINDS), https://www.ninds.nih.gov/Disorders/...nes-Work-Brain, All NINDS-prepared information is in the public domain and may be freely copied.

"Genetics and Behavior" by Boundless.com. **License**: *CC BY-SA: Attribution-ShareAlike*

From Openstax Psychology 2e Human Genetics License: Creative Commons Attribution Non-Commercial

This page titled 3.9: Genes and Behavior is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

CHAPTER OVERVIEW

4: Nervous System Anatomy

> This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

4.1: Nervous System Tissue
4.2: Nervous System Structure and Terminology
4.3: Nervous System Development
4.4: The Central Nervous System
4.5: The Peripheral Nervous System
4.6: Communication and the Endocrine System

This page titled 4: Nervous System Anatomy is shared under a mixed license and was authored, remixed, and/or curated by Naomi Bahm (ASCCC Open Educational Resources Initiative (OERI)) .

4.1: Nervous System Tissue

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Differentiate the central nervous system from the peripheral nervous system
2. Name the two types of cells found in nervous tissue
3. Explain the basic functions of neurons and glia
4. Describe the structures of a neuron
5. Give an example of a type of neuron that is classified by its structure, function, or other characteristic
6. Give an example of a glial cell and its function

4.1.1 Overview

This module presents a brief overview of the nervous system and goes into greater detail on nervous tissue cell types, including neurons (structure, types, and neurogenesis) and glia (types and functions).

In the Blink of an Eye

As you drive into a parking lot, a skateboarder suddenly flies in front of your car across your field of vision (Figure **4.1.1**). You see the skateboarder in the nick of time and react immediately. You slam on the brakes and steer sharply to the right — all in the blink of an eye. You avoid a collision, but just barely. You're shaken up but thankful that no one was hurt. How did you respond so quickly? Such rapid responses are controlled by your nervous system.



Figure **4.1.1**: Skateboarder

4.1.2 Overview of the Nervous System

The **nervous system**, illustrated in Figure **4.1.2**, is the human organ system that coordinates all of the body's voluntary and involuntary actions by transmitting electrical and chemical signals to and from different parts of the body. Specifically, the nervous system extracts information from the internal and external environments using sensory receptors. It then usually sends signals encoding this information to the brain, which processes the information to determine an appropriate response. Finally, the brain sends signals to muscles, organs, or glands to bring about the necessary response. The two main divisions of the nervous system are the central nervous system (CNS), consisting of the brain and the spinal cord, and the peripheral nervous system (PNS), which includes all other nervous tissue, such as ganglia and nerves, outside the brain and spinal cord. The CNS and PNS are covered in greater detail in separate sections. In the example above, your eyes detected the skateboarder, the information traveled to your brain, and your brain instructed your body to act to avoid a collision.

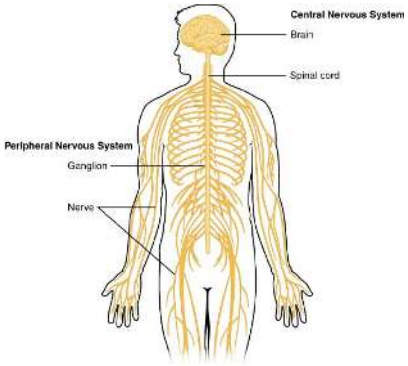


Figure **4.1.2**: Overview of the Nervous System. The two main divisions of the nervous system are the central nervous system (the brain and the spinal cord) and the peripheral nerve system (which includes ganglia and nerves).

4.1.3 Nervous Tissue Cell Types

Nervous tissue is composed of two types of cells: neurons (also called nerve cells) and glia (also called glial cells or neuroglia), as shown in Figure **4.1.3**. Neurons are responsible for the computation and communication that the nervous system provides. They are electrically active and release chemical signals to target cells. Glia are known to play a supporting role for nervous tissue. Ongoing research pursues an expanded role that glial cells might play in signaling, but neurons are still considered the basis of this function. Neurons are important, but without glial support they would not be able to perform their function.

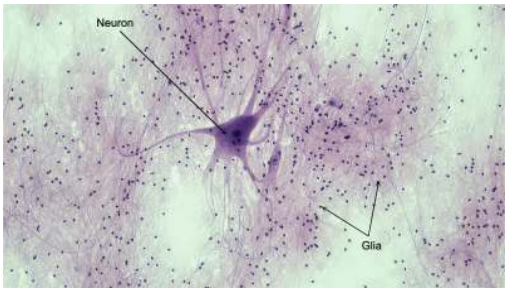


Figure **4.1.3**: Labeled Nervous Tissue Smear. Nervous tissue from the spinal cord shows a neuron and glial cells. Glial cells are smaller and in larger quantities than neurons. The neuron shown is large and has multiple processes coming out of the soma (cell body). The magnification of the picture is 100X.

4.1.3.1 Neurons

Neurons, also called nerve cells, are electrically excitable cells that are the main functional units of the nervous system. Their function is to transmit nerve impulses. They are the only type of human cells that can carry out this function. Neurons are large cells with a high metabolic rate and they depend on a continuous and abundant supply of oxygen and glucose. Neurons are responsible for the electrical signals that communicate information about sensations, produce movements in response to those stimuli, and induce thought processes within the brain. An important part of the function of neurons is in their structure, or shape. The three-dimensional shape of these cells makes the immense numbers of connections within the nervous system possible. In the micrograph of human nervous tissue, Figure **4.1.4**, the neon green structures in the forefront of the image are neurons.

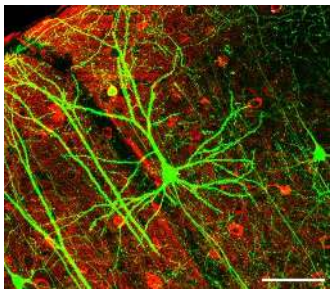


Figure **4.1.4**: Interneurons of adult visual cortex.

4.1.3.1.1 Neuron Structure

The main thing that makes neurons special and differentiates them from other cells in the body is that they have many extensions of their cell membranes, generally referred to as processes. Neurons are usually described as having one, and only one, **axon**—a fiber that emerges from the cell body and projects to target cells. That single axon can branch repeatedly to communicate with many target cells. It is the axon that propagates the nerve impulse (also called an action potential), which is communicated to one or more cells. The other processes of the neuron are **dendrites**, which receive information from other neurons across specialized areas called **synapses**. The dendrites are usually highly branched processes, providing locations for other neurons to communicate with the cell body. Information flows through a neuron from the dendrites, across the cell body, and down the axon.

Figure **4.1.5** shows the structure of a typical neuron. The main parts of a neuron are labeled in the figure and described below.

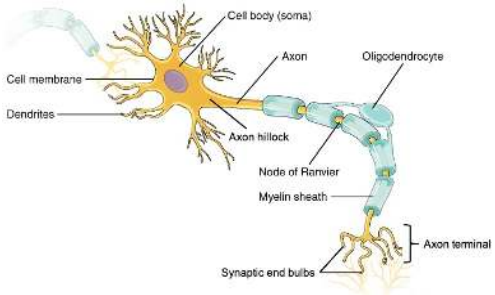


Figure **4.1.5**: Neuron. A somatic motor neuron (in the CNS) with labeled structures: cell membrane, dendrites, cell body (soma), axon, axon hillock, node of Ranvier, myelin sheath, axon terminal, and synaptic end bulbs (also called axon terminal buttons). The oligodendrocyte shown to the side of the axon is a glial cell that forms the myelin sheath surrounding the axon.

- The **cell body** (or soma; soma = "body") is the part of a neuron that contains the nucleus (shown as an oval structure in the center of the cell body, but not labeled) and most of the major organelles. The cell body is usually quite compact, and may not be much wider than the nucleus. The **cell membrane** is the structure that surrounds all the surfaces of the cell (including the dendrites and axon) and separates the inside of the cell from the outside of the cell.
- **Dendrites** are thin structures that are extensions of the cell body. Their function is to receive messages (excitatory and inhibitory post-synaptic potentials, EPSPs/IPSPs- see the nervous system communication chapter) from other cells and carry them to the cell body. A neuron may have many dendrites, and each dendrite may branch repeatedly to form a dendrite “tree” with more than 1,000 **dendritic branches**. **Dendritic spines** (small extensions on the surface of the dendritic branches) further increase surface area for receiving messages, allowing a given neuron to communicate with thousands of other cells.
- The **axon** is a long, thin extension of the cell body. It transmits nerve impulses away from the cell body and toward other cells. The **axon hillock** is a small bulge found at the base of motor neuron axons. The nerve impulse (or action potential) starts from the axon hillock.
- The axon branches at the end, forming the **axon terminal**. Branches of the axon terminal end in **axon terminal buttons** (also called axon endings, synaptic end bulbs, synaptic buttons/boutons, bouton terminaux, etc.) These are the points where the message is transmitted to other cells (via the release of chemicals called neurotransmitters), often to the dendrites of other neurons. A small gap called a **synapse** (also called a synaptic gap or synaptic cleft) is located between the end of the axon terminal and the surface of the receiving cell. An axon may branch hundreds of times, but there is never more than one axon per neuron.
- Many axons (especially the long axons of nerves in the peripheral nervous system) are covered by sections of myelin (also called the **myelin sheath**). The myelin sheath is composed of lipid layers that surround the axon. Myelin is a very good electrical insulator, like the plastic or rubber that encases an electrical cord. Axons that are covered by sections of myelin are called **myelinated**, whereas axons without myelin sheaths are called **unmyelinated**.
- Regularly spaced gaps between sections of myelin occur along the axon. (The gaps are actually much further apart than is shown in the figure- it is necessary to shrink the distance to fit all the structures in a diagram!) These gaps are called **nodes of Ranvier**, and they allow the transmission of nerve impulses along the axon. Nerve impulses jump from node to node in a process called

saltatory conduction, allowing nerve impulses to travel along the axon very rapidly.

- The **oligodendrocyte** shown in the figure is a glial cell that produces myelin sheaths in the central nervous system (brain and spinal cord)- see the Glia section below.

4.1.3.1.2 Classification of Neurons

There are many neurons in the nervous system- an estimated 86 billion in the brain alone (Lent et al., 2012). And there are many different types of neurons, exhibiting a variety of structures and functions. Neurons can be classified depending on their structure, function, or other characteristics.

4.1.3.1.3 Structural Classification

One type of structural classification depends on the number of processes attached to the cell body: one, two or multiple (Figure **4.1.6**). The structures of these three different types of vertebrate neurons support their unique functions, as described below.

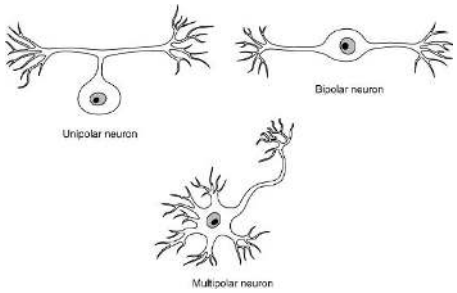


Figure **4.1.6**: Neuron Classification by Structure. Unipolar neurons have one process emerging from the soma which contains both the axon and the dendrite. Bipolar neurons have two processes emerging from the soma: one is an axon and the other is a dendrite. Multipolar neurons have more than two processes emerging from the soma: one is an axon and two or more others are dendrites. (Note: True unipolar neurons are only found in invertebrates. The unipolar neuron pictured is technically a pseudo-unipolar neuron.)

Unipolar neurons have only one process emerging from the cell. True unipolar cells are only found in invertebrate animals, so the unipolar cells in humans are more appropriately called “pseudo-unipolar” cells. (Invertebrate unipolar cells do not have dendrites.) Human unipolar cells have an axon that emerges from the cell body, but it splits so that the axon can extend along a very long distance. At one end of the axon are dendrites, and at the other end, the axon forms synaptic connections with a target. Unipolar neurons are structured in a way that is ideal for relaying information forward. Unipolar cells are exclusively **sensory neurons** (see functional classification below), involved in processes like transmission of physiological information from the body’s periphery (such as communicating body temperature through the spinal cord up to the brain).

Bipolar neurons have two processes which extend from each end of the cell body opposite to each other. One is the axon and the other is the dendrite. Bipolar neurons help acquire and pass sensory information to various centers in the brain, and are not very common. Bipolar cells are are involved in sensory perception such as smell (from the olfactory epithelium), and light (in the retina of the eye). The bipolar cells of the retina are an intermediate layer of neurons between the deepest layer, the visual receptors (photoreceptors, comprised of the rods and cones), and the retinal ganglion cells, the most superficial layer of retinal neurons, whose axons bundle together to form the optic nerve (see the chapter on sensory systems for more information).

Multipolar neurons are all of the neurons that are not unipolar or bipolar, and are also the most common. They have one axon (which can branch forming multiple endings) and two or more dendrites (usually many more), which allows them to communicate with many other neurons. Multipolar neurons convey sensory and motor information in the brain, spinal cord, and throughout the body. With the exception of the unipolar sensory ganglion cells, and the two specific bipolar cells mentioned above, all other neurons are multipolar. Multipolar neurons can have multiple functions and include **motor neurons** (see functional classification below), which carry commands from the brain and spinal cord to muscles and glands, and **interneurons** (see functional classification below), which constitute the majority of neurons. One of the most prominent neuron types is a pyramidal neuron, which falls under the multipolar category. As mentioned below, it gets its name from the triangular or pyramidal shape of its soma (for examples see, Furtak, Moyer, & Brown, 2007).

4.1.3.1.4 Functional Classification

The main functional classification depends on which function the neuron is carrying out: sensation, integration or motor. A mnemonic to remember the functional classification and which direction the nerve impulses travel (towards the CNS= afferent; away from the CNS= efferent) is "***SAME***" (***S***ensory -- ***A***fferent and ***M***otor -- ***E***fferent).

Sensory (also called afferent) **neurons** carry nerve impulses from sensory receptors in tissues and organs to the central nervous system. They convert physical stimuli such as touch, light, and sound into nerve impulses.

Motor (also called efferent) **neurons**, like the one in Figure **4.1.5**, carry nerve impulses from the central nervous system to muscles and glands. These signals then stimulate or inhibit the activity of these structures.

Interneurons carry out integrative functions (such as retrieving, processing, and storing information) and facilitate communication between sensory and motor neurons (inter = "between"). Interneurons are found in the brain and spinal cord, and their processes do not extend outside of the structure they occupy. In other words, a given interneuron will have its dendrites, cell body, and axon all contained within the brain region or spinal cord segment where it is located.

4.1.3.1.5 Other Classifications

Neurons can also be classified on the basis of what they look like, where they are found, who found them, what they do, or even what chemicals (neurotransmitters) they use to communicate with each other

. Some of these classifications will be discussed in greater detail in the communication in the nervous system chapter. As a brief example, a multipolar neuron that has a very important role to play in a pa

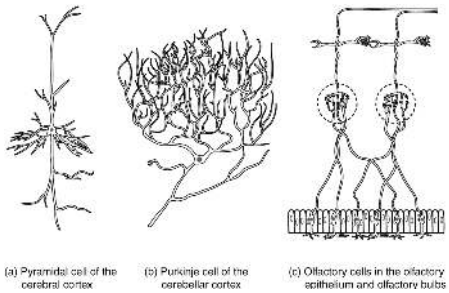


Figure **4.1.7**: Other Neuron Classifications. Three examples of neurons that are named on the basis of other criteria. (a) The pyramidal cell is a multipolar cell with a cell body that is shaped something like a pyramid. (b) The Purkinje cell in the cerebellum was named after the scientist who originally described it. (c) Olfactory cells (located in the olfactory epithelium and olfactory bulbs) are named for the functional group to which they belong.

4.1.3.1.6 Gray Matter and White Matter

The nervous tissue in the brain and spinal cord consists of **gray matter** and **white matter**. Gray matter contains mainly the cell bodies and dendrites of neurons. It is gray only in cadavers; living gray matter is actually more pink than gray (Figure **4.1.8**). White matter consists mainly of myelinated axons, giving it a white color. In the brain, the gray matter is on the outside surface (the **cortex**) and the white matter is on the interior; in the spinal cord it is reversed- the gray matter is inside and the white matter surrounds it.

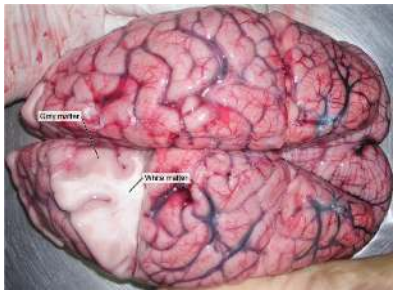


Figure **4.1.8**: White and gray matter. You can see the layers of (pinkish) gray matter and white matter in this photo of a brain from a recently deceased human patient.

Nerves in the peripheral nervous system are also white matter. Nerves consist of long bundles of myelinated axons that extend to muscles, organs, or glands throughout the body. The axons in each nerve are bundled together like wires in a cable. Axons in nerves may be more than a meter long in an adult. The longest nerve runs from the base of the spine to the toes.

4.1.3.1.7 Neurogenesis

Fully differentiated neurons, with all their special structures, cannot divide and form new daughter neurons. Until recently, scientists thought that new neurons could no longer be formed after the brain developed prenatally (see the section on nervous system development). In other words, they thought that people were born with all the brain neurons they would ever have, and as neurons died, they would not be replaced. However, new evidence shows that additional neurons can form in certain areas of the brain, even in adults, from the division of undifferentiated neural stem cells that are found throughout the brain. The production of new neurons is called **neurogenesis.** The extent to which it can occur in adult humans is not known, but it is not likely to be very great.

Recent research indicated that new neurons formed in the amygdala of adult mice (Jhaveri at al., 2018). As the amygdala is important for emotional memory, particularly related to fearful experiences, future research may suggest new treatments for disorders such as post-traumatic stress disorder (PTSD) and depression.

If research on rats is found to apply also to humans, then sustained aerobic exercise such as running can increase neurogenesis in the adult brain, specifically in the hippocampus, a brain structure important for memory as well as learning temporally and/or spatially complex tasks. Although the research is still in the beginning stages, it suggests that exercise may actually lead to a “smarter” brain. Exercise is certainly beneficial for your body, even if it ends up not being confirmed to result in neurogenesis in humans, so it can’t hurt to get more aerobic exercise!

4.1.3.2 Glia

Glia (also called neuroglia or glial cells), are the other type of cell found in nervous tissue. They are considered to be vital supporting cells, and many functions are directed at helping neurons complete their function for communication. Although they do not participate in the communication between cells in the same fashion that neurons do, some researchers have found evidence that suggests that some glia may participate in information processing activities along with neurons (Fields & Stevens-Graham, 2002; Perea, et al., 2014). Some glial functions include digesting debris from dead neurons, carrying nutritional support from blood vessels to the neurons, and helping to regulate the ionic composition of the extracellular fluid.

The name glia comes from the Greek word that means “glue,” and was coined by the German pathologist Rudolph Virchow, who wrote in 1856: “This connective substance, which is in the brain, the spinal cord, and the special sense nerves, is a kind of glue (neuroglia) in which the nervous elements are planted.” Modern research into nervous tissue has shown that there are many deeper roles that these cells play. After his death in 1955, Albert Einstein's brain was studied by scientists worldwide—all wanting to gain insight into the anatomy of a genius. But it wasn't until the 1980s when Marian Diamond reported that Einstein had more glial cells than the average male brain (Diamond et al., 1985). Now it is clear that glia may play a more active role in brain activity, and research may reveal much more about them in the future.

One type of glial cell, radial glia, exist only during prenatal development, guiding migrating neurons to their final destinations. After this task is complete, most of these cells become neurons, but some become astrocytes or oligodendrocytes (Kalat, 2019), described below. There are six types of glial cells in the adult nervous system. Four of them are found in the CNS- astrocytes, oligodendrocytes, microglia and ependymal cells, and two are found in the PNS- satellite cells and Schwann cells. Table **4.1.1** outlines some common characteristics and functions of these glia.

Table **4.1.1**: Glial Cell Types by Location and Basic Function.

CNS glia	PNS glia	Basic function
Astrocyte	Satellite cell	Support of the nervous tissue, form the blood-brain barrier (BBB)
Oligodendrocyte	Schwann cell	Insulation, myelination
Microglia	-	Immune surveillance and phagocytosis
Ependymal cell	-	Creating cerebrospinal fluid (CSF)

4.1.3.2.1 Important Functions of Some Glia

Astrocytes, named for their star-shaped appearance under the microscope (astro = “star”), have many processes extending from their main cell body (not axons or dendrites like neurons, just cell extensions). Those processes interact with neurons, blood vessels, or the connective tissue covering the CNS known as the pia mater. One important role of astrocytes is their contribution to the **blood-brain barrier**. The brain and spinal cord are isolated from the circulation — and most toxins or pathogens in the blood — by the blood-brain barrier, a highly selective membrane that separates the circulating blood from the extracellular fluid in the CNS. The barrier allows water, certain gases, glucose, and some other molecules needed by the brain and spinal cord to cross from the blood into the CNS while keeping out potentially harmful substances. Very little can pass through by diffusion. Most substances that cross the wall of a blood vessel into the CNS must do so through an active transport process. While this barrier protects the CNS from exposure to toxic or pathogenic substances and makes the CNS less susceptible to injury, it also keeps out the cells that could protect the brain and spinal cord from disease and damage (such as white blood cells, one of the body’s main lines of defense). Damage to the CNS is thus likely to have more serious consequences. The blood-brain barrier also causes problems with drug delivery to the CNS. Pharmaceutical companies are challenged to design drugs that can cross the blood-brain barrier as well as have an effect on the nervous system. Aside from finding efficacious substances, the means of delivery is also crucial.

Schwann cells and **oligodendrocytes** are the glia that produce the myelin sheath insulating axons. A single entire Schwann cell wraps around and surrounds a portion of only one axon segment in the peripheral nervous system, whereas oligodendrocytes have processes that reach out to multiple axon segments in the central nervous system. Oligodendrocyte (sometimes called just “oligo”), means “cell of a few branches” (oligo = “few”; dendro = “branches”; cyte = “cell”). Each dendritic process extends from the cell body and wraps around an axon many times to insulate it in myelin. One oligodendrocyte will provide the myelin for many axon segments- 30 to 50 (Kalat, 2019), either along the same axon or by branching out to separate axons.

The diagram (Figure **4.1.9**) shows several types of central nervous system cells associated with two multipolar neurons. Astrocytes are star-shaped cells with many dendrite-like projections but no axon. They are connected with the multipolar neurons and other cells in the diagram through their dendrite-like projections. Ependymal cells have a teardrop shaped cell body and a long tail that branches several times before connecting with astrocytes and the multipolar neuron. Microglial cells are small cells with rectangular bodies and many dendrite-like projections stemming from their shorter sides. The projections are so extensive that they give the microglial cell a fuzzy appearance. The oligodendrocytes have circular cell bodies with dendrite-like projections. Each projection is

connected to a segment of myelin sheath on the axons of the multipolar neurons. In the diagram, the oligodendrocytes are the same color as the myelin sheath segment and are adding layers to the sheath using their projections.

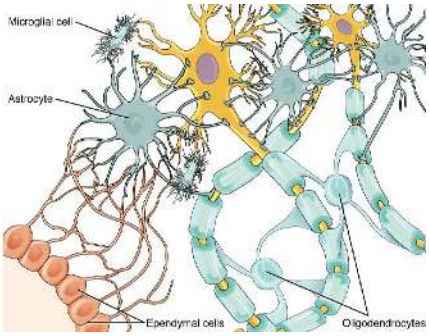


Figure **4.1.9**: Glial Cells of the CNS. This diagram shows several types of central nervous system cells (astrocytes, oligodendrocytes, microglia, and ependymal cells) associated with two multipolar neurons.

4.1.4 Summary

The nervous system coordinates all of the body's voluntary and involuntary actions by transmitting electrical and chemical signals to and from different parts of the body. The two main divisions of the nervous system are the central nervous system (CNS, the brain and the spinal cord), and the peripheral nervous system (PNS, all other nervous tissue in the body).

Nervous tissue contains two major cell types, neurons and glial cells. Neurons are the cells responsible for communication through electrical signals. Glial cells are supporting cells, maintaining the environment around the neurons.

The structures that differentiate neurons from other body cells are the extensions of their cell membranes, namely one axon that projects to target cells, and one or more dendrites, which receive information from other neurons across specialized areas called synapses. The axon propagates nerve impulses (action potentials), which are communicated to one or more cells.

Neurons can be classified depending on their structure, function, or other characteristics. One structural classification is based on the number of processes the neuron has- one (unipolar), two (bipolar) or many (multipolar). One functional classification groups neurons into those that participate in sensation (sensory neurons), integration (interneurons) or motor (motor neurons) functions. Some other ways of classifying neurons include what they look like, where they are found, who found them, what they do, or what neurotransmitters they use.

The nervous tissue in the brain and spinal cord consists of gray matter and white matter. Gray matter contains the cell bodies and dendrites of neurons and white matter contains myelinated axons. Typically, neurons cannot divide to form new neurons. Recent animal research indicates that some limited neurogenesis is possible, but the extent to which this applies to adult humans is unknown.

Several types of glial cells are found in the nervous system, including astrocytes, oligodendrocytes, microglia, and ependymal cells in the CNS, and satellite cells and Schwann cells in the PNS. Astrocytes contribute to the blood-brain barrier that protects the brain. Oligodendrocytes and Schwann cells create the myelin that insulates many axons, allowing nerve impulses to travel along the axon very rapidly.

4.1.5 References

Note: These references are specifically those added to the content of this page by Naomi Bahm and do not include citations from the outside sources used. Refer to the text attributions to locate citations for articles from other sources.

Diamond, M.C., Scheibel, A.B., Murphy, G.M., & Harvey, T. (1985). On the brain of a scientist: Albert Einstein. Experimental Neurology, 88 (1), 198-204. https://doi.org/10.1016/0014-4886(85)90123-2. (https://www.sciencedirect.com/science/article/pii/0014488685901232)

Editors of Encyclopedia Britannica (2021). Jan Evangelista Purkinje. *Encyclopedia Britannica*. https://www.britannica.com/biography...lista-Purkinje. Accessed June 2, 2022.

Fields, R. D., & Stevens-Graham, B. (2002). New insights into neuron-glia communication. *Science*, *298* (5593), 556-562.

Jhaveri, D.J., Tedoldi, A., Hunt, S., Sullivan, R., Watts, N.R., Power, J.M., Bartlett, P.F., & Sah, P. (2018) Evidence for newly generated interneurons in the basolateral amygdala of adult mice. *Molecular Psychiatry 23*, 521–532. doi.org/10.1038/mp.2017.134

Lent, R., Azevedo, F.A., Andrade-Moraes, C.H., & Pinto, A.V. (2012) How many neurons do you have? Some dogmas of quantitative neuroscience under revision. *European Journal of Neuroscience 35*(1),1-9. doi: 10.1111/j.1460-9568.2011.07923.x. Epub 2011 Dec 13. PMID: 22151227.

Perea, G., Sur, M., & Araque, A. (2014). Neuron-glia networks: integral gear of brain function. *Frontiers in cellular neuroscience*, *8*, 378.

4.1.6 Additional Resources

4.1.7 Attributions

Figures:

1. Skateboarder by JESHOOTS-com via Pixabay license
2. Overview of nervous system by OpenStax, licensed CC BY 4.0 via Wikimedia Commons
3. Labeled Nervous Tissue Smear by Chiara Mazzasette adapted from OpenStax, licensed CC BY 4.0 via Wikimedia Commons
4. Interneurons of Adult Visual Cortex by Wei-Chung Allen Lee, Hayden Huang, Guoping Feng, Joshua R. Sanes, Emery N. Brown, Peter T. So, Elly Nedivi, licensed CC BY 2.5 via Wikimedia. Commons
5. Neuron by Chiara Mazzasette adapted from OpenStax, licensed CC BY 4.0 via Wikimedia Commons
6. Neuron Shape Classification by OpenStax is licensed under CC BY 4.0
7. Other Types of Neurons by OpenStax is licensed under CC BY 4.0
8. White and gray matter by OpenStax, licensed CC BY 4.0 via Wikimedia Commons
9. Glial Cells of the CNS by OpenStax is licensed under CC-BY-4.0 (Human Anatomy)

Text adapted from:

1. Introduction to the Nervous System by Suzanne Wakim & Mandeep Grewal, LibreTexts is licensed under CC BY-NC .
2. Neurons by Suzanne Wakim & Mandeep Grewal, LibreTexts is licensed under CC BY-NC .
3. Anatomy of Nervous Tissue by Whitney Menefee, Julie Jenks, Chiara Mazzasette, & Kim-Leiloni Nguyen, LibreTexts is licensed under CC BY .
4. Neurons and Glial Cells by OpenStax, LibreTexts is licensed under CC BY .
5. Central Nervous System by Suzanne Wakim & Mandeep Grewal, LibreTexts is licensed under CC BY-NC .
6. Original material written by Kenneth A. Koenigshofer, Ph.D. is licensed under CC BY 4.0
7. Neurons by Sharon Furtak, licensed CC BY-NC-SA 4.0 via Noba Project (originally curated by Kenneth Koenigshofer, as immediately above).

Changes: Text (and images) from sources 1 through 5 pieced together with some modifications, transitions and additional content added by Naomi I. Gribneau Bahm, PhD., Psychology Professor at Cosumnes River College, Sacramento, CA. Information also incorporated from content (sources 6 and 7) curated by Kenneth A. Koenigshofer, PhD., Chaffey College, specifically in the sections on Structural Classification (of Neurons) and Glia.

This page titled 4.1: Nervous System Tissue is shared under a mixed license and was authored, remixed, and/or curated by Naomi Bahm (ASCCC Open Educational Resources Initiative (OERI)) .

- **11.2: Introduction to the Nervous System** by Suzanne Wakim & Mandeep Grewal is licensed CK-12. Original source: https://www.ck12.org/book/ck-12-human-biology/.
- **11.3: Neurons** by Suzanne Wakim & Mandeep Grewal is licensed CK-12. Original source: https://www.ck12.org/book/ck-12-human-biology/.
- **11.3: Anatomy of Nervous Tissue** by Whitney Menefee, Julie Jenks, Chiara Mazzasette, & Kim-Leiloni Nguyen is licensed CC BY 4.0.
- **11.5: Central Nervous System** by Suzanne Wakim & Mandeep Grewal is licensed CK-12. Original source: https://www.ck12.org/book/ck-12-human-biology/.

4.2: Nervous System Structure and Terminology

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe anatomical position and describe the location of at least two structures using anatomical direction terms
2. Explain the difference between frontal (/coronal), transverse (/horizontal), and sagittal sectional planes
3. Understand the six common divisions of the nervous system
4. List the four major regions of the central nervous system
5. Explain the structures that protect the brain and spinal cord
6. Describe the components of the ventricular system in the brain and spinal cord

4.2.1 Overview

This module starts with anatomical position, sectional planes and directional terms and continues with the divisions of the nervous system. Some areas and structures of the central nervous system (brain and spinal cord) are then discussed, including the four major regions, the protective coverings, and the ventricular system.

4.2.2 Anatomical Orientation and Directions

4.2.2.1 Anatomical Position

When anatomists or health professionals identify the location of a structure in the human body, they do so in reference to a body in anatomical position. That is, they figure out the location based on the assumption that the body is starting out in anatomical position. Anatomical position for a human is when the human is standing up, facing forward, with arms extended to the sides, and palms facing out, as illustrated by two human models in Figure **4.2.1**.



Figure **4.2.1**:

Female (left) and male (right) human body in anatomical position seen from anterior (upper) and posterior (lower) views.

When referencing a structure that is on one side of the body or the other, we use the terms “anatomical right” and “anatomical left.” Anatomical right means that the structure is on the side that a person in anatomical position would consider their right-hand side (not necessarily on the right of the viewer) and anatomical left means that the structure is the side that a person in anatomical position would consider their left-hand side (which likewise is not necessarily the left side of the viewer.)

4.2.2.2 Anatomical Planes

To view the interior of a body, we expose the organs and structures that are visible when that body is cut open along one of three commonly used sectional planes. These planes are the different directions a body is cut to reveal different views of its internal structures.

- **Frontal** (or **coronal**) **plane—**A vertical cut that separates the front from the back of the individual/structure.
- **Transverse** (or **horizontal**) **plane—**A horizontal cut that separates the top from the bottom of the individual/structure. (May also be called cross-sections.)
- **Sagittal plane—**A vertical cut that separates the left half from the right half of the individual/structure.

- **Midsagittal** (or **median**) **plane—**A vertical cut down the exact center line of the individual/structure that separates the left half from the right half.
- **Parasagittal plane—**A vertical cut that is off-center that separates the left of the individual/structure from the right in unequal portions. It does not matter whether it is the left side or the right side that is larger, as long as they are not equal.

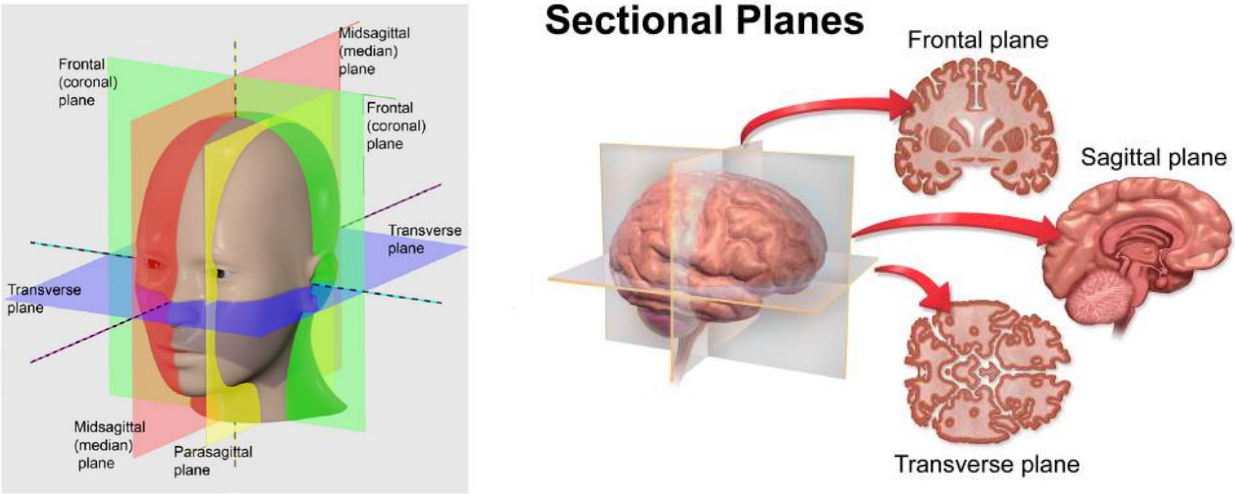


Figure 4.2.2: The different sectional planes used to expose internal structures, and examples of these planes through the human brain.

4.2.2.3 Anatomical Directional Terms

To be able to direct others to specific anatomical structures, or to find structures based on someone else's directions, it is useful to have specific pairs of terms that allow you to orient your search with respect to the location of another known structure. The following pairs of terms are used to make comparisons. Each term is used to orient a first structure or feature with respect to the position of a second structure or feature. All of these terms refer to the body in anatomical position, and are illustrated in Figure 4.2.3.

- **Superior/Inferior**–Equivalent to above (superior) versus below (inferior) when moving along the long axis of the body. A structure that is superior to another is above the second structure. A feature that is inferior to another is below the second feature.
- **Proximal/Distal**–Equivalent to near (proximal) versus far (distal). Usually used to orient the positions of structures and features along the limbs with respect to the trunk of the body. A feature that is proximal to something else is closer to the limb's point of attachment to the trunk. A structure that is distal to something else is farther away from the limb's point of attachment. Although less precise, these terms are occasionally used in the trunk of the body itself to indicate whether something is closer to (proximal) or farther away from (distal) something else.
- **Medial/Lateral**–Equivalent to towards the middle (medial) versus towards the edge (lateral). These terms are used with respect to the midline of the trunk of the body. A structure that is medial to another is closer to the midline of the body's trunk. A feature that is lateral to another is farther away from the midline of the trunk.
- **Anterior/Posterior**–Equivalent to the front (anterior) versus the back (posterior) of the body. A structure that is anterior to another is closer to the front of the body. A feature that is posterior to another is closer to the back of the body.
- **Ventral/Dorsal**–Equivalent to belly-side (ventral) versus back-side (dorsal) of the body. For a human in anatomical position, this pair of terms is largely equivalent to anterior versus posterior. However, for the head and brain, dorsal is on the top (superior) and ventral is on the bottom (inferior), whereas the front of the face would be anterior and the back of the head would be posterior. These terms match four-legged animals in what is considered their anatomical position, where the belly-side is not equivalent to the front of the animal. A structure below the human head that is ventral to another is closer to the belly-side of the body, whereas a feature that is dorsal to another is closer to the back of the body.
- **Superficial/Deep**–Equivalent to closer to the surface (superficial) versus farther from the surface (deep). A structure that is superficial to another is closer to the surface of the body. A feature that is deep to another is farther from the surface of the body.

- (**Cephalic/Caudal**–Equivalent to closer to the head (cephalic) versus closer to the tail (caudal). These terms are more useful for four-legged animals with tails than for upright humans with only a few vestigial tail bones and no visible tail- an area which is also not at the end of the body like the tail is in four-legged animals.)

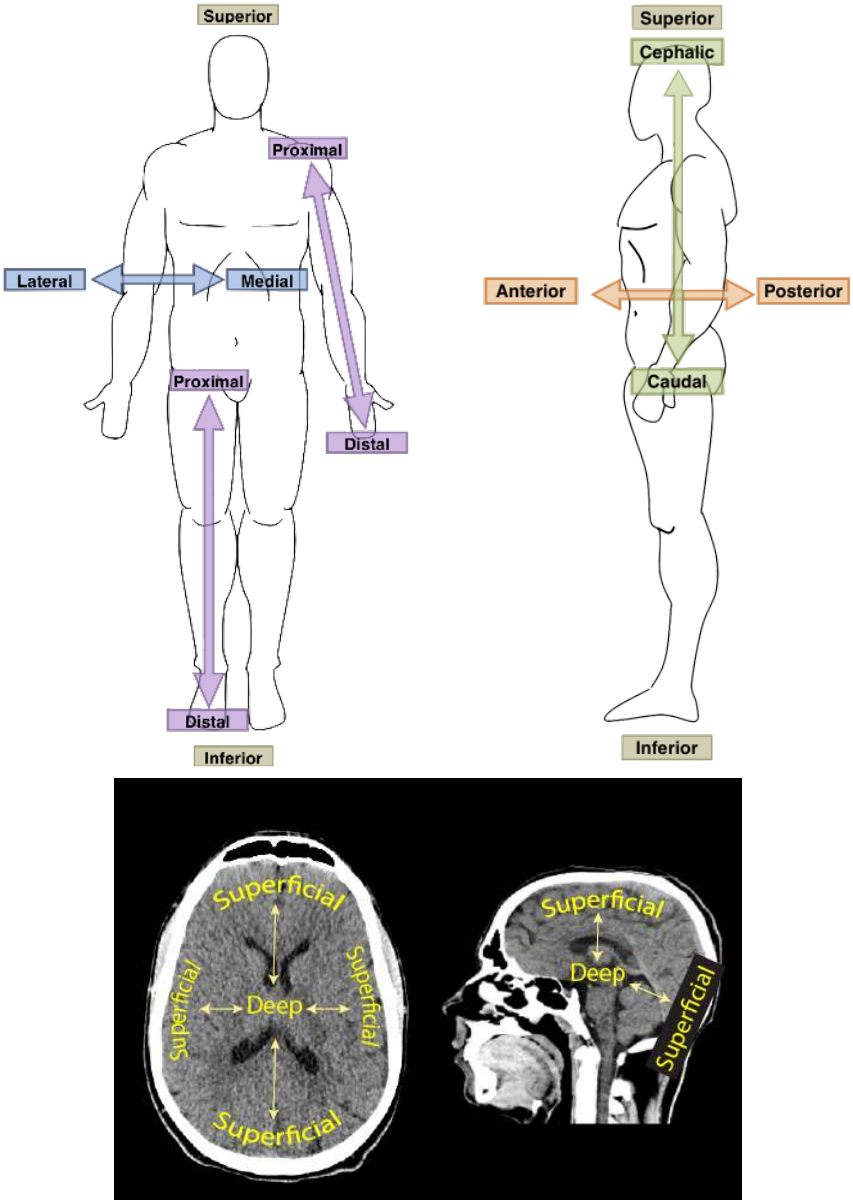


Figure **4.2.3**: Pairs of terms providing anatomical direction or orientation, shown on body outlines in the first image, and in a cross-section (left) and sagittal section (right) of the brain in the second image.

4.2.3 Organization of the Nervous System

As you might predict, the human nervous system is very complex. It has multiple divisions, beginning with its two main parts, the **central nervous system** (CNS) and the **peripheral nervous system** (PNS), as shown in Figure **4.2.4**. The CNS includes the brain and spinal cord, and the PNS consists of all other nervous tissue in the body- mainly nerves, which are bundles of axons from sensory and motor neurons. The nerves of the PNS connect the CNS to the rest of the body.

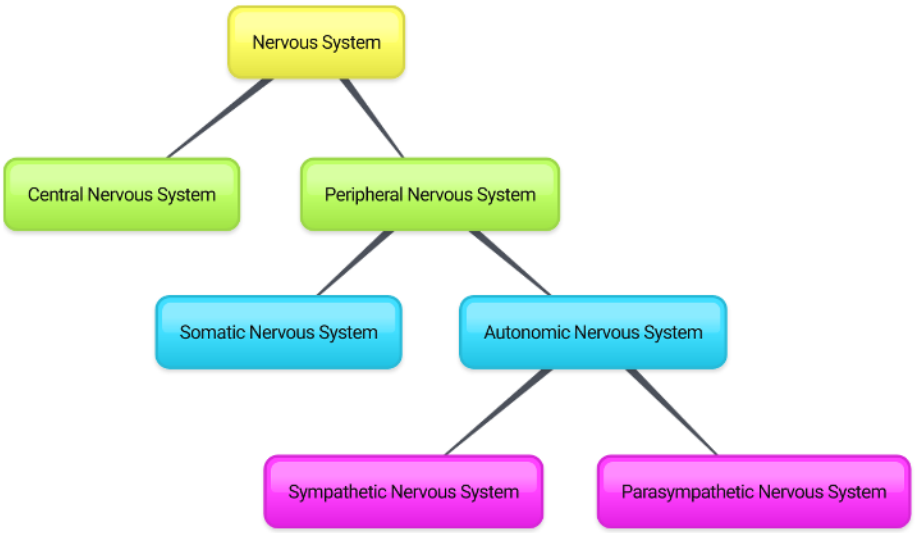


Figure **4.2.4**: This concept map illustrates that the nervous system is divided into the central and peripheral systems. The peripheral nervous system is further divided into the somatic and autonomic systems. The autonomic nervous system is further divided into the sympathetic and parasympathetic systems.

The PNS is divided into two major parts, called the autonomic and somatic nervous systems. The **somatic nervous system** (SNS) is responsible for activities that are under voluntary control and awareness, such as turning a steering wheel or feeling water splashing on your hands. The **autonomic nervous system** (ANS) controls involuntary activities, such as digesting a meal or regulating your heartbeat. The autonomic nervous system also has two divisions: the sympathetic and parasympathetic nervous systems. The **sympathetic nervous system** controls the "fight-or-flight" response during emergencies, and the **parasympathetic nervous system** controls the routine “housekeeping” functions of the body (sometimes referred to as the "rest and digest" division). Generally speaking, your body is usually activating a balance of sympathetic and parasympathetic functions.

The CNS and the sympathetic and parasympathetic branches of the ANS (itself a division of the PNS) are described in greater detail in other modules of this chapter. The other division of the PNS, the SNS, is composed largely of sensory receptors and their connections (covered in the sensory system chapter) and structures related to voluntary motor control (addressed in the movement chapter).

4.2.4 Central Nervous System Regions

Some commonly used terms referring to regions in the central nervous system (such as forebrain, midbrain, and hindbrain) derive from embryonic development and are discussed separately (see the module on development of the nervous system).

In its fully formed state, the CNS may be described in terms of four major regions: the cerebrum, the brainstem (comprised of the diencephalon, the midbrain, the pons, and the medulla- note that some definitions separate the diencephalon out as an additional separate region- Basinger and Hogg, 2021), the cerebellum, and the spinal cord. These regions are colored in on an MRI of an adult in Figure **4.2.5**.

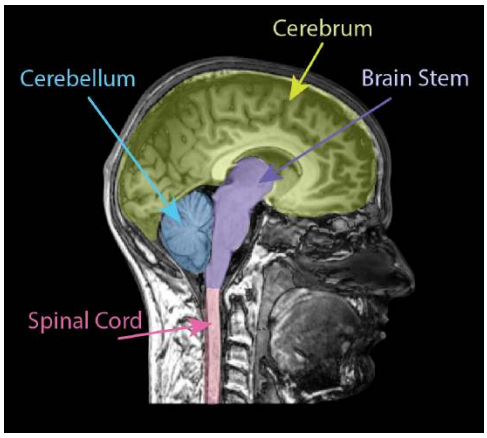


Figure 4.2.5: Regions of the adult central nervous system: The cerebrum, cerebellum, brainstem (including the diencephalon) and spinal cord.

4.2.5 Protective Coverings of the Central Nervous System

The central nervous system (CNS) is crucial to the operation of the body and any compromise of function in the brain and spinal cord can lead to severe difficulties. Both the brain and spinal cord are protected by multiple structures. First, the bones of the skull enclose and house the brain, and the bones of the vertebral column enclose and protect the spinal cord. Figure 4.2.6 shows a segment of the spinal cord within its protective vertebra.

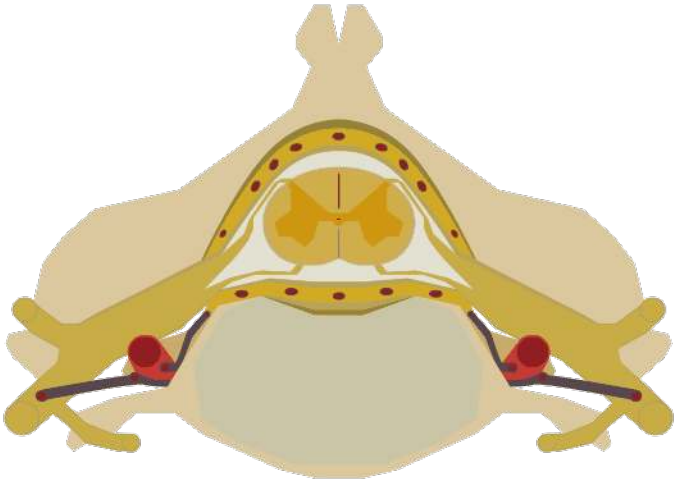


Figure 4.2.6: Cross-section of a spinal cord segment shown encased in its vertebra. The back of the person is at the top of the image.

Underneath the skeletal structures, the brain and spinal cord are surrounded and protected by tough **meninges**, a three-layer protective sheath made of connective tissue that surrounds, supports, stabilizes and partitions the nervous tissue. The meninges also contain cushioning **cerebrospinal fluid** (CSF). The three meningeal layers are the **dura mater**, the **arachnoid mater**, and the **pia mater**. Figure 4.2.7 shows the three layers of meninges covering the spinal cord and Figure 4.2.8 shows all the protective structures of the brain. The subdural cavity of the spinal cord and the subdural space of the brain lie between the dura mater and the arachnoid mater, whereas the subarachnoid cavity of the spinal cord and the subarachnoid space of the brain lie between the arachnoid mater and the pia mater. The **superior sagittal sinus** is a space between layers of dura mater immediately superior to the **longitudinal fissure** (the separation between the left and right hemispheres) of the brain.

In the spinal cord, the posterior (/dorsal) and anterior (/ventral) roots converge to form the **spinal nerve**. The **spinal ganglion** (also known as the **dorsal root ganglion**), an enlargement of the posterior (/dorsal) root, encases sensory neuron cell bodies.

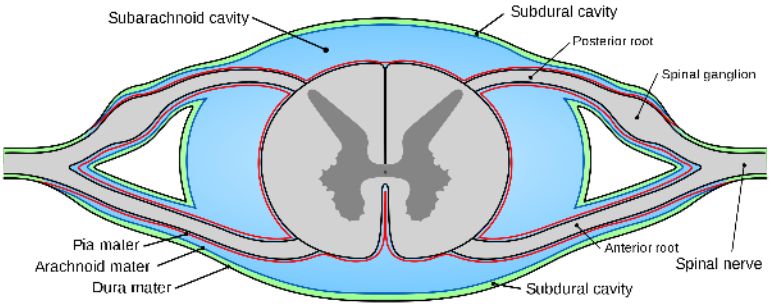


Figure 4.2.7: Cross-section of a spinal cord segment with the meninges and other structures labeled.

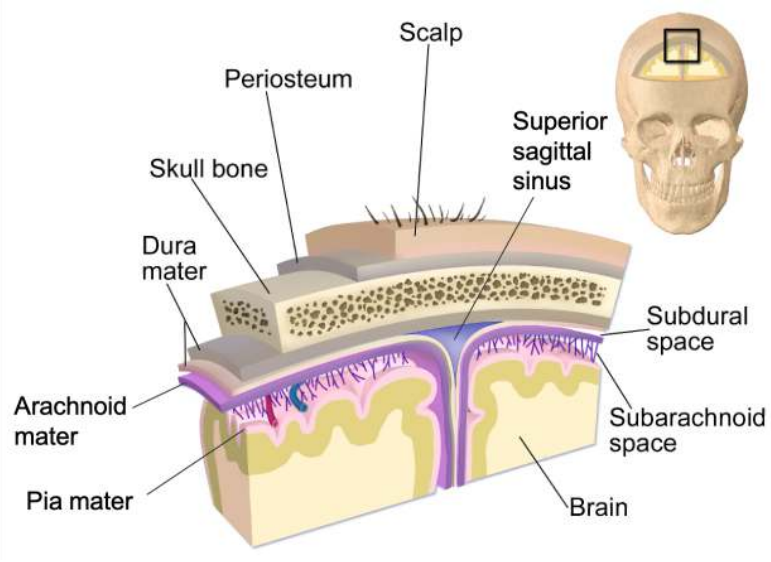


Figure 4.2.8: Protective structures of the brain. Superficially, the skin of the scalp and components of the skull bone create the first layer of protection. Deep to these layers, the meninges (the dura mater, arachnoid mater and pia mater) cover and partition the brain.

4.2.6 Ventricular System

There are four **ventricles** within the brain, all of which developed from the original hollow space within the neural tube, the **central canal**. The ventricles are lined with **ependymal cells** (a type of glia) that produce cerebrospinal fluid (CSF). The first two ventricles are named the **lateral ventricles** and are deep within the brain. The two lateral ventricles are shaped like a letter "C" and are located in the left and right hemispheres. (They were originally referred to as the first and second ventricles.) These ventricles are connected to the **third ventricle** by two openings called the **interventricular foramina** (plural; "foramen" is the singular term). The third ventricle opens into a canal called the **cerebral aqueduct** that passes through the midbrain and connects the third ventricle to the **fourth ventricle**. The fourth ventricle is the space between the cerebellum and the pons and upper medulla. From the fourth ventricle, CSF continues down the central canal of the spinal cord. Figure 4.2.9 illustrates the ventricles, from both a lateral and anterior viewpoint. CSF circulates through the brain and the spinal cord within these structures.

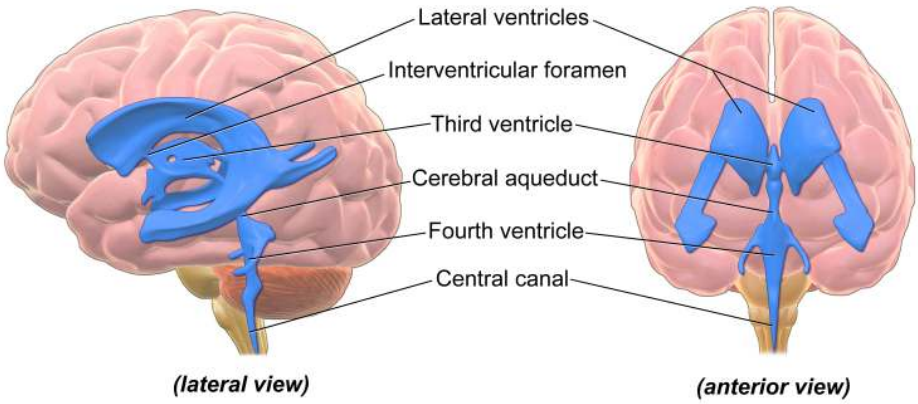


Figure **4.2.9**: Ventricles. Deep within the brain are the lateral ventricles, which are connected to the third ventricle by two openings called interventricular foramina. The third ventricle opens into a canal called the cerebral aqueduct which connects to the fourth ventricle. The fourth ventricle is connected to the central canal of the spinal cord.

4.2.7 Summary

Anatomical position for humans is standing up, facing forward, with arms extended to the sides, and palms facing out. Anatomical planes are the different directions a body is cut to reveal different views of its internal structures, which include frontal (coronal) planes, separating the front from the back, transverse (horizontal) planes, separating the top from the bottom, and sagittal planes, separating the left and right halves. A midsagittal (or median) plane is exactly on the center line and a parasagittal plane divides the left and right halves unevenly. Pairs of anatomical directional terms are used to make comparisons between one structure with respect to the position of a second structure. These include superior (above) versus inferior(below), proximal (near) versus distal (far), medial (towards the middle) versus lateral (towards the edge, anterior (front) versus posterior (back), ventral (belly-side) versus dorsal (back-side), and superficial (closer to the surface) versus deep (farther from the surface).

The organization of the nervous system includes six divisions. The CNS includes the brain and spinal cord, and the PNS consists of all other nervous tissue in the body. The nerves of the PNS connect the CNS to the rest of the body. The PNS is further divided into the autonomic and somatic nervous systems. The somatic nervous system (SNS) is responsible for activities that are under voluntary control and awareness. The autonomic nervous system (ANS) controls involuntary activities. The autonomic nervous system also has two divisions: the sympathetic and parasympathetic nervous systems. The sympathetic nervous system controls response during emergencies, and the parasympathetic nervous system controls the routine functions of the body.

The adult CNS is separated into four major regions: the cerebrum, the brainstem, the cerebellum, and the spinal cord. Anatomical structures help to protect and isolate the CNS, including the skull (surrounding the brain) and the vertebral column (surrounding the spinal cord). Layers of connective tissue called meninges support and stabilize the brain and spinal cord, as well as partition the brain into specific regions. The outer layer is the dura mater, the middle layer is the arachnoid mater and the inner layer is the pia mater.

The ventricular system is composed of four fluid-filled ventricles in the brain (two lateral ventricles, the third ventricle and the fourth ventricle) which connect to each other (via the interventricular foramina and the cerebral aqueduct) and the central canal of the spinal cord. Cerebrospinal fluid (CSF) circulates through the brain and the spinal cord within these structures.

4.2.8 References

Basinger H., & Hogg, J.P. (2021) Neuroanatomy, Brainstem. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK544297/

4.2.9 Additional Resources

If you are curious about anatomical terms used to describe different areas of the body, see "Anatomical Vocabulary" on: " Anatomical Position and Planes" by LibreTexts, licensed under CC BY-SA .

For more detail on the embryological development of the brain and the primary and secondary brain vesicles, see: " The Embryologic Perspective" by OpenStax, LibreTexts, licensed under CC BY .

For more detail on the meninges and the ventricles, see: " Support and Protection of the Brain" by Whitney Menefee, Julie Jenks, Chiara Mazzasette, & Kim-Leiloni Nguyen, LibreTexts, licensed under CC BY and " Circulation and the Central Nervous System" by OpenStax, LibreTexts, licensed under CC BY .

4.2.10 Attributions

1. Figures:
    1. Human Body by Mikael Häggström derivative work: Lamilli (talk)Taken at City Studios in Stockholm (www.stockholmsfotografen.se), September 29, 2011, with assistance from KYO (The organisation of life models) in Stockholm. Both models have consented to the licence of the image, and its usage in Wikipedia. Images uploaded by Mikael Häggström., Public domain, via Wikimedia Commons.
    2. Human Head Anatomical Planes by Jon Richfield is licensed under CC BY-SA 4.0 via Wikimedia Commons; modified labels from " Anatomical Position and Planes" by LibreTexts is licensed under CC BY-SA . And Sectional Planes of the Brain by Bruce Blausen is licensed under CC BY 3.0 via Wikimedia Commons. [Blausen.com staff (2014). "Medical gallery of Blausen Medical 2014". WikiJournal of Medicine 1 (2). DOI:10.15347/wjm/2014.010. ISSN 2002-4436.]
    3. CC-BY-SA, Osteomyoamare, WikiMedia. (both views of entire body).And CC-BY-SA, Marshall Strother, WikiMedia; modified labels from " Anatomical Position and Planes" by LibreTexts is licensed under CC BY-SA . Illustration of deep/superficial terms adapted by Suzanne Wakim from CT of a normal brain, axial 18.png by Mikael Häggström, M.D, licensed CC0 1.0 via Wikimedia Commons.
    4. Nervous System Flowchart by Suzanne Wakim dedicated CC0. From " Introduction to the Nervous System" by Suzanne Wakim & Mandeep Grewal, LibreTexts is licensed under CC BY-NC .
    5. Adapted by Suzanne Wakim from MRI brain sagittal section.jpg by everyone's idle, licensed CC BY-SA 2.0 via Wikimedia Commons.
    6. Time3000 is licensed under CC BY-SA 2.5 via Wikimedia Commons
    7. Mysid, Public domain, via Wikimedia Commons
    8. "Cranial meninges" by Chiara Mazzasette is licensed under CC BY 4.0 / A derivative of Blausen 0110 BrainLayers.png
    9. "Blausen 0896 Ventricles Brain" by BruceBlaus is in the Public Domain, CC0
2. Text adapted from:
    1. " Anatomical Position and Planes" by LibreTexts is licensed under CC BY-SA .
    2. " Introduction to the Nervous System" by Suzanne Wakim & Mandeep Grewal, LibreTexts is licensed under CC BY-NC .
    3. " The Central Nervous System" by OpenStax, LibreTexts is licensed under CC BY .
    4. " Support and Protection of the Brain" by Whitney Menefee, Julie Jenks, Chiara Mazzasette, & Kim-Leiloni Nguyen, LibreTexts is licensed under CC BY .
    5. " Central Nervous System" by Suzanne Wakim & Mandeep Grewal, LibreTexts is licensed under CC BY-NC .
3. Changes: Text (and most of the images) from above four sources pieced together with some modifications, transitions and additional content (and images) added by Naomi I. Gribneau Bahm, PhD., Psychology Professor at Cosumnes River College, Sacramento, CA.

This page titled 4.2: Nervous System Structure and Terminology is shared under a mixed license and was authored, remixed, and/or curated by Naomi Bahm (ASCCC Open Educational Resources Initiative (OERI)) .

- **11.2: Introduction to the Nervous System** by Suzanne Wakim & Mandeep Grewal is licensed CK-12. Original source: https://www.ck12.org/book/ck-12-human-biology/.
- **12.2: Support and Protection of the Brain** by Whitney Menefee, Julie Jenks, Chiara Mazzasette, & Kim-Leiloni Nguyen is licensed CC BY 4.0.

4.3: Nervous System Development

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Explain how the neural tube forms
2. Describe the growth and differentiation of the anterior neural tube into primary and secondary vesicles
3. Understand the mechanisms of postnatal brain development
4. Describe the development of the posterior neural tube into the adult spinal cord
5. Identify at least four stages of neuron development

4.3.1 Overview

This module starts with the value of an embryologic perspective, and then discusses the formation of the neural tube, embryonic brain development (the process of the anterior neural tube differentiates into primary and secondary vesicles), postnatal brain development, spinal cord development, and neuron development.

4.3.2 An Embryologic Perspective

The brain is a complex organ composed of gray and white matter, which can be hard to distinguish. Starting from an embryologic perspective allows you to understand more easily how the parts relate to each other. The embryonic nervous system begins as a very simple structure—essentially a plate of tissue, which then gets increasingly complex. Looking at the development of the nervous system through a few early snapshots makes it easier to understand the whole complex system. Many structures that appear to be adjacent in the adult brain are not connected, and the connections that exist may seem arbitrary. But there is an underlying order to the system that comes from how different parts develop. By following the developmental pattern, it is possible to distinguish the major regions of the nervous system.

4.3.3 Neural Tube

To begin, a sperm cell and an egg cell fuse to become a fertilized egg. The fertilized egg cell, or zygote, starts dividing to generate the cells that make up an entire organism. Sixteen days after fertilization, the developing embryo's cells belong to one of three germ layers that give rise to the different tissues in the body (Figure **4.3.1**). The endoderm, or inner tissue, is responsible for generating the lining tissues of various spaces within the body, such as the mucosae of the digestive and respiratory systems. The mesoderm, or middle tissue, gives rise to most of the muscle and connective tissues. Finally the ectoderm, or outer tissue, develops into the integumentary system (the skin) and the nervous system. It is probably not difficult to see that the outer tissue of the embryo becomes the outer covering of the body. But how is it responsible for the nervous system?

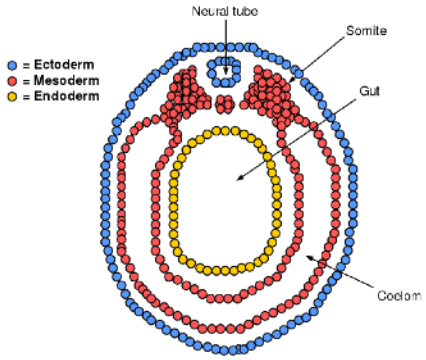


Figure **4.3.1**: Cross section of a vertebrate embryo in the neurula stage. This stage is characterized by development of the nervous system through the formation of the neural tube, shown as a small circle of ectoderm tissue at the top of the image, just inside the large outer circle of ectoderm tissue. The neural tube later develops into the brain and the spinal cord. The middle horseshoe-shaped loop of two layers of mesoderm tissue has clumps of cells at the upper closed ends representing somites. The somites become the vertebral column and the skeletal muscles. The coelom is the open space inside the two layers of mesoderm forming the loop-internal organs develop here. The innermost circle is endoderm tissue; the gut is enclosed in the center.

As the embryo develops, a portion of the ectoderm differentiates into tissue that will become the nervous system. Molecular signals induce cells in this region to form a **neural plate**. The cells then begin to change shape, causing the tissue to buckle and fold inward (Figure **4.3.2**). A **neural groove** forms, visible as a line along the dorsal surface of the embryo. The ridge-like edge on either side of the neural groove is referred to as the **neural fold**. As the neural folds come together and converge, the underlying structure forms into a tube called the **neural tube,** located just beneath the remaining ectoderm (which will form the epidermis).

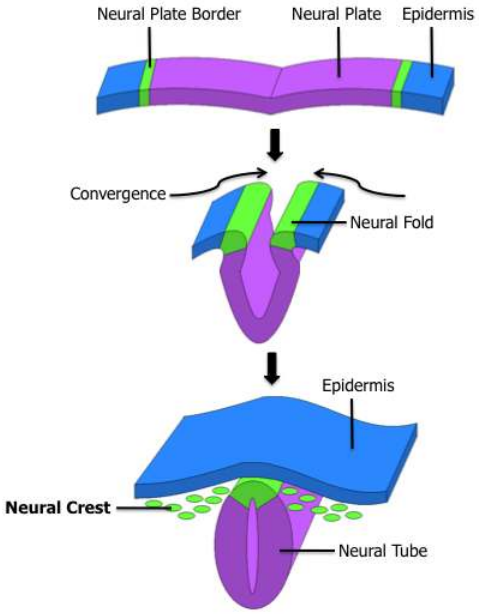


Figure **4.3.2**: Formation of the Neural Tube. The ectoderm layer forms the epidermis and the nervous system. The neural plate (with a border on either side separating that tissue from tissue that forms the epidermis) begins to fold inward to form the neural groove (visible in the middle drawing but not labeled), with the neural plate borders becoming the neural folds on either side. As the two sides of the neural groove converge, the neural folds fuse together and form the neural tube, which lies beneath the epidermis tissue (similarly merged to form a continuous outer plate). The anterior end of the neural tube will develop into the brain, and the posterior portion will become the spinal cord. The neural crest (formed from part of the neural folds) sits laterally to the neural tube and develops into structures in the peripheral nervous system.

At this point, the early nervous system is a simple, hollow tube. It runs from the anterior end of the embryo to the posterior end. Beginning at 25 days, the anterior end develops into the brain, and the posterior portion becomes the spinal cord. This is the most basic arrangement of tissue in the nervous system, and by the fourth week of development it gives rise to the more complex structures.

4.3.4 Brain Development

4.3.4.1 Embryonic Brain Development

4.3.4.1.1 Overview

The vertebrate brain has three major regions based on embryonic development: the forebrain (including the cerebrum, thalamus, hypothalamus, and limbic system structures), the midbrain, and the hindbrain (including the medulla, pons, and cerebellum). Figure **4.3.3** shows the embryologic vesicles (originating from bulges in the neural tube, explained in greater detail below) that contribute to these regions. The prosencephalon (forebrain) is composed of the telencephalon and the diencephalon, the mesencephalon is the midbrain, and the rhombencephalon (hindbrain) is composed of the metencephalon and the myelencephalon. The spinal cord extends below the hindbrain.

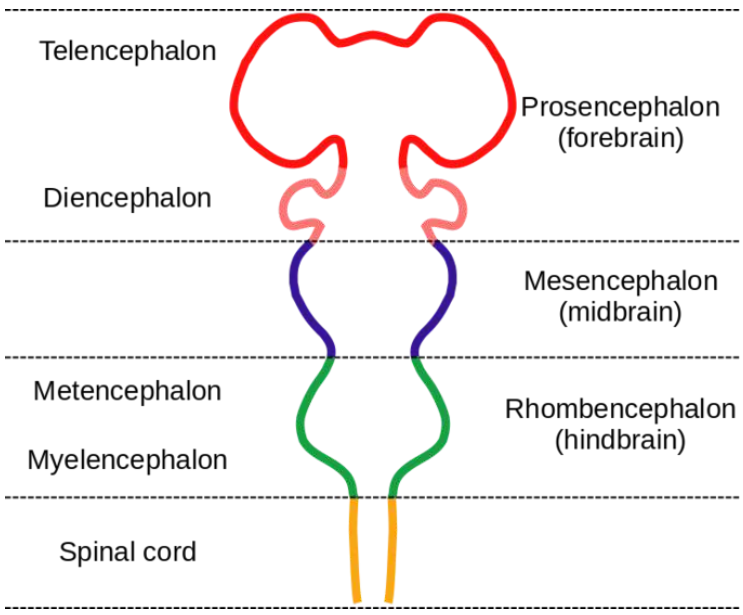


Figure **4.3.3**: Terms for the regions of the central nervous system during embryonic development. The anterior end of the neural tube forms vesicles that become the brain. The left side of the figure lists the five secondary vesicles, which differentiate from the three primary vesicles listed on the right side. The telencephalon and diencephalon differentiate from the prosencephalon (forebrain), The mesencephalon (midbrain) is fully differentiated at the primary vesicle stage (and thus does not form any additional secondary vesicle divisions). The metencephalon and myelencephalon differentiate from the rhombencephalon (hindbrain). The posterior end of the neural tube forms the spinal cord.

The diencephalon is the one region of the adult brain that retains its (Greek) embryonic name in common usage. This is because there is no better term for it (dia = “through”). The diencephalon is between the cerebrum and the rest of the nervous system and can be described as the region through which all projections have to pass. All vertebrate brains have these regions, but in humans the anterior forebrain is enlarged to the extent that the diencephalon (posterior forebrain), the midbrain, and even part of the hindbrain is hidden from view by the cerebrum.

Another aspect of the adult CNS structures that relates to embryonic development is the cerebral ventricles—open spaces within the CNS where cerebrospinal fluid circulates. They are the remnant of the hollow center of the neural tube. The four ventricles and the tubular spaces associated with them can be linked back to the hollow center of the embryonic brain.

4.3.4.1.2 Primary Vesicles

When the embryo is three to four weeks of age, the anterior end of the neural tube starts to develop into the brain. It undergoes a couple of enlargements; the result is the production of sac-like vesicles. Similar to a child’s balloon animal, the long, straight neural tube begins to take on a new shape. Three vesicles form at the first stage, which are called **primary vesicles**. These vesicles are given names that are based on Greek words, the main root word being *enkephalon*, which means “brain” (en- = “inside”; kephalon = “head”). The prefix to each generally corresponds to its position along the length of the developing nervous system.

The **prosencephalon** (pros- = “in front”) is the forward-most vesicle, and the term can be loosely translated to mean **forebrain**. The **mesencephalon** (mes- = “middle”) is the next vesicle, which can be called the **midbrain**. The third vesicle at this stage is the

rhombencephalon. The first part of this word is also the root of the word rhombus, which is a geometrical figure with four sides of equal length (a square is a rhombus with 90° angles). Whereas prosencephalon and mesencephalon translate into the English words forebrain and midbrain, there is not a word for "four-sided-figure-brain." Instead, the third vesicle is called the **hindbrain**. One way of thinking about how the brain is arranged is to use these three regions—forebrain, midbrain, and hindbrain—which are based on the primary vesicle stage of development (Figure **4.3.4**a).

4.3.4.1.3 Secondary Vesicles

The brain continues to develop in a five-week old embryo, and the vesicles differentiate further (Figure **4.3.4**b). The three primary vesicles become five **secondary vesicles**. The prosencephalon enlarges into two new vesicles called the **telencephalon** and the **diencephalon**. The telecephalon will become the cerebrum. The diencephalon gives rise to several adult structures; two that will be important are the thalamus and the hypothalamus.

The mesencephalon does not differentiate any further, but remains an established region of the brain. The rest of the brain develops around it and constitutes a large percentage of the mass of the brain. Dividing the brain into forebrain, midbrain, and hindbrain is useful in considering its developmental pattern, but the midbrain is a relatively small proportion of the entire brain.

The rhombencephalon develops into the **metencephalon** and **myelencephalon**. The metencephalon corresponds to the adult structure known as the pons and also gives rise to the cerebellum. The cerebellum (from the Latin meaning "little brain") accounts for about 10 percent of the mass of the brain and is an important structure in itself. The most significant connection between the cerebellum and the rest of the brain is at the pons, because the pons and cerebellum develop out of the same vesicle. The myelencephalon corresponds to the adult structure known as the medulla (or medulla oblongata).

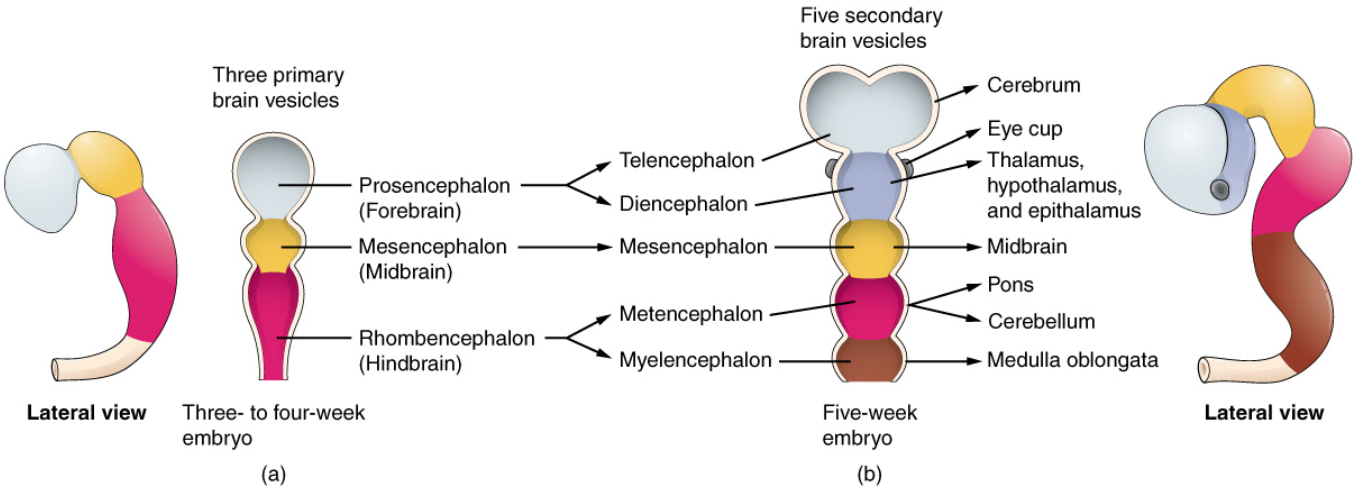


Figure **4.3.4**: Primary and Secondary Vesicle Stages of Development. The embryonic brain develops complexity through enlargements of the neural tube called vesicles. In part a, the primary vesicle stage which happens around 3 to 4 weeks of embryonic development shows three vesicles: the prosencephalon (forebrain) is superior and anterior; the mesencephalon (midbrain) is inferior to the prosencephalon; and the rhombencephalon (hindbrain) is the most inferior vesicle. In part b, the secondary vesicle stage which happens at 5 weeks of embryonic development shows five regions. From anterior and superior to posterior and inferior, these regions (and structures that form from them) are: telencephalon (cerebrum), diencephalon (eye cup; thalamus, hypothalamus and epithalamus), mesencephalon (midbrain), metencephalon (pons; cerebellum), and myelencephalon (medulla oblongata). Lateral views of each stage are also depicted.

4.3.4.2 Postnatal Brain Development

As summarized by Stiles & Jernigan (2010, p. 328), "Human brain development is a protracted process that begins in the third gestational week ... and extends at least through late adolescence, arguably throughout the lifespan". During the preschool years (roughly age three to five years), the brain quadruples in size, attaining about 90% of adult volume by age six (Stiles & Jernigan, 2010). Figure **4.3.5** illustrates the growth between a one-month old baby and a six-year-old child, in both brain size and neuron complexity.

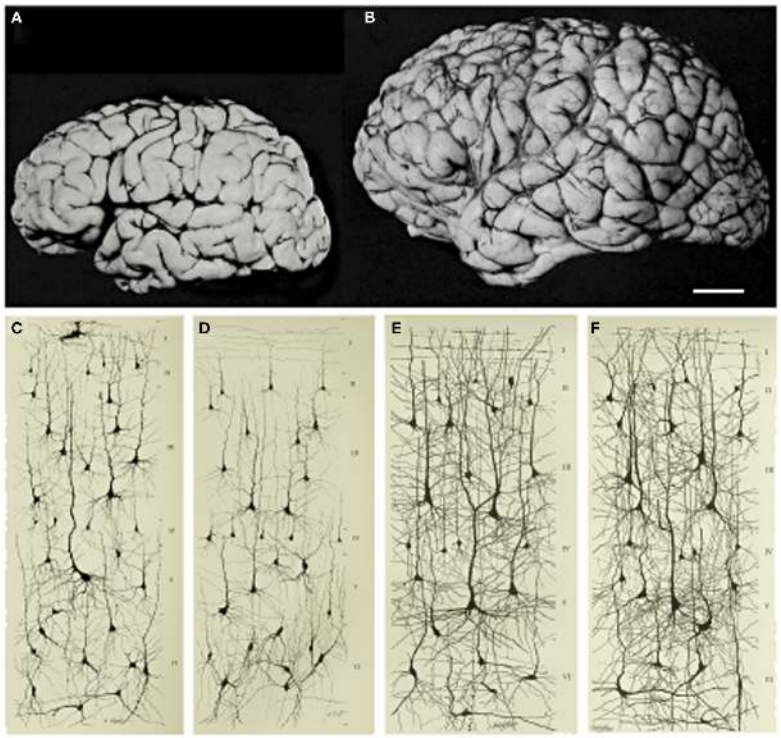


Figure **4.3.5**: Increase in brain size and the maturation of cortical circuits. The maturation of mental processes and motor skills is associated with an approximately four-fold enlargement in brain size. Photographs of the brains of a A. 1-month and B. 6-year-old child. This size increase is accompanied by a dramatic development in the complexity of the neuronal processes, which in turn is influenced by the genetic background and the surroundings. This increase in the complexity is clearly evident in the drawings of Golgi-stained cortical neurons from the cerebral cortex of a 1-month-old child (C. showing the cortical area known as “pars triangularis of gyrus frontalis inferior”; D. showing “orbital gyrus”) and a 6-year-old child (E. “pars triangularis of gyrus frontalis inferior”; F. “orbital gyrus”). Scale bar for A and B: 2 cm.

Changes in the structure of both gray matter and white matter components of the brain continue through childhood and adolescence, accompanied by changes in both functional organization and behavior. Young children, particularly infants, have much higher connectivity between neurons in the brain than adults. As development continues, these connections become more specialized (through pruning of less efficient pathways, as determined by experience- see neuron development below). Correspondingly, "plasticity and capacity for adaptation ... is the hallmark of early brain development" (Stiles & Jernigan, 2010, p. 328). Plasticity (the ability of the brain to change based on experience) continues throughout life, but slows down with age. It is still possible to "teach an old dog new tricks"- it just takes longer!

Although the human brain increases five-fold in volume between infancy and adulthood, there is little change in the number of neurons present. Brain growth occurs as a result of axon myelination and increased connections between neurons (Budday et al., 2015).

4.3.5 Spinal Cord Development

While the brain is developing from the anterior neural tube, the spinal cord is developing from the posterior neural tube. However, its structure does not differ from the basic layout of the neural tube. It is a long, straight cord with a small, hollow space down the center. The neural tube is defined in terms of its anterior versus posterior portions, but it also has a dorsal–ventral dimension. As the neural tube separates from the rest of the ectoderm, the side closest to the surface is dorsal (toward the back), and the deeper side is ventral (toward the belly).

As the spinal cord develops, the cells making up the wall of the neural tube proliferate and differentiate into the neurons and glia of the spinal cord. The dorsal tissues will be associated with sensory functions, and the ventral tissues will be associated with motor functions

Development of the Nervous System: Spina Bifida

Early formation of the nervous system depends on the formation of the neural tube. A groove forms along the dorsal surface of the embryo, which becomes deeper until its edges meet and close off to form the tube. If this fails to happen in the posterior region where the spinal cord forms, a developmental defect called spina bifida occurs. The closing of the neural tube is important for more than just the proper formation of the nervous system. The surrounding tissues are dependent on the correct development of the tube. The connective tissues surrounding the CNS can be involved as well.

There are three classes of this disorder: occulta, meningocele, and myelomeningocele (Figure **4.3.6**). The first type, spina bifida occulta, is the mildest because the vertebral bones do not fully surround the spinal cord, but the spinal cord itself is not affected. No functional differences may be noticed, which is what the word occulta means; it is hidden spina bifida. The other two types both involve the formation of a cyst—a fluid-filled sac of the connective tissues that cover the spinal cord called the meninges. “Meningocele” means that the meninges protrude through the spinal column but nerves may not be involved and few symptoms are present, though complications may arise later in life. “Myelomeningocele” means that the meninges protrude and spinal nerves are involved, and therefore severe neurological symptoms can be present.

Often surgery to close the opening or to remove the cyst is necessary. The earlier that surgery can be performed, the better the chances of controlling or limiting further damage or infection at the opening. For many children with meningocele, surgery will alleviate the pain, although they may experience some functional loss. Because the myelomeningocele form of spina bifida involves more extensive damage to the nervous tissue, neurological damage may persist, but symptoms can often be handled. Complications of the spinal cord may present later in life, but overall life expectancy is not reduced.

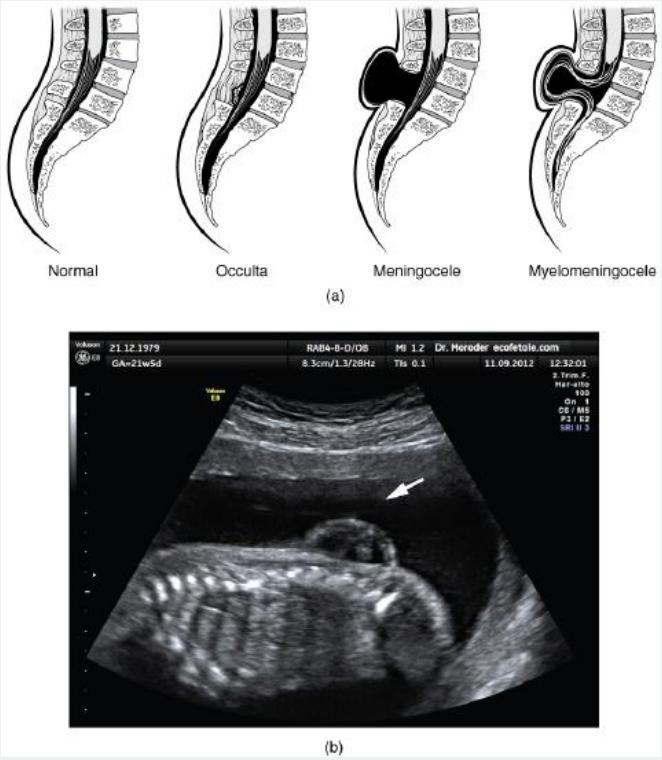


Figure **4.3.6**: Spina Bifida. Spina bifida is a birth defect of the spinal cord caused when the neural tube does not completely close, but the rest of development continues. The result is the emergence of meninges and neural tissue through the vertebral column. Part (a) shows the difference between a normal spinal cord, and the three types of spina bifida: occulta, meningocele and myelomeningocele. In part (b), an ultrasound taken at 21 weeks shows a fetal myelomeningocele.

4.3.6 Neuron Development

Neuron development is typically divided into several stages, although the names of the stages may vary somewhat from one source to another. **Neuron production** (also called proliferation) is the first stage, when **neurogenesis**, the production of new neurons from stem cells, occurs (Kalat, 2019). In humans, this begins around the fifth week of gestation (Budday et al., 2015), or on the 42nd day after conception (Stiles & Jernigan, 2010). It is mostly finished by 28 weeks gestational age, and if premature birth occurs before 28 weeks, further neurogenesis is inhibited (Kalat, 2019). Figure **4.3.7** shows a photograph of neurons developing in an embryo.

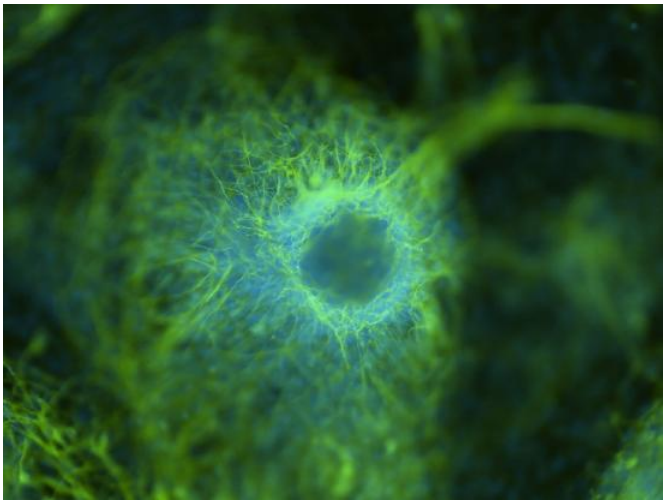


Figure **4.3.7**: Neuron cluster. Human embryonic tissue can be differentiated into many different cell types, including neurons. BIII-tubulin (green) is used as a marker for immature neurons. DAPI (blue) staining identifies the cells' nuclei.

As the neurons are forming, **migration** (movement to their final destinations) starts, following chemical signals (Kalat, 2019). The six-layered mature structure of the cerebral cortex is formed via the orderly migration of neurons (Stiles & Jernigan, 2010). **Differentiation** is the process of a neuron achieving the features that make it distinct from other body cells- forming an axon and one or more dendrites (Kalat, 2019). The term differentiation also encompasses the formation of different types of neurons (Stiles & Jernigan, 2010), or gaining the features that distinguish one type of neuron from another.

Synaptogenesis is the creation of connections between neurons by forming synapses (Kalat, 2019). It begins mid-gestation, and as connectivity progresses, axons form and reach out to numerous targets "until each neuron connects with thousands of other neurons" by birth (Budday et al., 2015, p. 5). Synaptogenesis "continues throughout life, as neurons form new synapses and discard old ones" (Kalat, 2019, p. 118).

Myelination is the process of axons becoming insulated with a layer of myelin, which speeds up the transmission of nerve impulses. Myelination starts in the spinal cord and then progresses from the hindbrain to the midbrain and finally the forebrain. This occurs gradually over decades and may also be implicated in learning new motor skills (Kalay, 2019).

Synaptic **pruning** is the process of removing synapses that are not useful or efficient, based on the specific experiences of the individual. Pruning starts around birth and is completed during adolescence, by the time sexual maturity is attained. It is thought that learning corresponds with pruning (Budday et al., 2015).

An estimated timeline for these stages of neuron development is shown in Figure **4.3.8**.

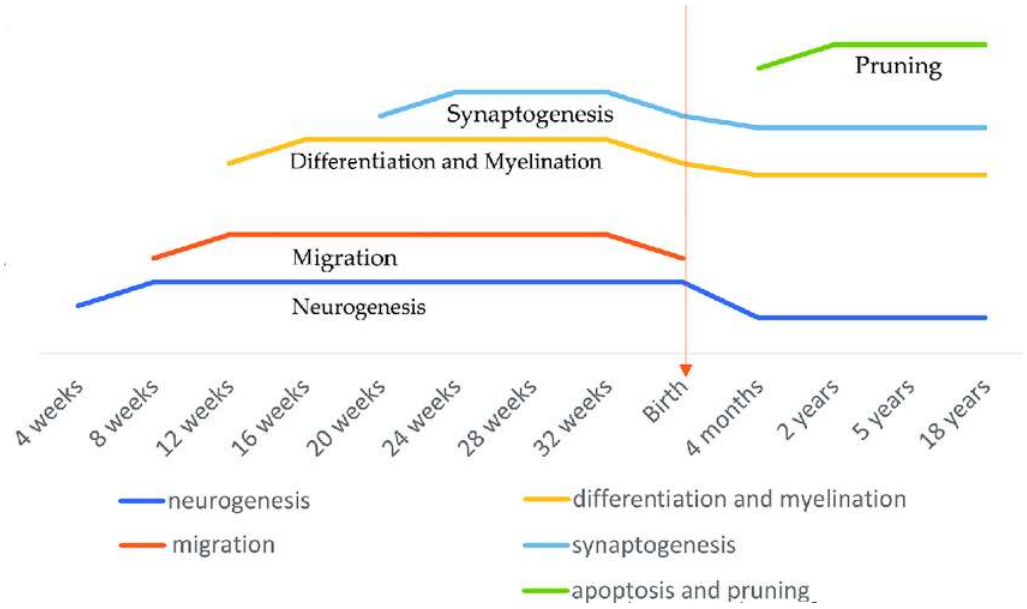


Figure **4.3.8**: Timeline for stages of neuron development. GA= gestational age. The timeline ends with age 18 years, but some of these processes may continue throughout life. Neurogenesis starts at four weeks GA and decreases sharply after birth, to almost nothing by four months of age and beyond. Migration starts at eight weeks GA, increases through 12 weeks GA, stays steady through 32 weeks GA, and then decreases, ending at birth. Differentiation and Myelination start at 12 weeks GA, increase through 16 weeks GA, stay steady through 32 weeks GA, decrease through four months of age, and then stay steady through 18 years of age. Synaptogenesis starts at 20 weeks GA, increases through 24 weeks GA, stays steady through 32 weeks GA, decreases through four months of age, and then stays steady through 18 years of age. Pruning starts at four months of age, increases through two years of age, and then stays steady through 18 years of age.

4.3.7 Summary

The development of the nervous system starts early in embryonic development. The outer layer of the embryo, the ectoderm, gives rise to the skin and the nervous system. A specialized region of this layer becomes a groove that folds in and becomes the neural tube beneath the dorsal surface of the embryo. The anterior end of the neural tube develops into the brain, and the posterior region becomes the spinal cord.

The brain develops from this early tube structure and gives rise to specific regions of the adult brain. As the neural tube grows and differentiates, it enlarges into three vesicles that correspond to the forebrain, midbrain, and hindbrain regions of the adult brain. Later in development, two of these three vesicles differentiate further, resulting in five vesicles. Those five vesicles can be aligned with the four major regions of the adult brain. The cerebrum is formed directly from the telencephalon. The diencephalon is the only region that keeps its embryonic name, and includes the thalamus and the hypothalamus. The mesencephalon becomes the midbrain, the metencephalon forms the pons and the cerebellum, and the myelencephalon becomes the medulla.

Brain development is a lifelong process, and the brain retains plasticity (the ability to change based on experience) throughout life. The increase in brain size during postnatal brain development occurs largely due to the myelination of axons and increased connections between neurons.

The spinal cord develops from the remainder of the neural tube and retains the tube structure, with the nervous tissue thickening and the hollow center becoming a very small central canal through the cord. The rest of the hollow center of the neural tube corresponds to open spaces within the brain called the ventricles, where cerebrospinal fluid is found.

Several stages of neuron development have been identified- neuron production (or proliferation), migration, differentiation, synaptogenesis (increased connectivity), myelination, and synaptic pruning.

4.3.8 References

Budday, S., Steinmann, P., & Kuhl, E. (2015). Physical biology of human brain development. *Frontiers in Cellular Neuroscience 9*, 1-17. DOI=10.3389/fncel.2015.00257

Kalat, J.W. (2019). *Biological Psychology* (13th ed.). Cengage.

Stiles, J. & Jernigan, T.L. (2010). The basics of brain development. *Neuropsychology Review 20*, 327–348. DOI 10.1007/s11065-010-9148-4

4.3.9 Attributions

1. Figures:
    1. Vertebrate embryo by Jlesk is licensed under CC BY-SA 3.0, via Wikimedia Commons
    2. Neural crest via Wikimedia Commons has been released into the public domain by its author, Abitua at English Wikipedia. This applies worldwide.
    3. Embryologic Brain Vesicles- no specific attribution (from " Development of the Human Brain" by LibreTexts, licensed under notset). Note: Text from this source is imprecise, sometimes to the point of being inaccurate, and is not included in the text adaptation on this page.
    4. Brain Vesicle Development by OpenStax is licensed under CC BY 4.0, via Wikimedia Commons
    5. Brain maturation by Javier DeFelipe is licensed CC BY 3.0, via Wikimedia Commons
    6. Spina Bifida by OpenStax is licensed under CC BY-SA 3.0 / Ultrasound image: "Spina bifida lombare sagittale" by Wolfgang Moroder is licensed under CC BY-SA 3.0; both are via Wikimedia Commons
    7. Neuron cluster by M. Oktar Guloglu, CC BY-SA 4.0, via Wikimedia Commons
    8. Timeline of neuron development adapted by Naomi Bahm (spelling corrected) from Timeline of brain development by Merve Çikili Uytun is licensed under CC BY 3.0, via Çikili Uytun, Merve. (2018). Development Period of Prefrontal Cortex. (From Prefrontal Cortex, edited by Ana Starcevic and Branislav Filipovic.) http://dx.doi.org/10.5772/intechopen.78697
2. Text adapted from:
    1. 11.5: Development of the Nervous System by Whitney Menefee, Julie Jenks, Chiara Mazzasette, & Kim-Leiloni Nguyen is licensed under CC BY.
3. Changes: Text (and some of the images) from above source pieced together with some modifications, transitions and additional content and images (particularly the embryonic brain development overview, and sections on postnatal brain development, and neuron development) added by Naomi I. Gribneau Bahm, PhD., Psychology Professor at Cosumnes River College, Sacramento, CA.

This page titled 4.3: Nervous System Development is shared under a mixed license and was authored, remixed, and/or curated by Naomi Bahm (ASCCC Open Educational Resources Initiative (OERI)) .

- **11.5: Development of the Nervous System** by Whitney Menefee, Julie Jenks, Chiara Mazzasette, & Kim-Leiloni Nguyen is licensed CC BY 4.0.

4.4: The Central Nervous System

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Name the structures that comprise the central nervous system
2. Identify the three main parts of the brain
3. Describe the lobes of the cerebrum and their locations
4. Identify the components of the diencephalon and describe their functions
5. Identify the regions of the brainstem and describe their functions
6. Identify the major structures of the limbic system and describe their functions
7. Describe the structure and function of the cerebellum
8. Describe the structure and function of the spinal cord

4.4.1 Overview

This module defines the central nervous system (the brain and spinal cord) and goes into detail on its components. Structures and functions of the brain are discussed, including the cerebrum: hemispheres and lateralization, the cerebral cortex, the frontal, parietal, temporal and occipital lobes; the brainstem: diencephalon, midbrain, pons and medulla; the limbic system; and the cerebellum; as well as structures, function and injury of the spinal cord.

Homunculus



Figure **4.4.1**: Front and rear views of a sensory homunculus model.

Figure **4.4.1** is a very odd-looking model called a homunculus. This model shows the relative representation of different parts of the body in the primary sensory cortex, an area in the parietal lobe of the brain. As you can see, larger areas of the brain in this region are associated with the hands, face, and tongue than the arms, torso, legs and feet. Larger areas of representation in the brain result in greater sensitivity in those body areas. Given the importance of speech, manual dexterity, and face-to-face social interactions in human beings, it is not surprising that relatively large areas of the brain are needed to relay sensation from these body parts. (There is a similarly odd-looking representation of voluntary motor functions called the motor homunculus that is located in the primary motor cortex, an area in the frontal lobe of the brain.) The brain is the most complex organ in the human body and is part of the central nervous system.

4.4.2 What Is the Central Nervous System?

The central nervous system (CNS) is the part of the nervous system that includes the brain and spinal cord. Figure **4.4.2** shows the central nervous system both as it appears in the body dissected from the backside and also completely removed from the body. The

central nervous sytem and the peripheral nervous system (PNS, the other main division, which is addressed in a separate module), work together to control virtually all body functions.

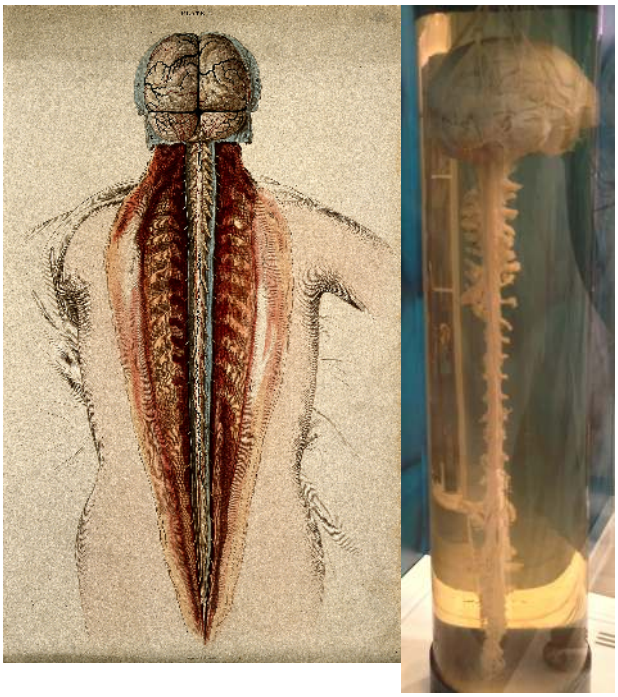


Figure **4.4.2**: The human central nervous system. A back view dissection of the brain and spinal cord (colored line engraving by W.H. Lizars, circa 1827) and a preserved brain and spinal cord.

4.4.3 The Brain

The **brain** is the control center not only of the rest of the nervous system but of the entire organism. The adult brain makes up only about two percent of the body's weight, but it uses about 20 percent of the body's total energy. The brain contains an estimated 86 billion neurons (Lent et al., 2012), and each neuron has thousands of synaptic connections to other neurons. It is estimated that the brain also has about the same number of glial cells as neurons. No wonder the brain uses so much energy! In addition, the brain uses mostly glucose for energy. As a result, if the brain is deprived of glucose, it can lead to unconsciousness.

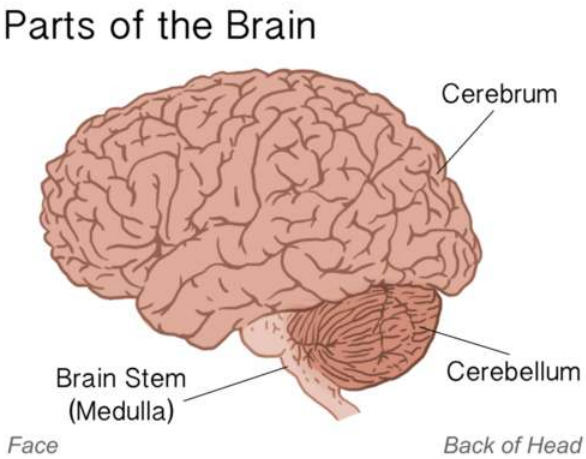


Figure **4.4.3**: The major parts of the brain are the cerebrum, cerebellum, and brainstem. The medulla is part of the brainstem.

The brain controls such mental processes as reasoning, imagination, memory, and language. It also interprets information from the senses and commands the body's responses. It controls basic physiological processes such as breathing and heartbeat as well as voluntary activities such as walking and writing. The brain has three major parts: the cerebrum, cerebellum, and brainstem (Figure **4.4.3**). The figure shows the brain from the left side of the head. It shows how the brain would appear if the skull and meninges

were removed. The brainstem links to the spinal cord via the medulla. The cerebellum is a small structure at the back of the brain. The largest part of the brain is the cerebrum.

4.4.4 Cerebrum

The iconic gray mantle of the human brain, which appears to make up most of the mass of the brain, is the **cerebrum**. It controls conscious, intellectual functions, such as reasoning, language, memory, sight, touch, and hearing. When you read a book, play a video game, or recognize a classmate, you are using your cerebrum. The wrinkled portion is the **cerebral cortex** (see below), and the rest of the structure is deep to that outer covering.

4.4.4.1 Hemispheres and Lateralization of the Cerebrum

There is a large separation between the two sides of the cerebrum called the **longitudinal fissure**. It separates the cerebrum into two distinct halves, the right and left **cerebral hemispheres**. The two hemispheres are connected by a thick bundle of axons, known as the **corpus callosum**, which lies deep within the cerebrum. The corpus callosum is the major pathway for communication between the two hemispheres, connecting each point in the cerebral cortex to the mirror-image point in the opposite hemisphere. The location of the corpus callosum is shown with a dotted outline in Figure 4.4.4, as though you are looking through the outside and deep inside the brain. Figure 4.4.5 shows a three-dimensional image of the corpus callosum in relation to other deep brain structures and surrounded by the cerebrum.

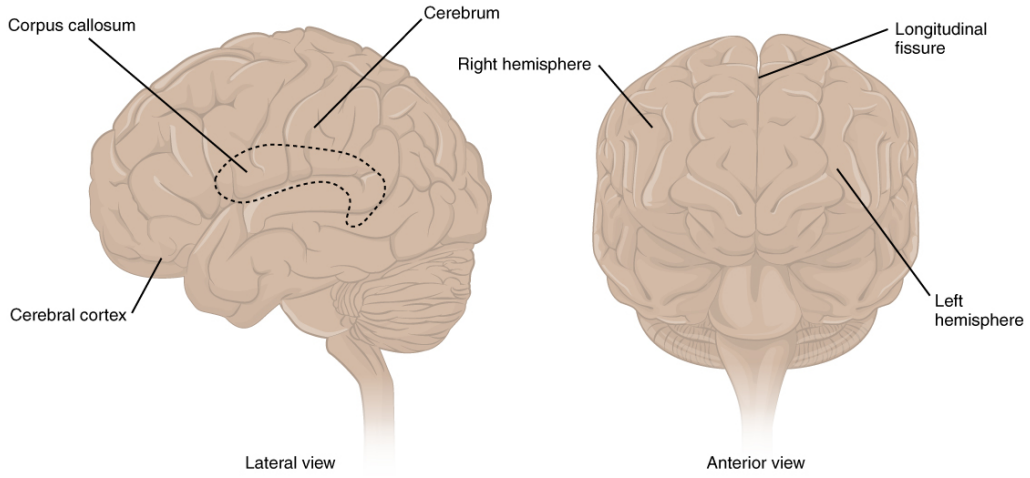


Figure 4.4.4: The cerebrum is a large component of the CNS in humans, and the most obvious aspect of it is the folded surface called the cerebral cortex.

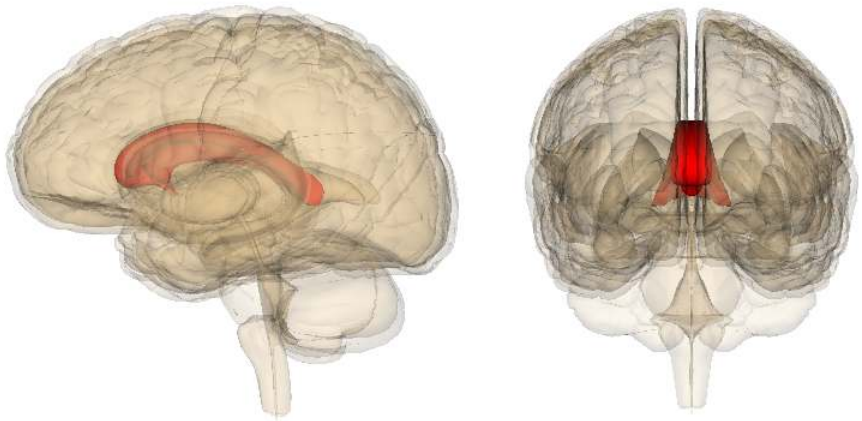


Figure 4.4.5: The corpus callosum (shown in red) is a thick bundle of axons connecting the left and right hemispheres of the brain. It is located deep within the cerebrum.

The right and left hemispheres of the cerebrum are similar in shape, and most areas of the cerebrum are found in both hemispheres. Some areas, however, show **lateralization,** or a concentration in one hemisphere or the other. For example, in most people, language functions are more concentrated in the left hemisphere, whereas abstract reasoning and visual-spatial abilities are more concentrated in the right hemisphere.

For reasons that are not yet clear, each hemisphere of the brain interacts primarily with the **contralateral** (opposite) side of the body. The left side of the brain receives messages from and sends commands to the right side of the body, and the right side of the brain receives messages from and sends commands to the left side of the body. Sensory nerves from the spinal cord to the brain and motor nerves from the brain to the spinal cord both cross to the other side of the body.

4.4.4.2 Cerebral Cortex

The cerebrum is covered by a continuous layer of gray matter that wraps around either side of the forebrain—the **cerebral cortex**. This thin (just a few millimeters thick), extensive region of wrinkled gray matter is responsible for most of the information processing and higher functions of the nervous system. The cerebral cortex has many folds in it that greatly increase the surface area of the brain. A **gyrus** (plural = gyri) is the ridge of one of those wrinkles, and a **sulcus** (plural = sulci) is the groove between two gyri. The pattern of these folds of tissue indicates specific regions of the cerebral cortex. At birth, the head of the newborn is limited by the size of the birth canal, and the brain must fit inside the cranial cavity of the skull. Extensive folding in the cerebral cortex enables more gray matter to fit into this limited space. If the gray matter of the cerebral cortex in an adult were peeled off of the cerebrum and laid out flat, its surface area would be about 2,500 cm^2 (2.5 ft^2), or about the size of the top of a card table.

4.4.4.3 Lobes of the Cerebral Cortex

The surface of the brain can be mapped on the basis of the locations of large gyri and sulci. Using these landmarks, the cortex can be separated into four major regions, or lobes (Figure **4.4.6**). The **lateral sulcus** that separates the **temporal lobe** from the other regions is one such landmark.

Superior to the lateral sulcus are the **frontal lobe** and **parietal lobe**, which are separated from each other by the **central sulcus**. The most posterior gyrus of the frontal lobe is the **precentral gyrus**, which is where the primary motor cortex is located, while the most anterior gyrus of the parietal lobe is the **postcentral gyrus**, which is where the primary somatosensory cortex is located. (The primary somatosensory cortex is the area that the homonculus in Figure **4.4.1** represents.)

The posterior region of the cortex is the **occipital lobe**, which has no obvious anatomical border between it and the parietal or temporal lobes on the lateral surface of the brain. From the medial surface, the **parieto-occipital sulcus** separates the parietal and occipital lobes.

Figure **4.4.7** has only the four lobes labeled for quicker reference. The lobes are associated with multiple functions. The image shows one function of each lobe: frontal is associated with reasoning, parietal with touch, occipital with sight, and temporal with hearing. The lobes (found on each hemisphere of the cerebrum) are described in greater detail below.

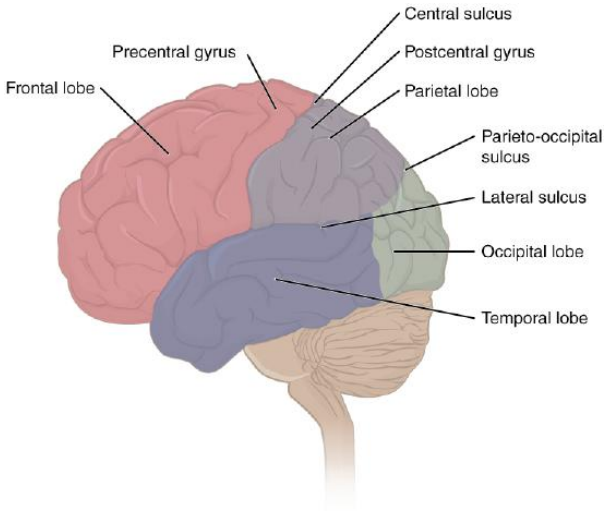


Figure 4.4.6: Major gyri, sulci and lobes of the cerebral cortex.

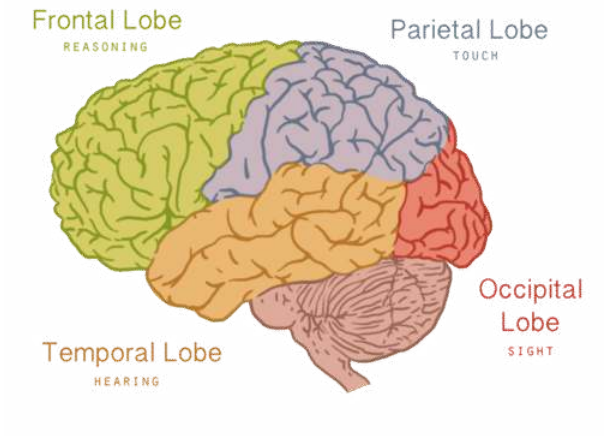


Figure 4.4.7: The four major lobes of the brain.

4.4.4.3.1 Frontal Lobe

The **frontal lobes** are located at the front of the brain behind the forehead. The frontal lobes are associated with executive functions such as attention, self-control, planning, problem-solving, reasoning, abstract thought, language, and personality. Another important function of the frontal lobes is movement- as mentioned above, the precentral gyrus (indicated on Figure 4.4.6) contains the primary motor cortex. Like the sensory homunculus pictured in the beginning of this section (Figure 4.4.1), there is a similarly disproportionate representation of voluntary motor control (the motor homunculus) located in the precentral gyrus. Figure 4.4.8 shows the areas of the primary motor cortex dedicated to movement of various body areas. Notice that like the sensory homunculus, the hands, fingers, thumb, and facial areas have a greater share of the cortex than areas with less fine-tuned movement (like the shoulders, trunk, and hips, for example).

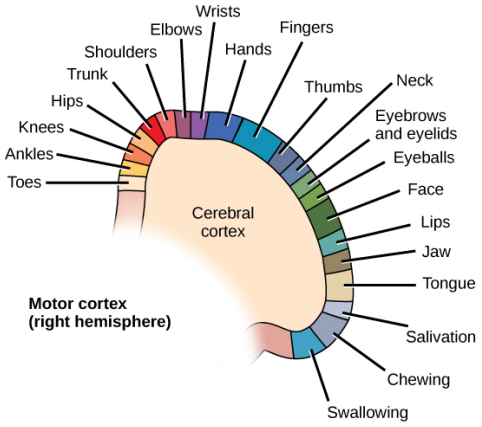


Figure 4.4.8: Different regions of the motor cortex control different muscle groups and their neighboring regions control neighboring muscle groups. Although the right hemisphere is shown, a mirror-image exists in the left hemisphere as well. Note that all areas of the body that have voluntary motor control are represented on the primary motor cortex.

4.4.4.3.2 Parietal Lobe

The **parietal lobes** are located posterior to the frontal lobes at the top of the head. The parietal lobes are involved in body sensations, including temperature, touch, and pain. As mentioned above, the postcentral gyrus (indicated on Figure 4.4.6) contains the primary somatosensory cortex. The sensory homunculus model pictured in the beginning of this section (Figure 4.4.1) shows the strange character that results from making the body areas proportional to their representation in the brain. The drawing in Figure 4.4.9 shows the sensory homunculus stretched out along the postcentral gyrus of the parietal lobe. Notice that areas that have sensation but no voluntary motor control (like the genitals and the intra-abdominal organs) are present on the primary somatosensory cortex but not on the primary motor cortex (Figure 4.4.8). Reading and arithmetic are also functions of the parietal lobes.

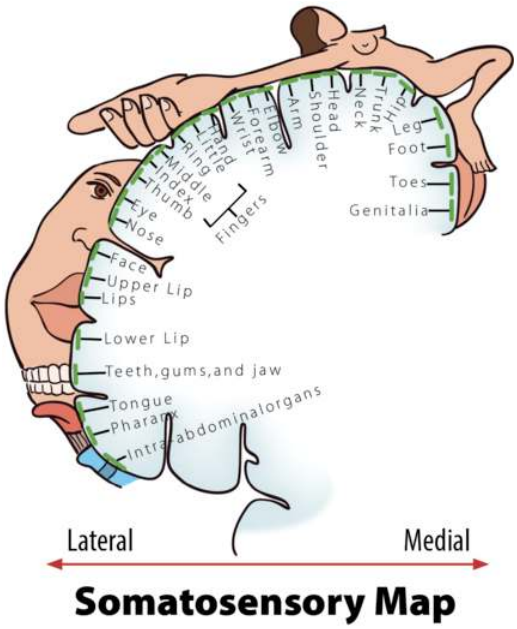


Figure 4.4.9: Somatosensory Map. The sensory map of your body represented on the primary somatosensory cortex in your brain is not proportional to the actual size of your body parts. Sensitive areas like your face (particularly your lips) and hands (particularly your fingers) occupy a larger section of cortex than less sensitive areas like your limbs (arms and legs) or trunk. Although the left hemisphere is shown, a mirror-image exists in the right hemisphere as well. Note that all areas of the body that register touch sensations are represented on the primary somatosensory cortex.

4.4.4.3.3 Temporal Lobe

The **temporal lobes** are located at the sides of the head below the frontal and parietal lobes. The temporal lobes enable the formation and retrieval of memories, the integration of memories and sensations, and perception of the senses of hearing (audition),

and smell (olfaction)- primary auditory cortex and primary olfactory cortex are both located in the temporal lobes.

4.4.4.3.4 Occipital Lobe

The **occipital lobes** are located at the back of the head below the parietal lobes. The occipital lobes are the smallest of the four pairs of lobes. Primary visual cortex is located in the occipital lobes, and they are dedicated almost solely to vision.

4.4.5 Brainstem

As noted separately (see the nervous system structure module), sometimes the diencephalon is considered as a separate additional region, and sometimes it is included as the uppermost part of the brainstem (Basinger and Hogg, 2021), as it is here. The **brainstem** is the central region of the brain that connects the cerebrum (on the superior end) to the diencephalon, midbrain, pons, medulla, and ultimately joins the spinal cord (on the inferior end). The **cerebellum** is attached to the brainstem, but is considered a separate region of the adult brain.

One of the brainstem's most important roles is that of an "information highway." That is, all of the information coming from the body to the brain and the information from the cerebrum to the body go through the brainstem. Sensory pathways for such things as pain, temperature, touch, and pressure sensation go upward to the cerebrum, and motor pathways for movement and other body processes go downward to the spinal cord. Most of the axons in the motor pathways cross from one side of the CNS to the other as they pass through the medulla. As a result, the right side of the brain controls much of the movement on the left side of the body, and the left side of the brain controls much of the movement on the right side of the body. Similarly, axons in sensory pathways cross from one side of the CNS to the other either in the spinal cord or in the medulla (depending on the specific sensory information carried, such as pain versus localized touch), such that the right side of the brain receives sensation primarily from the left side of the body and vice versa.

4.4.5.1 Diencephalon

The diencephalon ("through brain") is the one region of the adult brain that retains its name from embryonic development. It connects the cerebrum and the rest of the nervous system, with one exception. The brainstem, the spinal cord, and the PNS all send information to the cerebrum through the diencephalon, and output from the cerebrum then passes back through the diencephalon. The single exception to this pattern is the sense of smell (olfaction), which connects directly to the cerebrum first. The diencephalon comprises the ventral surface of the forebrain (and connects to the tapering cone of the midbrain, pons, and medulla, the remaining structures in the brainstem).

The diencephalon is located deep within the cerebrum and constitutes the walls of the third ventricle. The diencephalon can be described as any region of the brain with "thalamus" in its name. The two major regions of the diencephalon are the thalamus itself and the hypothalamus (Figure **4.4.10**). There are other structures as well, such as the epithalamus, which contains the pineal gland.

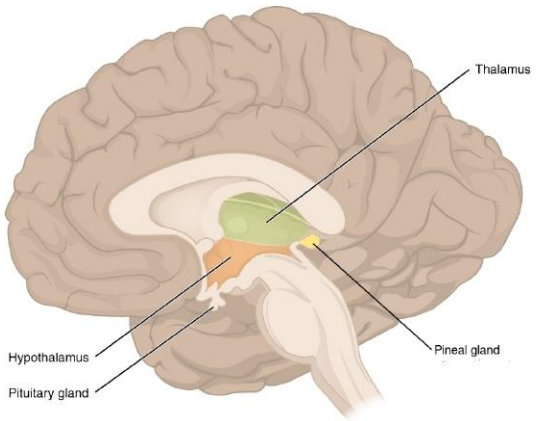


Figure **4.4.10**: The diencephalon.

4.4.5.1.1 Thalamus

The **thalamus,** which is located above the hypothalamus (Figure **4.4.10** and Figure **4.4.11**), is a major hub for information traveling back and forth between the spinal cord, brainstem, and cerebrum. It filters sensory information traveling to the cerebrum. It relays

sensory signals to the cerebral cortex and motor signals to the spinal cord. It is also involved in the regulation of consciousness, sleep, and alertness. As noted above, olfaction (smell) is the only sense that travels directly to the cortex and does not travel through the thalamus (as part of the diencephalon) first. However, olfactory messages do travel from the cortex back to the thalamus, so smell is still a participant in the "sensory relay" functions of the thalamus.

4.4.5.1.2 Hypothalamus

The **hypothalamus** ("below thalamus") is about the size of an almond and is responsible for certain metabolic processes and other activities of the autonomic nervous system, including body temperature, heart rate, hunger, thirst, fatigue, sleep, wakefulness, and circadian (24-hour) rhythms. The hypothalamus is also an important emotional center of the brain. The hypothalamus can regulate so many body functions because it responds to many different internal and external signals, including messages from the brain, light, steroid hormones, stress, and invading pathogens, among others.

One way the hypothalamus influences body functions is by synthesizing hormones that directly influence body processes. For example, it synthesizes the hormone oxytocin, which stimulates penile and vaginal contractions during orgasm, uterine contractions during childbirth, and the letdown of milk during lactation. It also synthesizes the hormone vasopressin (also called antidiuretic hormone), which stimulates the kidneys to reabsorb more water and excrete more concentrated urine. These two hormones are sent from the hypothalamus via a stalk-like structure called the **infundibulum** ("pituitary stalk", see Figure **4.4.11**) directly to the posterior (back) portion of the **pituitary gland**, which secretes them into the blood.

The main way the hypothalamus influences body functions is by controlling the pituitary gland, known as the "master gland" of the **endocrine system**. (The endocrine system is considered separately in the communication and the endocrine system module.) The hypothalamus synthesizes neurohormones called releasing factors that travel through the infundibulum directly to the anterior (front) part of the pituitary gland. The releasing factors generally either stimulate or inhibit the secretion of anterior pituitary hormones, most of which control other glands of the endocrine system.

Figure **4.4.11** shows where the thalamus, hypothalamus, and pituitary gland (typically considered an extension of the hypothalamus) are located in the brain. The cerebrum, hypothalamus, and thalamus exist in two halves, one in each hemisphere. The pituitary gland is a single structure in the center, connected to the hypothalamus by the infundibulum (or "pituitary stalk").

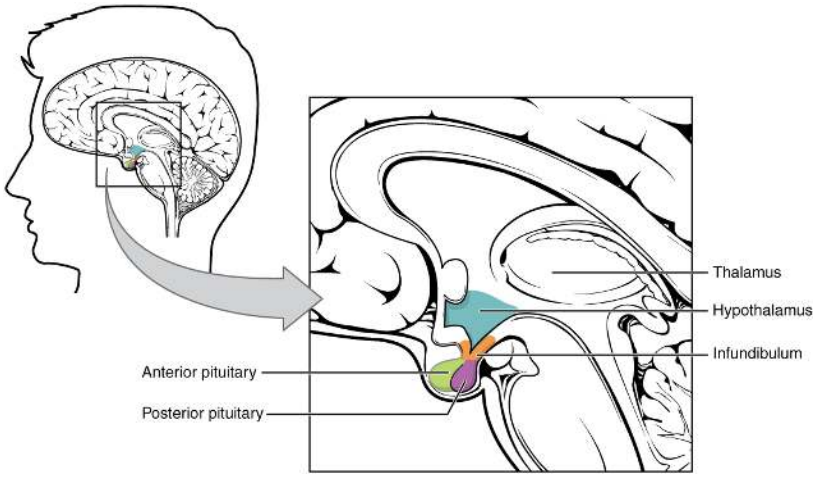


Figure **4.4.11**: Just inferior (and deep) to the cerebrum is the thalamus. The hypothalamus is located a little bit anterior and inferior to the thalamus. The pituitary gland is attached to the hypothalamus via a stalk called the infundibulum.

The remaining parts to the brainstem are the midbrain, the pons, and the medulla, which are shown in Figure **4.4.12** below. These structures are primarily involved in the unconscious functions of the autonomic nervous system as well as several types of sensory information. They also help coordinate large body movements such as walking and running.

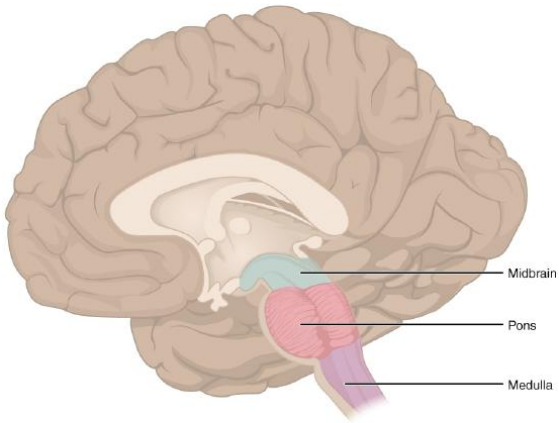


Figure 4.4.12: The brainstem includes the midbrain, pons, and medulla

4.4.5.2 Midbrain

The **midbrain** coordinates sensory representations of sight (visual system- the **superior colliculi**), sound (auditory system- the **inferior colliculi**), and somatosensory perceptual spaces, translating these inputs before sending them to the forebrain.

4.4.5.3 Pons

The **pons** ("bridge") is the main connection with the cerebellum. It relays messages between other parts of the brain (primarily the cerebrum and cerebellum) and sends messages to paralyze major body muscles during REM sleep. Some researchers have hypothesized that the pons plays a role in dreaming. Some of the functions of the pons are shared by the medulla, such as regulating several crucial functions including cardiovascular system and respiratory systems, and heartbeat and breathing rates.

4.4.5.4 Medulla

The **medulla** (also called the medulla oblongata) controls several subconscious homeostatic functions such as breathing, heart and blood vessel activity, swallowing, and digestion.

4.4.6 Limbic System

The **limbic system** is a collection of structures of the cerebrum and diencephalon that are involved in emotion, motivation and memory. Although still under debate, the structures most often recognized as belonging to this system are the cingulate gyrus, hippocampus, amygdala, olfactory structures, and various nuclei of the diencephalon. Here we will focus on the cingulate gyrus, the hippocampus, and the amygdala (Figure 4.4.13). The **cingulate gyrus** (colored green in the figure) is located on the midsagittal surface of the cortex, immediately superior to the corpus callosum and surrounding the diencephalon. The cingulate gyrus focuses attention on events that are emotionally salient. The **hippocampus** (colored purple in the figure) is a nucleus shaped like a seahorse, hence its name. The hippocampus is essential for forming memories and storing them long-term and is located deep in the temporal lobe. Lastly, the **amygdala** (also called the amygdaloid body, colored blue in the figure) is connected to the anterior end of the hippocampus and is involved in multiple aspects of emotion, including encoding memories related to highly emotional states. The amygdala is especially important for the emotion fear.

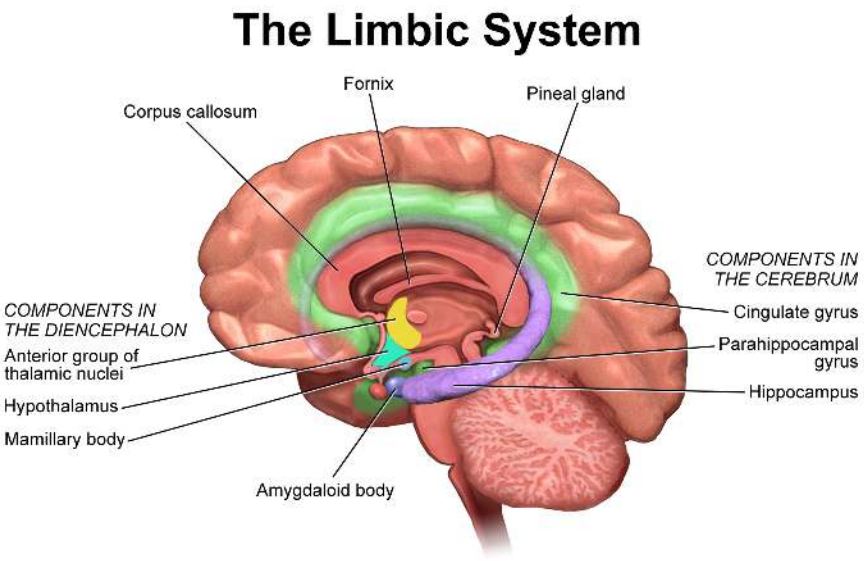
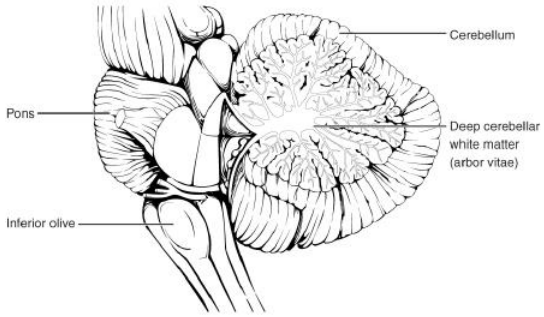


Figure 4.4.13: The Limbic System includes components from both the diencephalon (anterior group of thalamic nuclei, hypothalamus, and mammilary body) and the cerebrum (cingulate gyrus, parahippocampal gyrus, hippocampus, and amygdaloid body). Structures additionally labeled for reference are the corpus callosum, fornix, and pineal gland.

4.4.7 Cerebellum

The **cerebellum**, as the name suggests, is the "little brain", just below the cerebrum and at the back of the brain behind the brainstem. It is covered in gyri and sulci like the cerebrum, and looks like a miniature version of that part of the brain Figure 4.4.14. The cerebellum is largely responsible for comparing information from the cerebrum with sensory feedback from the periphery through the spinal cord. It accounts for approximately 10 percent of the mass of the brain. The cerebellum coordinates body movements and is involved in movements that are learned with repeated practice. For example, when you hit a softball with a bat or touch type on a keyboard you are using your cerebellum. Many nerve pathways link the cerebellum with motor neurons throughout the body.



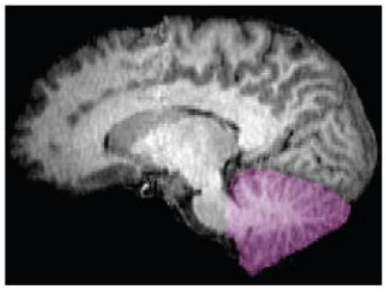


Figure 4.4.14: **The cerebellum.** The cerebellum is situated on the posterior surface of the brainstem. Descending input from the cerebellum enters through the large white matter structure of the pons. Ascending input from the periphery and spinal cord enters through the fibers of the inferior olive. Output goes to the midbrain, which sends a descending signal to the spinal cord.

4.4.8 Spinal Cord

The above description of the CNS is concentrated on the structures of the brain, but the **spinal cord** is another major organ of the system. The spinal cord is a long, thin, tubular bundle of nervous tissues that extends from the brainstem down through the center of the back and pelvis. It is highlighted in yellow in Figure 4.4.15.



Figure 4.4.15: The spinal cord (shown in yellow) runs from the inferior aspect of the brain to the lower back. Spinal nerves are shown in black, and bones are shown in blue.

4.4.8.1 Structure of the Spinal Cord

The center of the spinal cord consists of **gray matter**, which is made up mainly of neuron cell bodies, including interneurons and motor neurons. The gray matter is surrounded by **white matter** that consists mainly of myelinated axons from motor and sensory neurons. Sensory neuron axons enter the posterior side through the **dorsal (/posterior) nerve root**. Motor neuron axons emerge from the anterior side through the **ventral (/anterior) nerve root**. Note that it is common to see the terms dorsal (dorsal = back) and ventral (ventral = belly) used interchangeably with posterior and anterior, particularly in reference to nerves and the structures of the spinal cord. The **central canal** travels through the center of the spinal cord. The central canal is filled with **cerebrospinal fluid** (CSF). Figure 4.4.16 shows the gray matter, white matter, central canal, dorsal roots and ventral roots of the spinal cord.

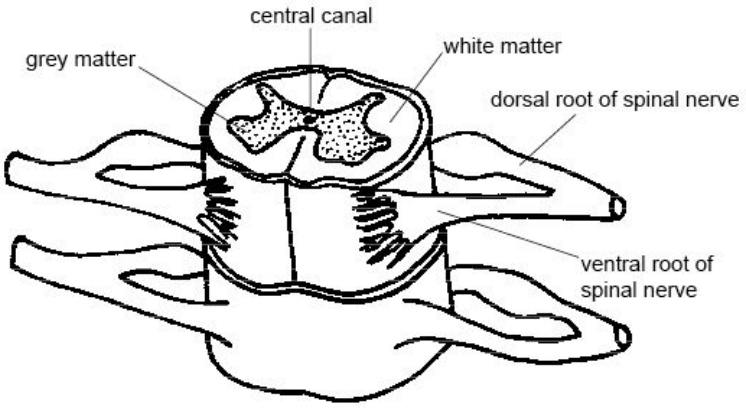


Figure 4.4.16: Cross section of spinal cord showing the central canal, gray and white matter, and the dorsal and ventral roots of the spinal nerves.

At each level of the spinal cord, the dorsal and ventral roots come together to form spinal nerves. Because the spinal cord is shorter than the vertebral column it is enclosed within, nerve roots must travel lower to exit at the associated vertebra. Spinal nerves connect the spinal cord to the PNS and enter/exit from the spinal cord between vertebrae (Figure 4.4.17).

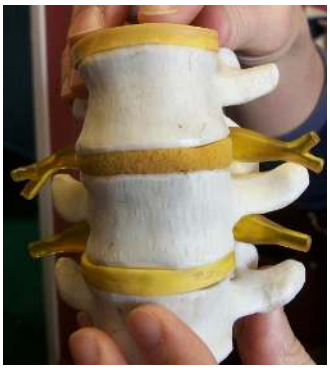


Figure **4.4.17**: This model shows three vertebrae (white) with branching spinal nerves (yellow) emerging from either side of the spinal cord between vertebrae. Also shown in yellow are pancake-like disks between vertebrae that help cushion and maintain space between adjacent bony vertebrae.

4.4.8.2 **Gray Horns**

In cross-section, the gray matter of the spinal cord has the appearance of an ink-blot test, with the spread of the gray matter on one side replicated on the other—a shape reminiscent of a bulbous capital "H" or a butterfly. As shown in Figure **4.4.18**, the gray matter is subdivided into regions that are referred to as horns.

The **posterior horn** is responsible for sensory processing. The **anterior horn** sends out motor signals to the skeletal muscles. The **lateral horn**, which is only found in the thoracic, upper lumbar, and sacral regions, is the central component of the sympathetic division of the autonomic nervous system.

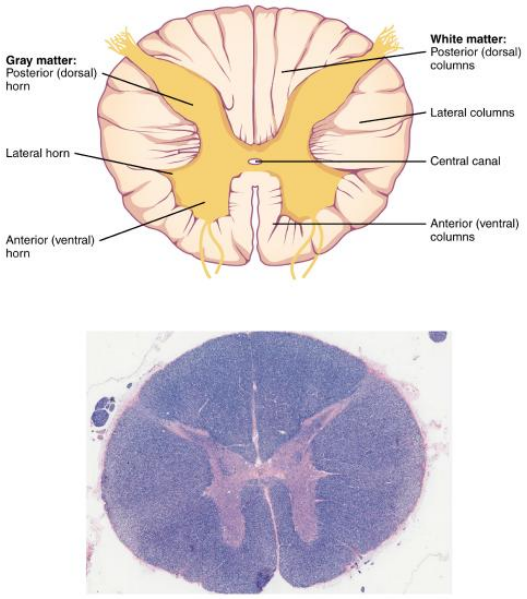


Figure **4.4.18**: **Cross-section of Spinal Cord** The cross-section of a thoracic spinal cord segment shows the posterior (dorsal), anterior (ventral), and lateral horns of gray matter, as well as the posterior (dorsal), anterior (ventral), and lateral columns of white matter. Light Microscopy × 40 magnification.

4.4.8.3 Functions of the Spinal Cord

The spinal cord serves as an information superhighway. It passes messages from the body to the brain and from the brain to the body. Sensory (afferent) nerves carry nerve impulses to the brain from sensory receptor cells everywhere in and on the body. Motor (efferent) nerves carry nerve impulses away from the brain to glands, organs, or muscles throughout the body.

The spinal cord also independently controls certain rapid responses called **reflexes** without any input from the brain. A sensory receptor responds to a sensation and sends a nerve impulse along a sensory nerve to the spinal cord. In the spinal cord, the message passes to an interneuron and from the interneuron to a motor nerve, which carries the impulse to a muscle. The muscle contracts in response. These neuron connections form a reflex arc, which requires no input from the brain. No doubt you have experienced such

reflex actions yourself. For example, you may have reached out to touch a pot on the stove, not realizing that it was very hot. Virtually at the same moment that you feel the burning heat, you jerk your arm back and remove your hand from the pot.

4.4.8.4 Injuries to the Spinal Cord

Physical damage to the spinal cord may result in **paralysis,** which is a loss of sensation and movement in part of the body. Paralysis generally affects all the areas of the body below the level of the injury because nerve impulses are interrupted and can no longer travel back and forth between the brain and body beyond that point. If an injury to the spinal cord produces nothing more than swelling, the symptoms may be transient. However, if nerve fibers (axons) in the spinal cord are badly damaged, the loss of function may be permanent. Experimental studies have shown that spinal nerve fibers attempt to regrow, but tissue destruction usually produces scar tissue that cannot be penetrated by the regrowing nerves, as well as other factors that inhibit nerve fiber regrowth in the central nervous system.

Feature: My Human Body

Each year, many millions of people have a stroke, and stroke is the second leading cause of death in adults. **Stroke,** also known as **cerebrovascular accident**, occurs when poor blood flow to the brain results in the death of brain cells. There are two main types of strokes:

- **Ischemic** strokes occur due to a lack of blood flow because of a blood clot in an artery going to the brain.
- **Hemorrhagic** strokes occur due to bleeding from a broken blood vessel in the brain.

Either type of stroke may result in paralysis, loss of the ability to speak or comprehend speech, loss of bladder control, personality changes, and many other potential effects, depending on the part of the brain that is injured. The effects of a stroke may be mild and transient or more severe and permanent. A stroke may even be fatal. It generally depends on the type of stroke and how extensive it is.

Are you at risk of stroke? The main risk factor for stroke is age: about two-thirds of strokes occur in people over the age of 65. There is nothing you can do about your age, but most other stroke risk factors can be reduced with lifestyle changes or medications. The risk factors include high blood pressure, tobacco smoking, obesity, high blood cholesterol, diabetes mellitus, and atrial fibrillation.

Chances are good that you or someone you know is at risk of a stroke, so it is important to recognize a stroke if one occurs. Stoke is a medical emergency, and the more quickly treatment is given, the better the outcome is likely to be. In the case of ischemic strokes, the use of clot-busting drugs may prevent permanent brain damage if administered within 3 or 4 hours of the stroke. Remembering the signs of a stroke is easy. They are summed up by the acronym FAST, as explained in Figure **4.4.19**.



Figure **4.4.19**: The signs of stroke are abbreviated as FAST, where F stands for droopy face, A stands for arm weakness, S stands for slurry speech, and T stands for time to call 911.

4.4.9 Summary

The central nervous system (CNS) includes the brain and spinal cord, described in greater detail in this module. The adult brain is separated into three major regions: the cerebrum, the brainstem, and the cerebellum. The cerebrum is the largest portion and is divided into two halves called hemispheres by the longitudinal fissure. The two hemispheres are connected by the corpus callosum, the major pathway for communication between the hemispheres.

The cerebral cortex, a continuous layer of gray matter covering the cerebrum, is the location of important cognitive functions. The cerebral cortex is separated into the frontal, parietal, temporal, and occipital lobes. The frontal lobe is responsible for executive functions and movements. The parietal lobe is involved in body sensations, reading, and arithmetic. The temporal lobe has regions crucial for memory formation, and contains cortical areas that process audition (hearing) and olfaction (smell). The occipital lobe is where visual processing begins.

The brainstem is composed of the diencephalon, midbrain, pons, and medulla. The brainstem serves as an "information highway", since all of the information coming from the body to the brain and the information from the cerebrum to the body go through the brainstem. The two major regions of the diencephalon are the thalamus and the hypothalamus. The thalamus is a relay between the cerebrum and the rest of the nervous system. The hypothalamus coordinates metabolic processes through the autonomic and endocrine systems. The midbrain coordinates sensory representations of sight, sound and somatosensation. The pons is the main connection between the cerebrum and the cerebellum. The pons and the medulla together regulate the cardiovascular and respiratory systems. The medulla also has important functions for swallowing and digestion.

The limbic system includes structures of the cerebrum and diencephalon that are responsible for emotion, motivation, and memory. The main structures are the cingulate gyrus, the hippocampus and the amygdala.

The cerebellum is connected to the brain stem and compares information from the cerebrum with sensory feedback from the spinal cord. The cerebellum coordinates body movements and is involved in movements that are learned with repeated practice.

The spinal cord extends from the brainstem down through the center of the back and pelvis. Sensory neuron axons enter through the dorsal (/posterior) nerve root and motor neuron axons emerge from the ventral (/anterior) nerve root. Gray matter (containing of neuron cell bodies) is located in the center of the spinal cord, surrounded by white matter (myelinated axons). The gray matter is subdivided into the posterior, anterior, and lateral horns. The spinal cord passes messages from the body to the brain and from the brain to the body. Sensory (afferent) nerves carry nerve impulses to the brain from sensory receptor cells everywhere in and on the body. Motor (efferent) nerves carry nerve impulses away from the brain to glands, organs, or muscles throughout the body. The spinal cord also independently controls reflexes without any input from the brain. Physical damage to the spinal cord may result in paralysis, which may be permanent.

4.4.10 References

Basinger H., & Hogg, J.P. (2021) Neuroanatomy, Brainstem. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK544297/

Lent, R., Azevedo, F.A., Andrade-Moraes, C.H., & Pinto, A.V. (2012) How many neurons do you have? Some dogmas of quantitative neuroscience under revision. *European Journal of Neuroscience 35*(1),1-9. doi: 10.1111/j.1460-9568.2011.07923.x. Epub 2011 Dec 13. PMID: 22151227.

4.4.11 Additional Resources

4.4.12 Attributions

1. Figures:
    1. Front of Sensory Homunculus and Rear of Sensory Homunculus by Mpj29, CC BY-SA 4.0, via Wikimedia Commons
    2. Brain and spinal cord: dissection, back view (coloured line engraving by W.H. Lizars, ca. 1827. Iconographic Collections) from Wellcome Images, a website operated by Wellcome Trust, a global charitable foundation based in the United Kingdom, CC BY 4.0, via Wikimedia Commons; AND Human brain and spinal cord by Z22, CC BY-SA 4.0, via Wikimedia Commons
    3. Brain by Laura Guerin, CC BY-NC 3.0 via CK-12
    4. Cerebrum by OpenStax, CC BY 4.0, via Wikimedia Commons

5. Corpus callosum by Images generated by Life Science Databases (LSDB), CC BY-SA 2.1 JP, via Wikimedia Commons
6. Lobes of Cerebral Cortex by OpenStax, CC BY 4.0, via Wikimedia Commons
7. Brain lobes by Laura Guerin, CC BY-NC 3.0 via CK-12
8. Primary motor cortex by CNX OpenStax, CC BY 4.0, via Wikimedia Commons
9. Somatosensory Map from NOBA module Human Sexual Anatomy and Physiology by Lucas, D. & Fox, J., CC BY-NC-SA 4.0, via NOBA
10. Thalamus/hypothalamus adapted by Naomi Bahm (areas adjusted, pineal gland added) from Diencephalon by OpenStax, CC BY 4.0, via Wikimedia Commons
11. Hypothalamus-Pituitary Complex by OpenStax, licensed CC BY 3.0 via Wikimedia Commons
12. Brainstem adapted by Naomi Bahm (midbrain extended to include colliculi) from Brain stem by OpenStax, licensed CC BY 4.0 via Wikimedia Commons
13. Limbic system by BruceBlaus licensed CC BY 3.0 via Wikimedia Commons
14. Cerebellum by OpenStax, CC BY 4.0, via Wikimedia Commons; Note: MRI originally by Semiconscious, Public domain, via Wikimedia Commons
15. Spinal cord by BruceBlaus licensed CC BY 3.0 via Wikimedia Commons
16. The spinal cord by Ruth Lawson Otago Polytechnic, CC BY 3.0, via Wikimedia Commons
17. Spinal readjustment by Tomwsulcer dedicated CC0 via Wikimedia Commons
18. Spinal cord cross section by OpenStax, CC BY 4.0, via Wikimedia Commons; Micrograph provided by the Regents of University of Michigan Medical School © 2012; CC-BY-4.0 Open Oregon, Anatomy and Physiology
19. Stroke Communications Kit by CDC, public domain

2. Text adapted from:
    1. " Central Nervous System" by Suzanne Wakim & Mandeep Grewal, LibreTexts is licensed under CC BY-NC .
    2. " The Cerebrum" by OpenStax, LibreTexts is licensed under CC BY.
    3. " The Diencephalon" by OpenStax, LibreTexts is licensed under CC BY.
    4. " The Brain Stem" by OpenStax, LibreTexts is licensed under CC BY.
    5. " The Cerebellum" by OpenStax, LibreTexts is licensed under CC BY.
    6. " The Spinal Cord" by OpenStax, LibreTexts is licensed under CC BY.
    7. " Brain- Cerebrum" by Whitney Menefee, Julie Jenks, Chiara Mazzasette, & Kim-Leiloni Nguyen, LibreTexts is licensed under CC BY .
    8. " Brain- Diencephalon, Brainstem, Cerebellum and Limbic System" by Whitney Menefee, Julie Jenks, Chiara Mazzasette, & Kim-Leiloni Nguyen, LibreTexts is licensed under CC BY .
3. Changes: Text (and images) from above sources were pieced together with some modifications, transitions and additional content and images added by Naomi I. Gribneau Bahm, PhD., Psychology Professor at Cosumnes River College, Sacramento, CA.

This page titled 4.4: The Central Nervous System is shared under a mixed license and was authored, remixed, and/or curated by Naomi Bahm (ASCCC Open Educational Resources Initiative (OERI)) .

- **12.3: Brain- Cerebrum** by Whitney Menefee, Julie Jenks, Chiara Mazzasette, & Kim-Leiloni Nguyen is licensed CC BY 4.0.
- **12.4: Brain- Diencephalon, Brainstem, Cerebellum and Limbic System** by Whitney Menefee, Julie Jenks, Chiara Mazzasette, & Kim-Leiloni Nguyen is licensed CC BY 4.0.
- **11.5: Central Nervous System** by Suzanne Wakim & Mandeep Grewal is licensed CK-12. Original source: https://www.ck12.org/book/ck-12-human-biology/.

4.5: The Peripheral Nervous System

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe the structures found in the PNS
2. Distinguish between somatic and autonomic structures
3. Name the twelve cranial nerves and explain the functions associated with each
4. Describe the sensory and motor components of spinal nerves
5. Explain the functions of the sympathetic and parasympathetic divisions of the autonomic system

4.5.1 Overview

This module discusses the peripheral nervous system (PNS) in greater depth, including nervous tissues, the two major systems the PNS divides into (the somatic nervous system and the autonomic nervous system), names and functions of the twelve cranial nerves, and characteristics of the 31 pairs of spinal nerves. The subdivisions of the autonomic nervous system (the sympathetic nervous system and the parasympathetic nervous system) and their functions are also covered, as well as a brief mention of disorders of the PNS.

4.5.2 What Is the Peripheral Nervous System?

The **peripheral nervous system (PNS)** consists of all the nervous tissue that lies outside of the central nervous system (CNS). The main function of the PNS is to connect the CNS to the rest of the organism. It serves as a communication relay, going back and forth between the CNS and muscles, organs, and glands throughout the body. Figure **4.5.1** illustrates both the central and peripheral nervous systems, including many representative **spinal nerves**. Although an overview of spinal nerves is discussed in greater detail below, the identification and functions of specific spinal nerves are beyond the scope of this section.

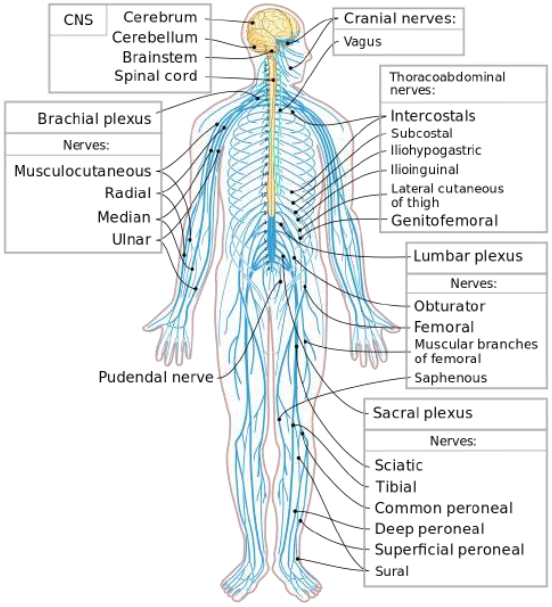


Figure **4.5.1**: The central nervous system (CNS) is shown in yellow: Cerebrum, Cerebellum, Brainstem, Spinal cord. Some representative nerves of the peripheral nervous system (PNS) are shown in blue.

The representative nerves pictured are Cranial nerves: Vagus. Brachial plexus nerves: Musculotaneous, Radial, Median, Ulnar; Thoracoabdominal nerves: Intercostals, Subcostal, Iliohypogastric, Ilioinguinal, Lateral cutaneous of thigh, Genitofemoral; Lumbar plexus nerves: Obturator, Femoral, Muscular branches of femoral, Saphenous; Sacral plexus nerves: Sciatic, Tibial, Common peroneal, Deep peroneal, Superficial peroneal, Sural. The Pudendal nerve is also pictured.

4.5.3 Tissues of the Peripheral Nervous System

The primary structures that make up the PNS are nerves and ganglia. Ganglia are clusters of neuron cell bodies outside the CNS; they act as relay points for messages transmitted through nerves of the PNS. Nerves are cable-like bundles of axons that make up the majority of PNS tissues. Nerves are generally classified based on the type of information they carry (which coincides with the direction in which their nerve impulses travel) as sensory, motor, or mixed nerves. See an example of a sensory and motor nerve in Figure **4.5.2**. (Note that the interneuron that is usually located between the sensory and motor neuron is not shown in this figure.)

- **Sensory nerves** transmit information from sensory receptors in the body to the CNS. Sensory nerves are also called afferent nerves.
- **Motor nerves** transmit information from the CNS to muscles, organs, and glands. Motor nerves are also called efferent nerves.
- **Mixed nerves** contain axons from both sensory and motor neurons, so they transmit information in both directions and have both afferent and efferent functions.

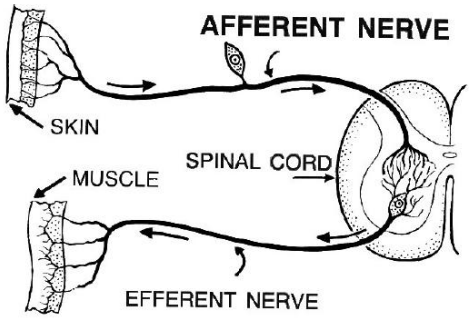


Figure **4.5.2**: A sensory nerve is an afferent nerve. The one pictured here sends nerve impulses from sensory receptors in the skin to the CNS. A motor nerve is an efferent nerve. This one sends nerve impulses from the CNS to a muscle.

4.5.4 Divisions of the Peripheral Nervous System

The PNS is divided into two major systems, called the autonomic nervous system and the somatic nervous system. Both systems of the PNS interact with the CNS and include sensory and motor neurons, but they use different circuits of nerves and ganglia.

4.5.4.1 Somatic Nervous System

The **somatic nervous system** primarily senses the external environment and controls voluntary activities for which decisions and commands come from the cerebral cortex of the brain. For example, when you feel too cold, decide to turn on the heater, and walk across the room to the thermostat, you are using your somatic nervous system. In general, the somatic nervous system is responsible for all of your conscious perceptions of the outside world, and all of the voluntary motor activities you perform in response. Whether it's playing piano, driving a car, or playing basketball, you can thank your somatic nervous system for making it possible. Structurally, the somatic nervous system consists of 12 pairs of cranial nerves and 31 pairs of spinal nerves.

4.5.4.1.1 Cranial Nerves

Cranial nerves (Figure **4.5.3**) are in the head and neck and connect directly to the brainstem. There are twelve cranial nerves, which are designated CN I through CN XII (the "CN" stands for "Cranial Nerve") using the Roman numerals for 1 through 12, based on their anatomical location on the inferior view of the brain, from anterior to posterior. Sensory cranial nerves sense smells, tastes, light, sounds, and body position and sensations. Motor cranial nerves control muscles of the face, tongue, eyeballs, throat, head, and shoulders. The motor nerves also control the salivary glands and swallowing. Four of the 12 cranial nerves participate in both sensory and motor functions as mixed nerves, having both sensory and motor functions. See Table **4.5.1** for a brief summary of the functions of each cranial nerve.

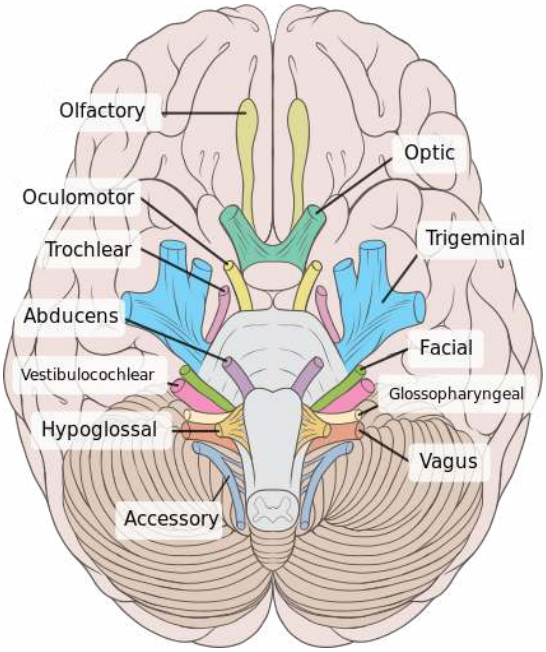


Figure **4.5.3**: The twelve pairs of cranial nerves are located on the inferior aspect of the human brain.

4.5.4.1.2

Table **4.5.1**: Cranial Nerves

Mnemonic	#	Name	Function (**S**ensory/**M**otor/**B**oth)	Central connection (nuclei)	Peripheral connection (ganglion or muscle)
On	I	Olfactory	Smell (S)	Olfactory bulb	Olfactory epithelium
Old	II	Optic	Vision (S)	Hypothalamus/ thalamus/ midbrain	Retina (retinal ganglion cells)
Olympus'	III	Oculomotor	Eye movements; lens and pupil movements (M)	Oculomotor nucleus	Four of the extraocular muscles; levator palpebrae superioris; ciliary ganglion (autonomic)
Towering	IV	Trochlear	Eye movements (M)	Trochlear nucleus	Superior oblique muscle (eyeball)
Tops	V	Trigeminal	Sensory/ motor – face (B)	Trigeminal nuclei in the midbrain, pons, and medulla	Trigeminal
A	VI	Abducens	Eye movements (M)	Abducens nucleus	Lateral rectus muscle (eyeball)
Finn	VII	Facial	Motor – face; Taste (B)	Facial nucleus, solitary nucleus, superior salivatory nucleus	Facial muscles; Geniculate ganglion; Pterygopalatine ganglion (autonomic)
And	VIII	Auditory (Vestibulocochlear)	Hearing/ balance (S)	Cochlear nucleus, Vestibular nucleus/ cerebellum	Spiral ganglion (hearing); Vestibular ganglion (balance)
German	IX	Glossopharyngeal	Motor – throat; Taste (B)	Solitary nucleus, inferior salivatory nucleus, nucleus ambiguus	Pharyngeal muscles; Geniculate ganglion; Otic ganglion (autonomic)
Viewed	X	Vagus	Motor/ sensory – viscera (autonomic) (B)	Medulla	Terminal ganglia serving thoracic and upper abdominal organs (heart and small intestines)
Some	XI	(Spinal) Accessory	Motor – head and neck (M)	Spinal accessory nucleus	Neck muscles
Hops	XII	Hypoglossal	Motor – lower throat (M)	Hypoglossal nucleus	Muscles of the larynx and lower pharynx

4.5.4.1.3 Spinal Nerves

The nerves connected to the spinal cord are the **spinal nerves**. The arrangement of these nerves is much more regular than that of the cranial nerves. All of the spinal nerves are combined sensory and motor axons that separate into two nerve roots. The sensory axons enter the spinal cord as the **dorsal nerve root**. The motor fibers, both somatic and autonomic, emerge as the **ventral nerve**

root. The sensory neuron cell bodies (soma) are grouped in enlargements of the dorsal nerve root called **dorsal root ganglia**. See Figure 4.5.4 to locate these structures on a segment of the spinal cord. (As noted in the central nervous system module, it is common to see the terms dorsal (dorsal = back) and ventral (ventral = belly) used interchangeably with posterior and anterior in reference to nerves and the structures of the spinal cord.)

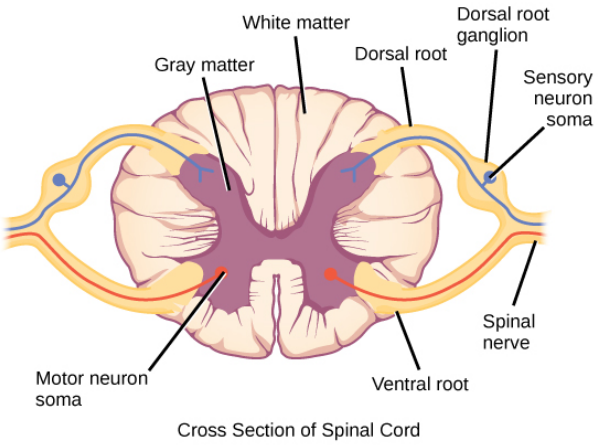


Figure 4.5.4: **Spinal nerves**: Spinal nerves contain both sensory and motor axons. The cell bodies of sensory neurons are located in dorsal root ganglia. The cell bodies of motor neurons are found in the ventral portion of the gray matter of the spinal cord.

Each sensory neuron has one projection with a sensory receptor ending in skin, muscle, or sensory organs, and another that synapses with a neuron in the dorsal spinal cord. Motor neurons have cell bodies (soma) in the ventral gray matter of the spinal cord that project to muscle through the ventral root. These neurons are usually stimulated by interneurons within the spinal cord, but are sometimes directly stimulated by sensory neurons. Spinal nerves include the motor nerves that stimulate skeletal muscle contraction, allowing for voluntary body movements.

There are 31 pairs of spinal nerves, named for the level of the spinal cord at which they emerge, one on each side of the body. There are eight pairs of cervical nerves designated C1 to C8, twelve thoracic nerves designated T1 to T12, five pairs of lumbar nerves designated L1 to L5, five pairs of sacral nerves designated S1 to S5, and one pair of coccygeal nerves. The nerves are numbered from the superior to inferior positions, and each emerges from the vertebral column through the intervertebral foramen at its level. Thus, each spinal nerve corresponds to a segment of the spinal cord, carrying both sensory and motor information (as mentioned above) for the areas of the body represented. As such, all of the spinal nerves are mixed nerves and contain both sensory and motor neuron axons. Figure 4.5.5 illustrates the body areas represented by the spinal nerves and Figure 4.5.6 lists some of the functions of the spinal nerves.

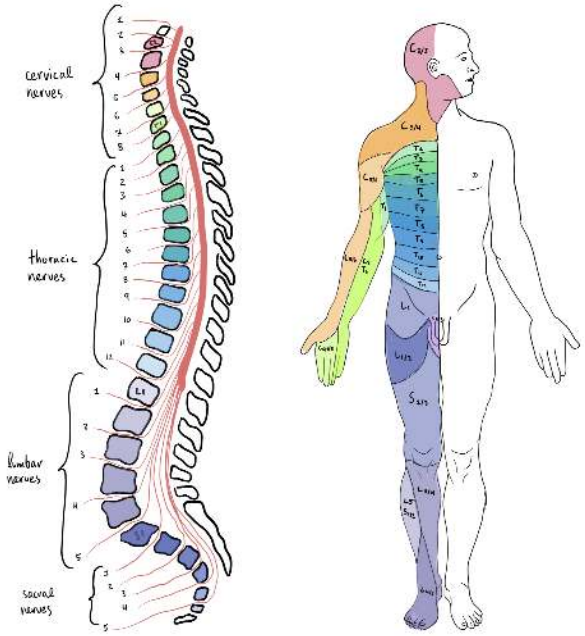


Figure 4.5.5: Spinal cord segments and body representation.

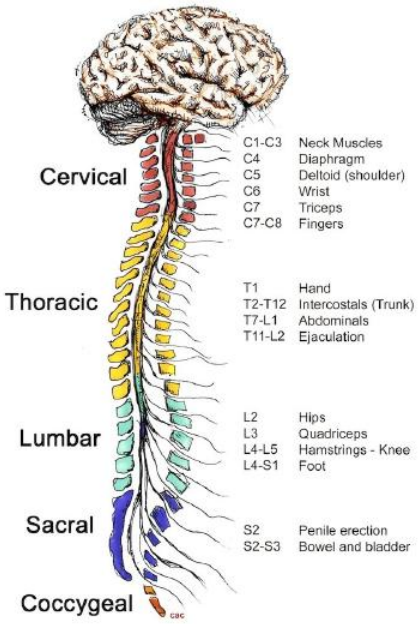


Figure 4.5.6: Vertebral location and some functions of spinal nerves.

4.5.4.2 Autonomic Nervous System

The **autonomic nervous system** primarily senses the internal environment and controls involuntary activities. It is responsible for monitoring conditions in the internal environment and bringing about appropriate changes in them. In general, the autonomic nervous system is responsible for all the activities that go on inside your body without your conscious awareness or voluntary participation.

Structurally, the autonomic nervous system consists of sensory and motor nerves that run between the CNS (especially the hypothalamus in the brain) and internal organs (such as the heart, lungs, and digestive organs) and glands (such as the pancreas and sweat glands). Sensory neurons in the autonomic system detect internal body conditions and send messages to the brain. Motor nerves in the autonomic system function by controlling the contractions of smooth or cardiac muscle or glandular tissue. For example, when sensory nerves of the autonomic system detect a rise in body temperature, motor nerves signal both the smooth muscles in blood vessels near the body surface to undergo vasodilation, and the sweat glands in the skin to secrete more sweat, to cool the body.

The autonomic nervous system, in turn, has two subdivisions: the sympathetic division and parasympathetic division. The functions of the two subdivisions of the autonomic system are summarized in Figure 4.5.5. Both affect most of the same organs and glands, but they generally do so in opposite ways.

- The **sympathetic division** controls the fight-or-flight response. Changes occur in organs and glands throughout the body that prepare the body to fight or flee in response to a perceived danger. For example, the pupils dilate, saliva production decreases, air passages in the lungs become wider (the bronchia dilate), heart rate speeds up, more blood flows to the skeletal muscles, and the digestive system temporarily shuts down. With regard to urinary and sexual functions, the sympathetic division relaxes the urinary bladder and stimulates orgasm in both men and women.
- The **parasympathetic division** returns the body to normal after the fight-or-flight response has occurred. For example, pupils constrict, saliva production increases, air passages in the lungs narrow (the bronchia constrict), heart rate slows down, blood flow to the skeletal muscles is reduced, and the digestive system is stimulated to start working again. The parasympathetic division also maintains the internal homeostasis of the body at other times. With regard to urinary and sexual functions, the parasympathetic division constricts the urinary bladder and stimulates genital tissue erection in both men and women.

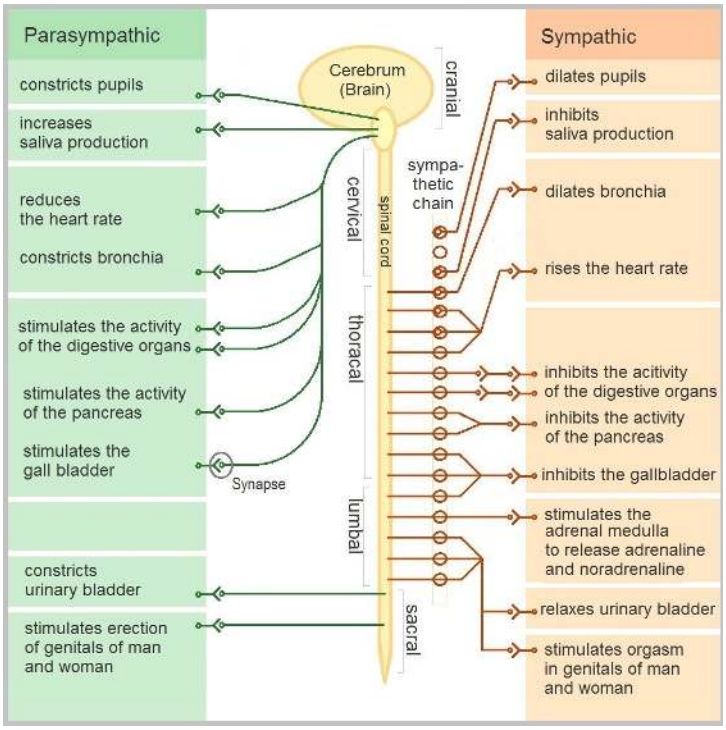


Figure **4.5.7**: This diagram summarizes the structures and functions controlled by the parasympathetic and sympathetic divisions of the autonomic nervous system. The image shows how eyes, salivary gland, heart rate, bronchi, digestive system, bladder, and genitalia are affected in opposite ways by the two systems.

4.5.5 Disorders of the Peripheral Nervous System

Unlike the CNS, which is protected by bones, meninges, and cerebrospinal fluid (CSF), the PNS has no such protections. The PNS also has no blood-brain barrier to protect it from toxins and pathogens in the blood. Therefore, the PNS is more subject to injury and disease than is the CNS. Causes of nerve injury include diabetes, infectious diseases such as shingles, and poisoning by toxins such as heavy metals. Disorders of the PNS often have symptoms such as loss of feeling, tingling, burning sensations, or muscle weakness. If a traumatic injury results in a nerve being transacted (cut all the way through), it may regenerate, but this is a very slow process and may take many months.

4.5.6 Summary

The peripheral nervous system (PNS) is composed of the groups of neurons (ganglia) and bundles of axons (nerves) that are outside of the brain and spinal cord. The PNS connects the CNS to the rest of the body.

Sensory (/afferent) nerves transmit information from sensory receptors in the body to the CNS. Motor (/efferent) transmit information from the CNS to muscles, organs, and glands. Mixed nerves transmit information in both directions and have both sensory (/afferent) and motor (/efferent functions). The PNS is divided into the autonomic nervous system and the somatic nervous system.

The somatic nervous system senses the external environment and controls voluntary activities, and consists of 12 pairs of cranial nerves (connected to the brain) and 31 pairs of spinal nerves (connected to the spinal cord). Cranial nerves can be strictly sensory, strictly motor, or a combination of the two functions. The olfactory nerve (CN I) is responsible for smell and the optic nerve (CN II) is responsible for vision. The oculomotor nerve (CN III) is responsible for eye movements, lifting the upper eyelid, and controlling the size of the pupil. The trochlear nerve (CN IV) and the abducens nerve (CN VI) are both responsible for eye movement, but control different extraocular muscles. The trigeminal nerve (CN V) is responsible for cutaneous (skin) sensations of the face and controlling the muscles of mastication (chewing). The facial nerve (VII) is responsible for the muscles involved in facial expressions, as well as part of the sense of taste. The vestibulocochlear nerve (VIII) is responsible for the senses of hearing and balance. The glossopharyngeal nerve (IX) is responsible for controlling muscles in the throat, as well as part of the sense of taste. The vagus nerve (CN X) is responsible for contributing to homeostatic control of the organs of the thoracic and upper

abdominal cavities. The accessory nerve (CN XI, along with cervical spinal nerves) is responsible for controlling the muscles of the neck. The hypoglossal nerve (CN XII) is responsible for controlling the muscles of the lower throat and tongue.

Spinal nerves are all mixed nerves with both sensory and motor fibers. The sensory axons enter the spinal cord as the dorsal nerve root and the motor fibers (both somatic and autonomic) emerge as the ventral nerve root. The sensory neuron cell bodies (soma) are located in dorsal root ganglia. Spinal nerves emerge from the spinal cord and are numbered from superior to inferior positions.

The autonomic nervous system senses the internal environment and controls involuntary activities. It is divided into the sympathetic division and parasympathetic division. Both affect most of the same organs and glands, but in opposite ways. The sympathetic division controls the fight-or-flight response and the parasympathetic division returns the body to normal after the fight-or-flight response has occurred.

The PNS does not have the same protections (bones, meninges, CSF, blood-brain barrier) that the CNS has, so it is more prone to injury and disease. If a nerve is transacted, it may regenerate, but this is a very slow process.

4.5.7 Additional Resources

Interested in the effects of meditation on the brain? See Integrative and Contemplative Neuroscience

Mindfulness techniques have been shown to reduce symptoms of depression as well as those of anxiety and stress. They have also been shown to be useful for pain management and performance enhancement. Specific mindfulness programs include Mindfulness-Based Stress Reduction (MBSR) and Mindfulness Mind-Fitness Training (MMFT). You can learn more about MBSR by watching the video below.

4.5.8 Attributions

1. Figures:
    1. The nervous system licensed CC BY-SA 4.0 via Lumen Learning
    2. Afferent nerve by Pearson Scott Foresman, Public domain via Wikimedia Commons
    3. Cranial nerves: Brain_human_normal_inferior_view_with_labels_en.svg: *Brain_human_normal_inferior_view.svg: Patrick J. Lynch, medical illustrator derivative work: Beao derivative work: Dwstultz, CC BY 2.5 via Wikimedia Commons
    4. Spinal nerves- no specific attribution (from " Sensory-Somatic Nervous System" by LibreTexts is licensed under notset).
    5. Spinal Cord Segments and body representation by David Nascari and Alan Sved, CC BY-SA 4.0 via Wikimedia Commons
    6. Brain Spinal Cord Labeled by Vankadara Bhavya sree 1840585, CC BY-SA 4.0 via Wikimedia Commons
    7. Autonomic nervous system by Geo-Science-International, dedicated CC0 via Wikimedia Commons
2. Text adapted from:
    1. " Peripheral Nervous System" by Suzanne Wakim & Mandeep Grewal, LibreTexts is licensed under CC BY-NC .
    2. " Sensory-Somatic Nervous System" by LibreTexts is licensed under notset .
    3. " Cranial Nerves" by Whitney Menefee, Julie Jenks, Chiara Mazzasette, & Kim-Leiloni Nguyen, LibreTexts is licensed under CC BY .
    4. " The Peripheral Nervous System" by OpenStax, LibreTexts is licensed under CC BY . (Table also from this source.)
3. Changes: Text (and images) from above four sources pieced together with some modifications, transitions and additional content (and images) added by Naomi I. Gribneau Bahm, PhD., Psychology Professor at Cosumnes River College, Sacramento, CA.

This page titled 4.5: The Peripheral Nervous System is shared under a mixed license and was authored, remixed, and/or curated by Naomi Bahm (ASCCC Open Educational Resources Initiative (OERI)) .

- **11.6: Peripheral Nervous System** by Suzanne Wakim & Mandeep Grewal is licensed CK-12. Original source: https://www.ck12.org/book/ck-12-human-biology/.

4.6: Communication and the Endocrine System

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Explain the difference between chemical signals in the nervous system and chemical signals in the endocrine system
2. Understand the reciprocal interactions between the influence of hormones on behavior and behavior on hormones
3. Identify at least three endocrine glands and describe their primary functions

4.6.1 Overview

Throughout the nervous system, neurons communicate via electrical and chemical signals. Another form of chemical communication is the secretion of hormones into the bloodstream, which is accomplished via endocrine glands located in the endocrine system. This chapter introduces the endocrine system and the major endocrine glands found in the brain and the body.

4.6.2 Communication in the Nervous System versus the Endocrine System

Nerve impulses are covered in depth in the nervous system communication chapter. Briefly, neurons in the nervous system communicate via electrical and chemical signals. When a neuron transmits an electrical signal, called a nerve impulse (or action potential), it travels down the axon and causes neurotransmitters to be released into the synapse. This chemical signal then influences the receiving cell in either an excitatory or inhibitory manner. A cell that receives nerve impulses from a neuron may be excited to perform a function, inhibited from carrying out an action, or otherwise controlled. In this way, the information transmitted by the nervous system is specific to particular cells and is transmitted very rapidly. In fact, the fastest nerve impulses travel at speeds greater than 100 meters per second! Compare this to the chemical messages carried by the hormones that are secreted into the blood by endocrine glands (below). These hormonal messages are “broadcast” to all the cells of the body, and they can travel only as quickly as the blood flows through the cardiovascular system.

4.6.2.1 Neural Transmission versus Hormonal Communication

Hormones are similar in function to neurotransmitters, the chemicals used by the nervous system in coordinating animals’ activities (as mentioned above). However, hormones can operate over a greater distance and over a much greater temporal range than neurotransmitters. Neural and hormonal communication both rely on chemical signals, but several important differences exist. Communication in the nervous system is analogous to traveling on a train. You can use the train in your travel plans as long as tracks exist between your intended origin and destination. Likewise, neural messages can travel only to destinations along existing nerve tracts. Hormonal communication, on the other hand, is like traveling in a car. You can drive to many more destinations than train travel allows because there are many more roads than railroad tracks. Similarly, hormonal messages can travel anywhere in the body via the circulatory system; any cell receiving blood is potentially able to receive a hormonal message.

Not all cells are influenced by each and every hormone. Rather, any given hormone can directly influence only cells that have specific hormone receptors for that particular hormone. Cells that have these specific receptors are called target cells for the hormone. When a hormone engages its receptor, a series of subsequent events either activates enzymes or turns genes on or off to regulate protein synthesis. The newly synthesized proteins may then activate or deactivate other genes, causing additional effects. Importantly, sufficient numbers of appropriate hormone receptors must be available for a specific hormone to produce any effects. For example, testosterone is important for male sexual behavior. If men have too little testosterone, their sexual motivation may be low, which can be restored by testosterone treatment. However, if men have normal or even elevated levels of testosterone yet display low sex drive, then it might be possible for a lack of receptors to be the cause, in which case treatment with additional hormones will not be effective.

Another difference is that neural messages are all-or-none events that have rapid onset and offset: neural signals can take place in milliseconds. For example, the nervous system regulates immediate food intake and directs body movement, changes in the body

that are relatively rapid. In contrast, hormonal messages are graded events that may take seconds, minutes, or even hours to occur. Hormones can mediate long-term processes, such as growth, development, reproduction, and metabolism.

4.6.2.2 Interactions Between Hormones and Behavior

The scientific study of the interaction between hormones and behavior is called behavioral endocrinology. This interaction is bidirectional: hormones can influence behavior, and behavior can sometimes influence hormone concentrations. Hormones are chemical messengers produced and released by specialized glands called endocrine glands (see below). Hormones are released from these glands into the blood, where they travel to act on target structures anywhere in the body, influencing the nervous system to regulate an individual's behaviors (such as aggression, mating, and parenting). Examples of hormones that influence behavior include steroid hormones such as testosterone (a common type of androgen), estradiol (a common type of estrogen), progesterone (a common type of progestin), and cortisol (a common type of glucocorticoid). Several types of protein or peptide (small protein) hormones also influence behavior, including oxytocin, vasopressin, prolactin, and leptin.

There are many ways that hormones influence behavior and behavior feeds back to influence hormone secretion. Some examples include hormones in the mediation of food and fluid intake, social interactions, salt balance, learning and memory, stress coping, as well as psychopathology including depression, anxiety disorders, eating disorders, postpartum depression, and seasonal depression. Some other hormone-behavior interactions that are related to reproductive behaviors are described in Chapter 13.1.

4.6.3 Introduction to the Endocrine System

As the endocrine system is not a part of the nervous system, this section may surprise you. However, as textbook author James W. Kalat states, “Hormonal influences resemble synaptic transmission in many ways, including the fact that many chemicals serve both as neurotransmitters and hormones” (Kalat, 2019, page 59). Thus, to understand many of the topics that will be covered in upcoming chapters, you will need a basic understanding of the endocrine system.

The **endocrine system** is a system of glands called **endocrine glands** that release chemical messenger molecules called **hormones** into the bloodstream. Other glands of the body, including sweat glands and salivary glands, also secrete substances but not into the bloodstream. Instead, they secrete them through ducts that carry them to nearby body surfaces. These other glands are called **exocrine glands.**

Endocrine hormones must travel through the bloodstream to the cells they affect, and this takes time. Because endocrine hormones are released into the bloodstream, they travel throughout the body wherever blood flows. As a result, endocrine hormones may affect many cells and have body-wide effects. Endocrine hormones may cause effects that last for days, weeks, or even months.

4.6.3.1 Major Glands of the Endocrine System

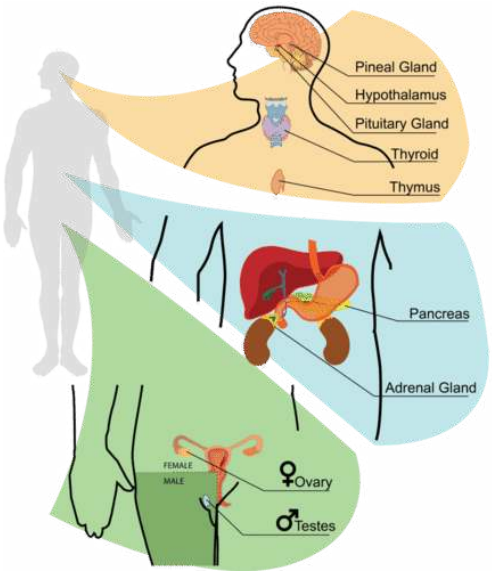


Figure 4.6.1: Endocrine Glands. All of the glands in this figure are part of the endocrine system. Note that the ovary and testis are the only endocrine glands that differ in males and females.

The major glands of the endocrine system are shown in Figure **4.6.1**. The glands in the figure are described briefly in the rest of this section. Refer to the figure as you read about the glands in the following text.

4.6.3.1.1 Endocrine Glands in the Brain

The **pituitary gland** is located at the base of the brain. It is controlled by the nervous system via the brain structure called the **hypothalamus**, to which it is connected by a thin stalk (called the **infundibulum**). The pituitary gland consists of two lobes, called the anterior (front) lobe and posterior (back) lobe. The posterior lobe, composed of nervous tissue, stores and secretes hormones synthesized by the hypothalamus, specifically the hormones oxytocin and vasopressin (also called antidiuretic hormone).

The anterior lobe, composed of glandular tissue, synthesizes and secretes many of its own endocrine hormones, also under the influence of the hypothalamus. One endocrine hormone secreted by the anterior pituitary gland is growth hormone, which stimulates cells throughout the body to synthesize proteins and divide. Most of the other endocrine hormones secreted by the anterior pituitary gland are called **tropic hormones** because they control other endocrine glands. Generally, tropic hormones direct other endocrine glands to secrete either more or less of their own hormones, such as adrenocorticotrophic hormone (ACTH) stimulating the adrenal gland to produce the "stress" hormone cortisol. This is why the pituitary gland is often referred to as the “master gland” of the endocrine system. Since the hypothalamus is the structure that controls both the anterior and posterior parts of the pituitary, you can think of it as the "master" of the "master gland".

The **pineal gland** (also called the **epithalamus**) is a tiny gland located near the center of the brain. It secretes the hormone melatonin, which controls the sleep-wake cycle and several other processes. The production of melatonin is stimulated by darkness and inhibited by light. Cells in the retina of the eye detect light and send signals to a structure in the hypothalamus named the **suprachiasmatic nucleus** (SCN). Nerve fibers carry the signals from the SCN to the pineal gland via the autonomic nervous system.

4.6.3.1.2 Endocrine Glands in the Body

Each of the other major glands of the endocrine system is summarized briefly below. Several of these endocrine glands are also discussed in greater detail as they relate to other topics in separate chapters.

- The **thyroid gland** is a large gland in the neck. Thyroid hormones such as **thyroxine** increase the rate of metabolism in cells throughout the body. They control how quickly cells use energy and make proteins.
- The **thymus gland** is located in front of the heart. It is the site where immune system cells called T cells mature. T cells are critical to the adaptive immune system, facilitating the body's adaption to specific pathogens.
- The **pancreas** is located near the stomach. Its endocrine hormones include **insulin** and **glucagon**, which work together to control the level of glucose in the blood. The pancreas also secretes digestive enzymes into the small intestine.
- The two **adrenal glands** are located above the kidneys. Adrenal glands secrete several different endocrine hormones, including the hormone **epinephrine** (also known as adrenaline), which is involved in the fight-or-flight response. Other endocrine hormones secreted by the adrenal glands have a variety of functions. For example, the hormone **aldosterone** helps to regulate the balance of minerals in the body. The "stress" hormone **cortisol** is also an adrenal gland hormone.
- The paired **gonads** include the **ovaries** in females and **testes** in males. They secrete sex hormones, such as **testosterone** (in males) and **estrogen** (in females). These hormones control sexual maturation during puberty and the production of **gametes** (sperm or egg cells) by the gonads after sexual maturation.

4.6.4 Summary

Neurons communicate via electrical and chemical signals. The endocrine system also communicates via chemical signals, using hormones traveling through the bloodstream. Hormones are secreted by endocrine glands, and have similar features to neurotransmitters (which are released at synapses between neurons in the nervous system). Both neural and hormonal communication rely on chemical signals. However, hormones can operate over a greater distance (anywhere in the body via the circulatory system) and over a much greater time length, and can thus mediate long-term processes. A particular hormone can only influence cells that have receptors for that specific hormone. The interaction between hormones and behavior is bidirectional: hormones can influence behavior, and behavior can sometimes influence hormone concentrations.

A basic understanding of the endocrine system is necessary to understand many of the topics covered in other chapters. Endocrine glands in the brain include the pituitary gland (located at the base of the brain), the hypothalamus (immediately above the pituitary

gland), and the pineal gland (near the center of the brain). The pituitary gland is controlled by the hypothalamus, to which it is connected by a thin stalk (the infundibulum). The pituitary gland consists of an anterior (front) lobe and a posterior (back) lobe. The posterior lobe (composed of nervous tissue) stores and secretes hormones synthesized by the hypothalamus, specifically oxytocin and vasopressin (also called antidiuretic hormone). The anterior lobe (composed of glandular tissue) synthesizes and secretes many of its own endocrine hormones, one of which is growth hormone, which stimulates cells throughout the body to synthesize proteins and divide. Most of the anterior pituitary gland hormones control other endocrine glands, such as adrenocorticotrophic hormone (ACTH), which stimulates the adrenal gland to produce cortisol. The pineal gland secretes the hormone melatonin, which controls the sleep-wake cycle.

Some major endocrine glands in the body include the thyroid gland, the thymus gland, the pancreas, the adrenal glands, and the gonads. The thyroid gland is in the neck and secretes hormones that control how quickly cells use energy and make proteins. The thymus gland is located in front of the heart, and is the site where T cells (of the immune system) mature. T cells facilitate the body's adaption to specific pathogens. The pancreas is located near the stomach. Its endocrine hormones (insulin and glucagon) work together to control the level of glucose in the blood. The two adrenal glands (located above the kidneys) secrete several different endocrine hormones, including the hormone epinephrine (adrenaline), involved in the fight-or-flight response, and the "stress" hormone cortisol. The paired gonads include the ovaries in females and testes in males. The gonads secrete sex hormones (such as testosterone in males and estrogen in females), which control sexual maturation during puberty and the production of gametes (sperm or egg cells).

4.6.5 References

Kalat, J.W. (2019). *Biological Psychology* (13th ed.). Cengage.

4.6.6 Additional Resources

Most people want to live a long, healthy life. Geneticist Cynthia Kenyon’s research suggests that endocrine hormones may be a key to human longevity. Watch this fascinating TED talk to learn how.

4.6.7 Attributions

1. Endocrine glands by Mariana Ruiz Villarreal CC BY-NC 3.0 via CK-12 Foundation
2. Text adapted from:
    1. " Introduction to the Nervous System" by Suzanne Wakim & Mandeep Grewal, LibreTexts is licensed under CC BY-NC .
    2. " Introduction to the Endocrine System" by Suzanne Wakim & Mandeep Grewal, LibreTexts is licensed under CC BY-NC .
    3. Hormones & Behavior by Randy J. Nelson, licensed CC BY-NC-SA 4.0 via Noba Project.
3. Changes: Text (and image) from above three sources with some minor modifications and additional content added by Naomi I. Gribneau Bahm, PhD., Psychology Professor at Cosumnes River College, Sacramento, CA.

This page titled 4.6: Communication and the Endocrine System is shared under a mixed license and was authored, remixed, and/or curated by Naomi Bahm (ASCCC Open Educational Resources Initiative (OERI)) .

- **12.2: Introduction to the Endocrine System** by Suzanne Wakim & Mandeep Grewal is licensed CK-12. Original source: https://www.ck12.org/book/ck-12-human-biology/.

CHAPTER OVERVIEW

5: Communication within the Nervous System

> This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

5.1: Neurons and their Basic Functions

5.2: Neurons Generate Voltage Changes to Code Information

5.3: Synaptic Transmission

This page titled 5: Communication within the Nervous System is shared under a mixed license and was authored, remixed, and/or curated by Kenenth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

5.1: Neurons and their Basic Functions

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

5.1.1 Learning Objectives

1. Describe the anatomy of neurons and the function of each of the three major parts of a neuron
2. Describe neurotransmitters, synapse, synaptic vesicles, resting potential, EPSP, IPSP, and action potential
3. Explain saltatory conduction and why it is important to communication within the nervous system; include description of the myelin sheath and nodes of Ranvier
4. Describe the role of post-synaptic receptor sites in communication between neurons
5. Describe neurons classified by shape and by their functions

5.1.2 **Overview**

In this section, we continue our exploration of neurons, the building blocks of the nervous system. We examine how they generate electrochemical signals, and how the billions of neurons in the nervous system communicate with one another, a process known as synaptic transmission. Before tackling these topics, we review and expand the basic anatomy and functioning of neurons covered in part in Chapter 4. A sound grasp of these facts provides the groundwork for understanding how neuron potentials are generated within neurons and how they combine to trigger synaptic transmission. As you read, remember that the voltages and chemical events we discuss in this section, operating in large populations of brain cells, somehow generate our perceptions, thoughts, emotions, and the entirety of our mental experience. To date, how this happens, how patterns of neuron potentials in brain circuits become conscious minds, remains the greatest mystery of all facing modern science.

5.1.3 Introduction to Neuron Anatomy, Neuron Potentials, and Synaptic Transmission

The nervous system of vertebrates like us is comprised of the **central nervous system**, made up of the brain and spinal cord, and the **peripheral nervous system**, which consists of all the nerves outside the brain and spinal cord.

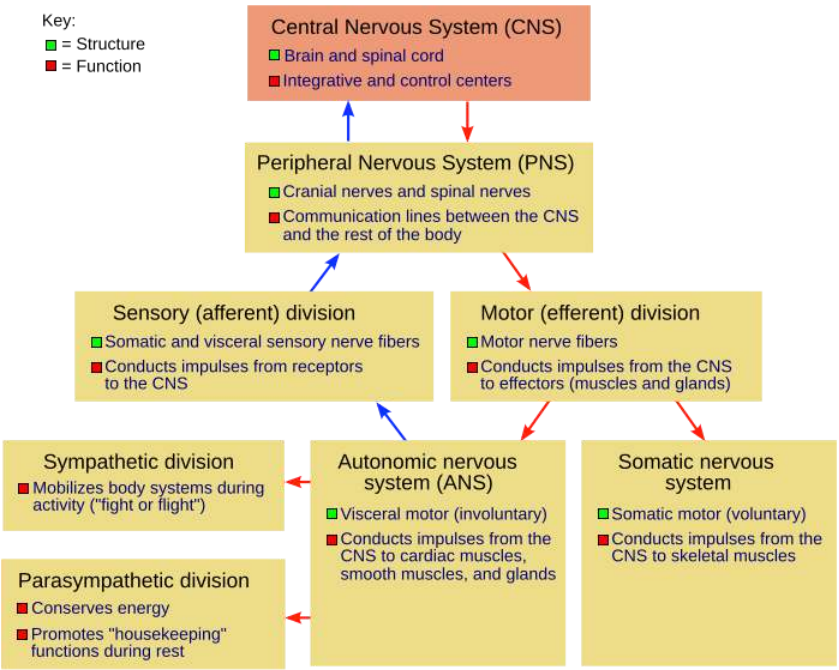


Figure 5.1.1: The human nervous system and its major divisions. All components consist of neurons--nerve cells. (NSdiagram NGB colors fixed.png adapted by Naomi Bahm (colors boxed and fixed) from File:NSdiagram.svg;

https://commons.wikimedia.org/wiki/File:NSdiagram.svg; by Fuzzform via Wikimedia Commons; Permission is granted to copy, distribute and/or modify this document under the terms of the **GNU Free Documentation License**, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled *GNU Free Documentation License*. Subject to disclaimers).

The peripheral nervous system is composed of the **somatic sensory nerves** (input nerves such as the auditory nerve, the optic nerve, spinal sensory nerves mediating skin sensations, etc.) and **somatic motor nerves** (output nerves, serving the skeletal muscles, activating or inhibiting them) as well as the **autonomic sensory and motor nerves** (serving the internal body organs such as the heart, lungs, blood vessels, digestive system, adrenal glands, etc.). These autonomic motor nerves, which either activate or inhibit the internal body organs, are of two types: the **sympathetic** (activate heart, lungs, constrict blood vessels, etc. to prepare the body for "fight or flight") and **parasympathetic** (conserve the body's resources during non-emergency, non-threatening situations) motor autonomic nerves. All of these parts of the nervous system, including the brain, spinal cord, and the peripheral nervous system, are made up of nerve cells, technically known as **neurons**. The **human brain** is estimated to have somewhere between **80 and 100 billion neurons**. Recall from Chapter 4 that neurons have three major parts: the soma or cell body, the dendrites which receive inputs from other neurons, and the axon, which carries the output of neurons, nerve impulses (action potentials), to other neurons, to muscles, or to glands.

You have probably seen photographs or drawings of neurons previously in your introductory psychology course. The images below show small sections of cerebral cortex in a human adult and a human infant. Note the density of neurons and connections in each brain slice. The complexity of interconnections among neurons is evident in the image at the far right which is based on use of Golgi stain.

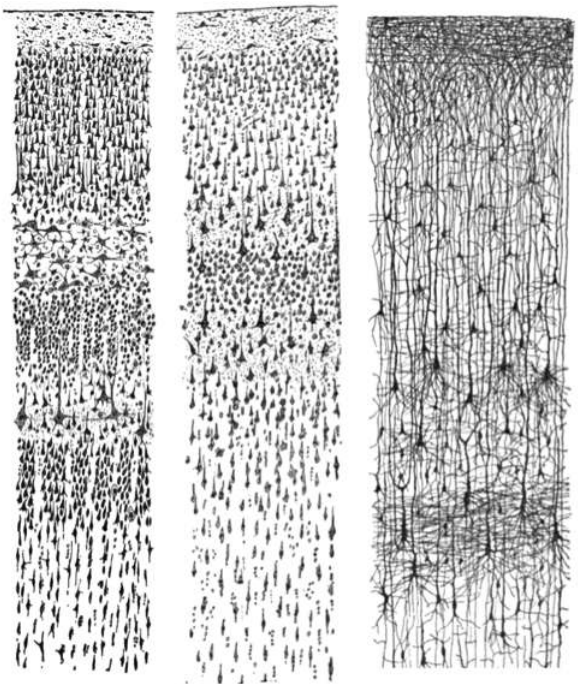


Figure **5.1.2**: Small sections of tissue from the brain of a human infant and a human adult. The image at the far right using Golgi stain highlights cell bodies and extensive branching of dendrites. Left: Nissl-stained visual cortex of a human adult. Middle: Nissl-stained motor cortex of a human adult. Right: Golgi-stained cortex of a 1 1/2 month old infant. [Image: Santiago Ramon y Cajal, https://goo.gl/zOb2l1, CC0 Public Domain, https://goo.gl/m25gce]

5.1.3.1 Nissl and Golgi Stains

Note that different chemicals were used to stain the neurons in the figure above so that they could be seen under the microscope. Nissl stain labels only the main part of the cell (i.e., the cell body; see left and middle panels of Figure 5.1.2). However, by contrast, a Golgi stain fills the cell body and all the processes that extend outward from the cell body (see right panel of Figure 5.1.2). A more notable characteristic of a Golgi stain is that it only stains approximately 1–2% of neurons (Pasternak & Woolsey,

1975; Smit & Colon, 1969), permitting the observer to distinguish one cell from another. These qualities of the Golgi stain allowed the pioneering neuroanatomist Santiago Ramon y Cajal to examine the full anatomical structure of individual neurons for the first time. This significantly enhanced our appreciation of the intricate networks their processes form. Based on his observation of Golgi-stained tissue, Cajal suggested neurons were distinguishable processing units rather than part of a continuous network of nerves, as had been previously believed. Cajal and Golgi shared the Nobel Prize for Medicine in 1906 for their combined contribution to the advancement of science and our understanding of the structure of the nervous system.

5.1.3.2 Neuron Anatomy and Basic Functions

To understand neuron function, it is important to be familiar with the anatomy involved. As noted above, neurons have three major structural parts - - the **soma** or cell body, the **axon** (carries the neuron's output), and the **dendrites** (the "receivers" of the neuron). The entire neuron is bounded by a **cell membrane**, the neural membrane. The cell membrane of a neuron has **channels or "doors" for ions** (electrically charged atoms) which can pass through the membrane when specific channels are opened for specific ions. Figure 5.1.4 shows this basic neuron anatomy. Various types of neurons are discussed below.

The soma or cell body contains organelles, common to all types of cells in the body. These are involved in the basic metabolism of the cell. The soma also contains the nucleus, where the genes and chromosomes (containing DNA) are located.

The second main part of the neuron is the **dendrites**, the receivers of the neuron. Dendrites in some neurons can branch profusely (large numbers of dendritic branches off the main shaft of a dendrite with their own branches are often collectively called **dendritic trees**), expanding the region of the neuron that can receive inputs from other neurons. The **receptor sites** (or more technically, **postsynaptic receptor sites** because of their location on the receiving or postsynaptic neuron) which receive molecules of **neurotransmitter** are located on the dendrites (and, to a lesser degree, on the soma) of the receiving neuron. On the dendrites are small **dendritic spines** which are associated with the connections between neurons (the **synapses**) and can change shape rapidly when learning occurs (see Chapter 10 for more detail). Note that the spines are not the same thing as dendritic branches. In Figure 5.1.3, a dendritic branch with dendritic spines is shown on the left in microscopic detail, and on the right, are dendritic trees of two types of neuron found in the retina. Spines are not visible in the images of dendritic trees on the right because the dendritic spines are too small, while dendritic branches comprising the dendritic trees can easily be seen (see caption for Figure 5.1.3).

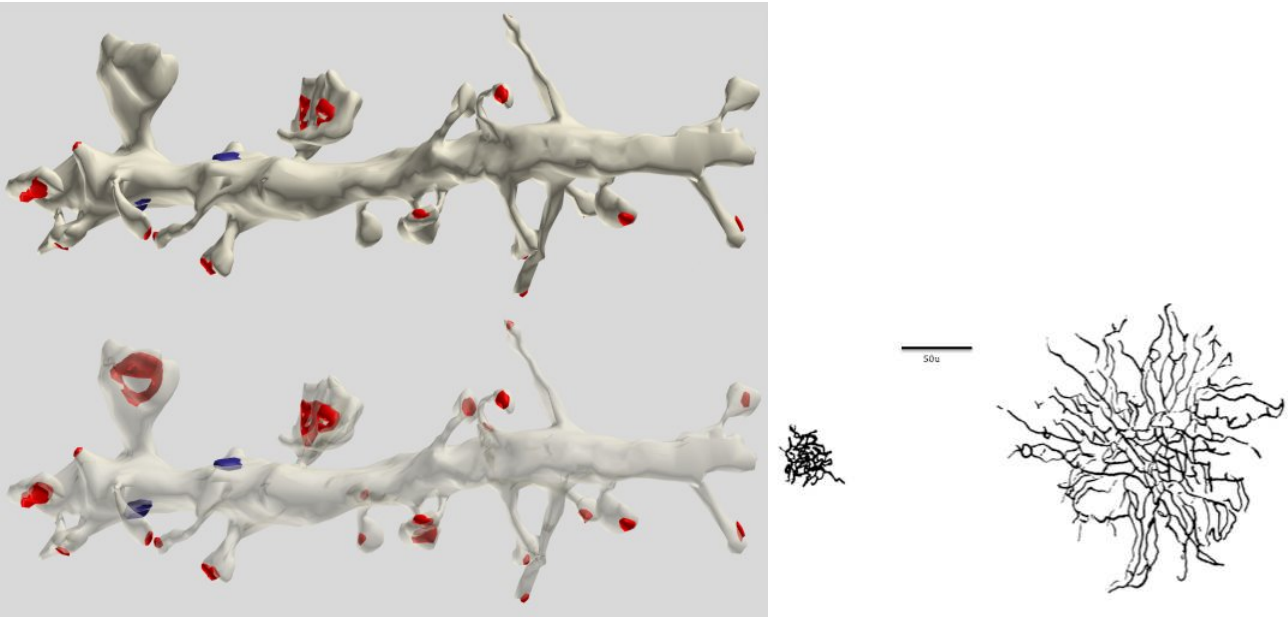


Figure **5.1.3**: (Left) A segment of pyramidal cell dendrite from stratum radiatum (CA1) of the hippocampus with thin, stubby, and mushroom-shaped spines. Spine synapses are colored in red, stem (or shaft) synapses are colored in blue. The dendrite was made transparent in the lower image to enable visualization of all synapses. (Right) A size comparison between midget and parasol cell dendritic trees. Dendritic spines are too small to be visible in the two images on the right. Parasol and midget neurons are found in the ganglion cell layer of the retina. (On left, Pyramidal cell dendrite and spines: Image and caption from Wikimedia Commons;

File:A segment of pyramidal cell dendrite from stratum radiatum (CA1).jpg; https://commons.wikimedia.org/wiki/F...atum_(CA1).jpg; by Synapse Web, Kristen M. Harris, PI, http://synapses.clm.utexas.edu/; licensed under Creative Commons Attribution 2.0 Generic license. On right, Midget and parasol cell dendritic trees: Image and one line caption from Wikimedia Commons; File:Midget vs Parasol cell.png; https://commons.wikimedia.org/wiki/F...rasol_cell.png; by Stromdabomb; licensed under the Creative Commons Attribution-Share Alike 4.0 International license; two sentence explanation by Kenneth A. Koenigshofer, PhD).

The third major part of the neuron is the **axon**, coming out of the soma like a hose. The axon carries the output messages of a neuron (nerve impulses) along its length to its axon terminal buttons (axon endings). There is only one axon per neuron, although it can branch into multiple axon terminal buttons. In a typical neuron, the root end of the axon emerges out of the soma at a small swelling called the **axon hillock**. Between the axon hillock and the first segment of the axon is where the nerve impulse is first generated (see discussion of the action potential, the nerve impulse, that follows below).

File:Components of neuron.jpg

Figure 5.1.4: Basic structure of a neuron (Image from Wikimedia Commons; File:Components of neuron.jpg; https://commons.wikimedia.org/wiki/F..._of_neuron.jpg; by Jennifer Walinga; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

Myelinated Axons and Saltatory Conduction

Some axons have a glial cell covering (glials are non-neural cells in the nervous system, see Chapter 4.1) known as the **myelin sheath**. This fatty, insulating, myelin sheath has gaps in it, revealing the bare axon, at regular intervals along the axon's length. These bare spots along the length of a **myelinated axon** are called nodes, or **nodes of Ranvier**, after their discoverer.

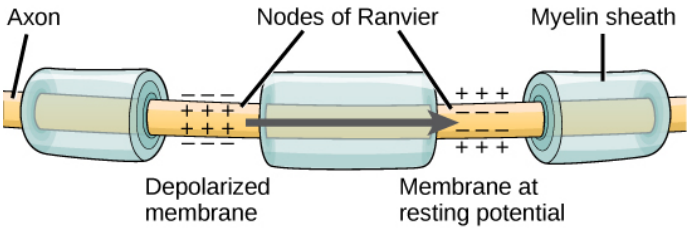


Figure 5.1.5: **Nodes of Ranvier.** Nodes of Ranvier are gaps in the myelin sheath which covers myelinated axons. Nodes contain voltage-gated potassium (K+) and sodium (Na+) channels. Action potentials travel down the axon by "jumping"from one node to the next, speeding conduction of the action potential down the length of the axon toward the axon ending, also known as the axon bouton, axon button, or axon terminal.

The function of the myelin sheath and the nodes is to speed up the rate at which nerve impulses travel down the length of the axon toward their destination, the axon ending (axon bouton). In myelinated axons, the impulses sort of "jump" from node to node allowing the action potential to move more rapidly down the axon. This leaping of the nerve impulse (action potential) from node to node is called **saltatory conduction,** from the Latin "saltatore" which means to dance. Imagine the romantic image of the impulse dancing from node to node.

A node of Ranvier is a natural gap in the myelin sheath along the axon. These unmyelinated spaces are about one micrometer long and contain voltage-gated (i.e., opened by voltage) sodium (Na^+) and potassium (K^+) ion channels (ions are electrically charged atoms). The flow of ions through these channels, particularly the Na^+ channels, regenerates the action potential over and over again along the axon at each successive node of Ranvier. As noted above, the action potential “jumps” from one node to the next in saltatory conduction. If nodes of Ranvier are not present along an axon (as is the case in unmyelinated axons; see below), the action potential propagates much more slowly; ion movement through Na^+ and K^+ channels continuously regenerates new action potentials at every successive point along the axon, using extra time to do so. In effect, conduction of the action potential along an axon involves voltage-gated channels (channels that respond to voltage, rather than to neurotransmitter), on the axon, responding to voltage arising from an electrical field which spreads from the action potential in the previous segment of the axon. In a myelinated axon, because the action potential jumps from node to node, it does not have to be regenerated at every successive point along the axon, but only at the nodes, skipping across myelinated segments of the axon between nodes. Because the action potential in a

myelinated axon must be regenerated fewer times to move any particular distance along the length of the axon, it reaches its destination faster, compared to the speed of conduction in an unmyelinated axon.

Nodes of Ranvier also save energy for the neuron since the ion channels only need to be present and opened and closed at the nodes and not along the entire axon. It is extraordinary that the nodes are placed along the axon's length at just the right spatial intervals to make impulse conduction down the axon the most efficient and speedy as possible. One can only wonder at the incredible precision with which natural selection operated on this feature of myelinated axons over the long course of animal evolution to create this optimal spacing of the nodes.

Neural Conduction in Unmyelinated Axons

Not all axons are myelinated. **Unmyelinated axons** tend to be older in evolution and to be the smaller diameter axons (classified as C fibers based on their small diameters; large diameter myelinated axons are called A fibers). In unmyelinated axons, in order to move, the nerve impulse must be regenerated at every successive point along the axon. This takes time and slows the conduction of the nerve impulse (the action potential) down the length of the axon. Therefore, conduction of the action potential down the length of an unmyelinated axon is relatively slow. An example of unmyelinated C fibers are axons that are part of slow pain pathways. These pathways mediate the slower aching pain that follows tissue damage. The quick, sharp pain from an injury is mediated by larger diameter A fibers (axons).

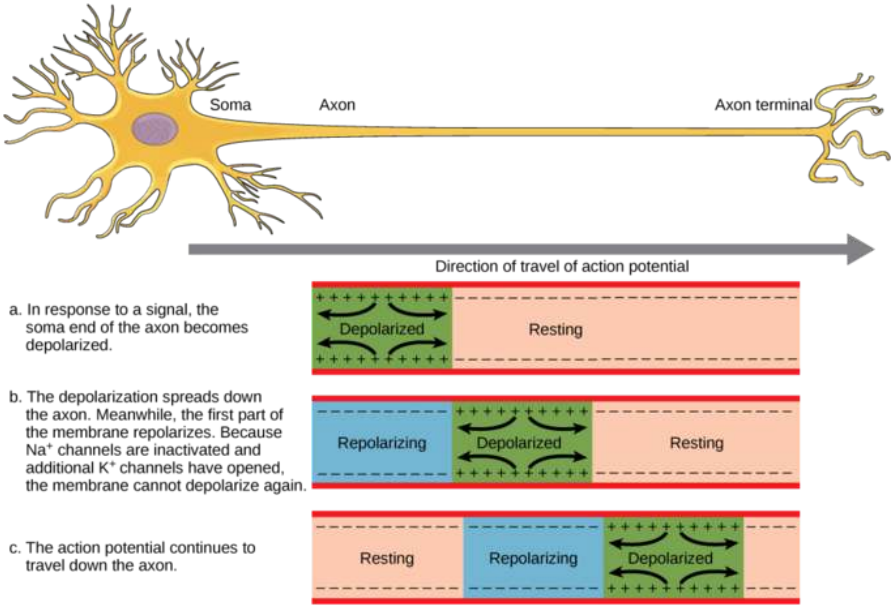


Figure **5.1.6**: Action potential traveling along an unmyelinated neuronal axon. The action potential is conducted down the axon as the axon membrane depolarizes, then repolarizes. Because of these dynamics, the action potential can only be conducted in one direction, away from the cell body. Image: figure-35-02-04_NGB_added_resting.png adapted by Naomi Bahm (resting added to part a) from https://s3-us-west-2.amazonaws.com/c...e-35-02-04.png

The ion channels along the length of the unmyelinated axon are called voltage-gated channels because the voltage generated by the action potential in the prior segment of the axon triggers opening of these channels in the next segment. This leads to regeneration of the action potential. As this process is repeated over and over along the length of the unmyelinated axon, the action potential is conducted along the length of the axon away from the cell body toward the axon endings of the neuron. Figure **5.1.7** shows the channels opening and closing and the ions moving across the cell membrane generating an action potential in successive segments of the axon. This causes movement of the action potential, its conduction, along the axon away from the cell body (from right to left in the figure, as the arrow shows).

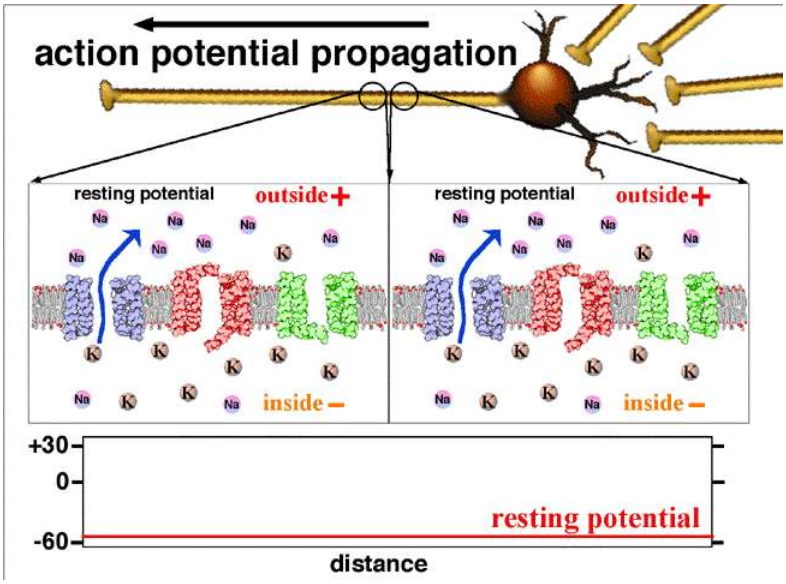


Figure 5.1.7: Another depiction of propagation of the nerve impulse (the axon potential) along an axon. This animation illustrates action potential propagation in an axon. Three types of ion channel are shown: potassium "leak" channels (blue), voltage-gated sodium channels (red) and voltage-gated potassium channels (green). The movement of positively-charged sodium and potassium ions through these ion channels controls the membrane potential of the axon. Action potentials are initiated in the axon's initial segment after neurotransmitter activates excitatory receptors in the neuron's dendrites and cell body. This depolarizes the axon initial segment to the threshold voltage for opening of voltage-gated sodium channels. Sodium ions entering through the sodium channels shift the membrane potential to positive-inside. The positive-inside voltage during the action potential in the initial segment causes the adjacent part of the axon membrane to reach threshold voltage. When positive-inside membrane potentials are reached, voltage-gated potassium channels open and voltage-gate sodium channels close. Potassium ions leaving the axon through voltage-gated potassium channels return the membrane potential to negative-inside values. When the voltage-gated potassium channels gate shut, the membrane potential returns to the resting potential. (Image and caption from Wkimedia Commons; File:Action potential propagation animation.gif; https://commons.wikimedia.org/wiki/F..._animation.gif; by John Schmidt; licensed under the Creative Commons Attribution-Share Alike 4.0 International, 3.0 Unported, 2.5 Generic, 2.0 Generic and 1.0 Generic license).

By contrast, as described above, in myelinated axons, the impulse gets regenerated only at the bare spots on the axon, the nodes of Ranvier. Because the action potential gets regenerated fewer times in order to travel a given distance, than is the case for unmyelinated axons, neural conduction is faster with the insulating myelin sheath. Myelinated axons tend to be large diameter (A and B fibers) and found in neural pathways mediating rapid behavioral response, such as the Pyramidal tract that runs from motor cortex to spinal cord motor neurons which generate voluntary action.

5.1.3.3 Synapses

The axon splits many times, so that it can communicate, or synapse, with several other neurons (see Figure 5.1.4). At the end of the axon, the **axon terminal button** (also called the axon ending, the terminal bouton, or the axon terminal) forms synapses with the dendritic **spines**, small protrusions mentioned above, on the dendrites of receiving neurons (postsynaptic neurons). Synapses form between the axon terminal button of the ***presynaptic neuron*** (neuron sending the signal) and the ***postsynaptic*** membrane (membrane of the neuron receiving the signal; see Figure 5.1.4). Here we will focus specifically on synapses between the axon terminal button of an axon and a dendritic spine; however, synapses can also form between the axon terminal button of the presynaptic neuron and the postsynaptic neuron's soma, dendritic shaft directly, or the axon of another neuron.

A very small space called a **synaptic gap** or a synaptic cleft exists between the pre-synaptic axon's terminal button and the post-synaptic neuron's dendritic spine. To give you an idea of the size of the synaptic gap, a dime is 1.35 mm (millimeter) thick. There are 1,350,000 nm (nanometers) in the thickness of a dime. The synaptic gap is only about 5 nm wide. In the pre-synaptic terminal button, there are **synaptic vesicles** that package together groups of chemicals, **neurotransmitters** (see Figure 5.1.4).

Neurotransmitters are released from the pre-synaptic axon's terminal button or axon ending into the synaptic gap; molecules of neurotransmitter then travel across the synaptic gap, and open ion channels on the post-synaptic spine by binding to *receptor sites* there. We will discuss the role of these receptors in more detail later in section 5.2.

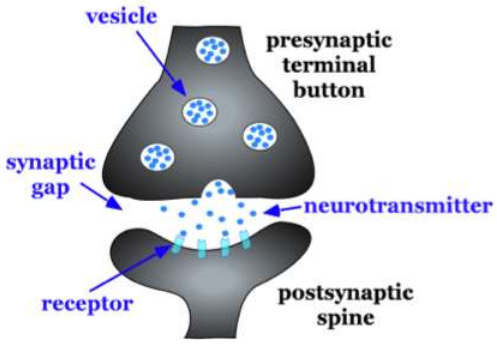


Figure 5.1.8: Basic characteristics of a typical synapse. Enlargement of the synapse between one of the axon terminal buttons (labeled presynaptic terminal button) and one of the dendrites of the second neuron shown (on the right) in Figure 5.1.4 above.

Axon Terminal Buttons, Synaptic Vesicles, PSPs, and Synaptic Transmission

Recall that the end of an axon or axon branch is called the axon terminal button (or terminal, or simply the axon ending). Within the axon terminal button are structures called **synaptic vesicles**, which contain neurotransmitter chemicals (which have been manufactured in the soma and transported to the axon ending and stored there in the synaptic vesicles, ready for release). When neurons communicate with one another across the synaptic gap which separates them, it is the **neurotransmitter**, released from the synaptic vesicles in the axon terminal button of the "sender" neuron (the presynaptic neuron) that transmits the neural message across the synaptic gap (the space between the membrane of the axon terminal button of the pre-synaptic neuron and the membrane of the dendrite or soma of the postsynaptic neuron). The **event that triggers release of neurotransmitter** from the synaptic vesicles in the axon terminal button is the **arrival of an action potential at the axon terminal button** of the sender neuron. The nerve cell releasing the neurotransmitter, the sender neuron, is known technically as the **presynaptic neuron**. The neuron receiving the neurotransmitter, the receiver cell, is called the **postsynaptic neuron**.

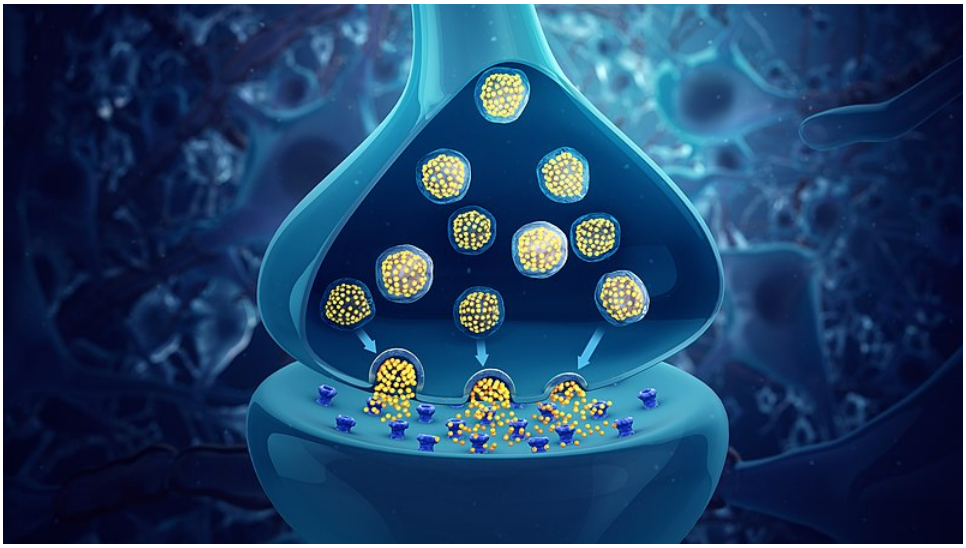


Figure 5.1.9: A synapse. Synaptic vesicles release neurotransmitters (small yellow balls) which bind to the receptors (blue peg-like structures) on the postsynaptic membrane. Synaptic vesicles inside a presynaptic axon terminal button (axon ending) releasing neurotransmitter molecules onto receptors on a dendrite of a receiving (post-synaptic) neuron. (Image from Wikimedia Commons; File:Neurotransmitters.jpg; https://commons.wikimedia.org/wiki/F...ansmitters.jpg; by https://www.scientificanimations.com/; by https://www.scientificanimations.com/; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

The neurotransmitter molecules attach to special sites on the membrane of the postsynaptic neuron. These special sites are called **receptor sites or postsynaptic receptor sites** and are typically located on the dendrites and dendritic spines of the receiving neuron, but can also be on its soma. Their molecular shapes match the shapes of the neurotransmitter molecules they receive, a kind of "lock and key" fit, which opens chemically-gated ion channels allowing ions with their electrical charges to move across the cell membrane creating voltage shifts called **postsynaptic potentials** (PSPs; additional details follow in Section 5.2).

These events from release of transmitter to generation of post-synaptic potentials comprise **synaptic transmission**. Many **psychoactive drugs** (drugs that alter mind and/or behavior) such as amphetamines, LSD, "magic mushrooms," etc. produce their effects by **blocking or activating specific receptor sites** (other psychoactives produce their effects by other mechanisms which affect synaptic transmission, some of which are discussed below and in Chapter 6 of this textbook).

As mentioned above, the cell membrane of a neuron has **channels or "doors" for ions** (electrically charged atoms) which can pass through the membrane when specific channels are opened for specific ions. Normally the channels are closed until acted upon by the attachment of neurotransmitter molecules to receptors sites on the membrane of the receiving neuron (the post-synaptic neuron). The **attachment of the molecules of transmitter to the receptors sites**, like a key into a lock, **opens the "doors" (ion channels) to specific ions** which pass through the cell membrane carrying with them their electrical charge. This results in a change in the electrical state of the neuron, a post-synaptic potential (i.e. a voltage shift in the post-synaptic neuron)--if a positive shift in voltage occurs, it is called an **EPSP**, an excitatory post-synaptic potential; if a negative shift in voltage occurs in the post-synaptic neuron, then it is called an **IPSP**, an inhibitory post-synaptic potential. **EPSPs are caused by excitatory neurotransmitters** (e.g. glutamate; acetylcholine; histamine; epinephrine), **IPSPs by inhibitory transmitters** (e.g. GABA, gamma amino butyric acid; serotonin, also known by its chemical name, 5-hydroxytryptamine, abbreviated 5-HT).

These **PSPs**, whether excitatory or inhibitory, are called **graded potentials** because they are **not of a fixed voltage, but instead vary in voltage** depending on the amount of neurotransmitter (and other factors) that has been released onto the receptor sites on the postsynaptic neuron's dendrite. This is in contrast to the **action potential** (the nerve impulse) which is of a **fixed voltage and is "all or none"** which means that if it occurs, it occurs at its full strength or not at all. Think of a gun firing. If you pull the trigger of the gun with sufficient force (a kind of "trigger threshold" of force), then the gun fires a bullet with its full strength. But if trigger threshold is not reached, the gun doesn't fire at all. You don't get half a shot or one-eighth of a shot if you pull with less force, you can only get an "all or none" result--the gun fires, or it doesn't. Either the force on the trigger reaches threshold and the gun fires at its full "strength," or if the force is insufficient, the gun does not fire a bullet. The same is true of the action potential; if an action potential occurs, it occurs "all or none," at its full strength (voltage), or not at all, and, like the gun, the neuron "fires" only if its trigger threshold or **firing threshold** (in voltage, **about -55 millivolts for most neurons**) is reached. Graded potentials (EPSPs and IPSPs) are akin to **analog** signals (of different voltages), whereas action potentials aresimilar to **digital** signals (fixed voltage).

5.1.4 Types of Neurons

Not all neurons are created equal! There are neurons that receive information about the world around us, ***sensory* neurons**. There are ***motor* neurons** that allow us to initiate movement and behavior, ultimately allowing us to interact with the world around us. Finally, there are ***interneurons***, which process the sensory input from our environment into meaningful representations, plan the appropriate behavioral response, and connect to the motor neurons to execute these behavioral plans.

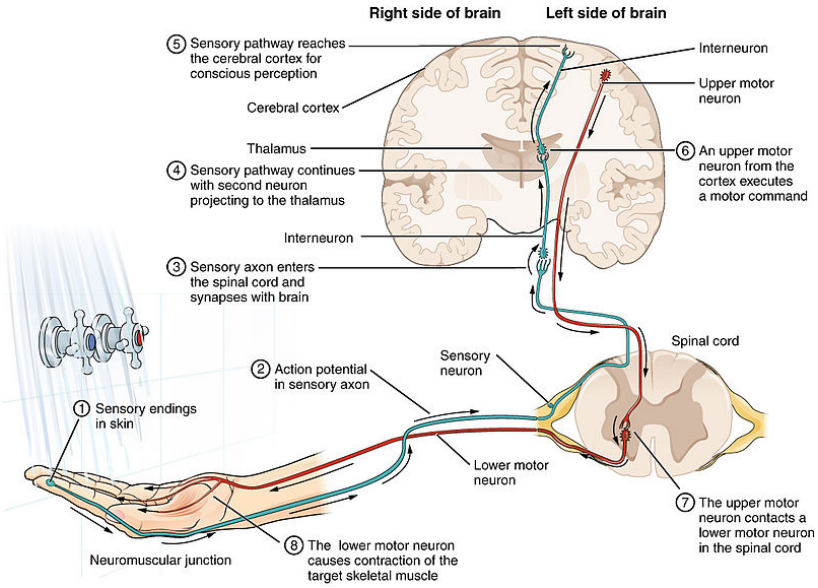


Figure 5.1.10: Sensory neurons carry information towards the CNS. Motor neurons carry information from the CNS. Interneurons carry information between sensory and motor neurons.

In addition to this general functional categorization of neurons (sensory, motor, and interneurons), neurons may also be classified by structure (e.g. unipolar, bipolar, multipolar), shape, or other characteristics. See Chapter 4.1 for an overview of these classifications. Since the shape of the neuron relates to its role in communication, this categorization is explored in greater depth below.

5.1.4.1 Categorizing Neurons by Shape

Neurons can also be classified by their shape. For example, Cajal used such names as basket, stellate, moss, pyramidal, etc. to describe various types of neurons he observed. However, this classification is anatomical and does not reflect whether the cell is, for example, a motor neuron or an interneuron. Nevertheless, the shape of neurons may reflect aspects of their function and their role in information processing in the nervous system. For example, **pyramidal neurons** (Figure 5.1.11) are "a common class of neuron found in the cerebral cortex of virtually every mammal, as well as in birds, fish and reptiles. Pyramidal neurons are also common in subcortical structures such as the hippocampus and the amygdala. They are named for their shape: typically they have a soma (cell body) that is shaped like a teardrop or rounded pyramid. They also tend to have a conical spray of longer dendrites that emerge from the pointy end of the soma (apical dendrites) and a cluster of shorter dendrites that emerge from the rounded end. . . They comprise about two-thirds of all neurons in the mammalian cerebral cortex . . . they are ‘projection neurons’ — they often send their axons for long distances, sometimes out of the brain altogether. For example, pyramidal neurons in layer 5 of the motor

cortex send their axons down the spinal cord to drive muscles" (Bekkers, 2011). They are the primary excitatory neurons in the motor cortex and the pre-frontal cortex. Table **5.1.1** summarizes features of pyramidal neurons.

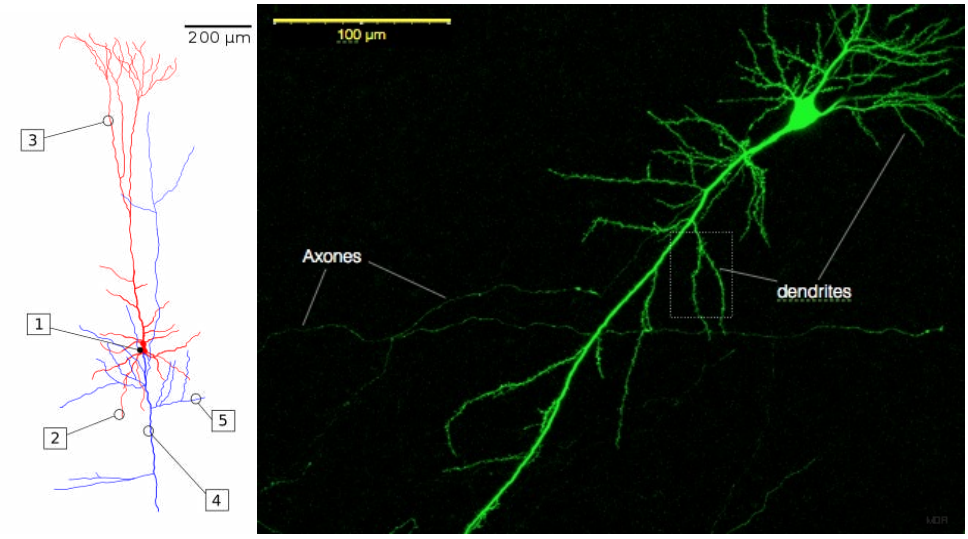


Figure **5.1.11**: (left) A reconstruction of a pyramidal cell. Soma and dendrites are labeled in red, axon arbor in blue. (1) Soma, (2) Basal dendrite, (3) Apical dendrite, (4) Axon, (5) Collateral axon. (right) Pyramidal neuron of a rat hippocampal organic culture. Axons shown (axones, in Spanish) are from another neuron most of which is not shown (Image on left, Wikipedia, The Pyramidal Cell, retrieved 8/30/21. Image on right from Wikimedia Commons, File:Neurone pyramidal.jpg; https://commons.wikimedia.org/wiki/F..._pyramidal.jpg; by Mathias De Roo; licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license.).

Table **5.1.1**. Features of pyramidal neurons. (adapted from John Bekkars; Pyramindal Neurons; Current Biology).

Location	Common in cerebral cortex of mammals, especially Layers III and V (cerebral cortex in mammals has six cell layers); also found in hippocampus, and amygdala. Also found in other vertebrates
Shape and numbers	Multipolar Pyramidal, but some variations among species and location in brain; about 2/3 of all neurons in the mammalian cerebral cortex
Function	Excitatory projection neurons with long axons carrying action potentials long distances, sometimes completely out of the brain; e.g. pyramidal neurons in layer 5 of motor cortex go all the way to spinal motor neurons; others that remain in the cortex interconnecting distant areas of cortex are critically important in many cognitive functions
Neurotransmitter	Glutamate, the most abundant excitatory transmitter in the vertebrate nervous system

By contrast, **stellate neurons** have a star-like shape by virtue of the dendrites which radiate out from the stellate neuron's cell body (see Figure 5.1.12(b)). Spiny stellate neurons, like pyramidal neurons, have large numbers of dendritic spines, but they lack the long apical dendrite characteristic of pyramidal neurons. Like pyramidal neurons, spiny stellate neurons in the cerebral cortex are excitatory and use glutamate as transmitter. Van Essen and Kelly (1973) reported different functions for visual cortical neurons depending upon whether they were pyramidal or stellate neurons. After pyramidal neurons, stellate neurons are the second most numerous type of cortical neuron. Another type of stellate neuron that has only sparse spines is inhibitory and uses GABA (gamma-amino-butyric-acid) as its transmitter.

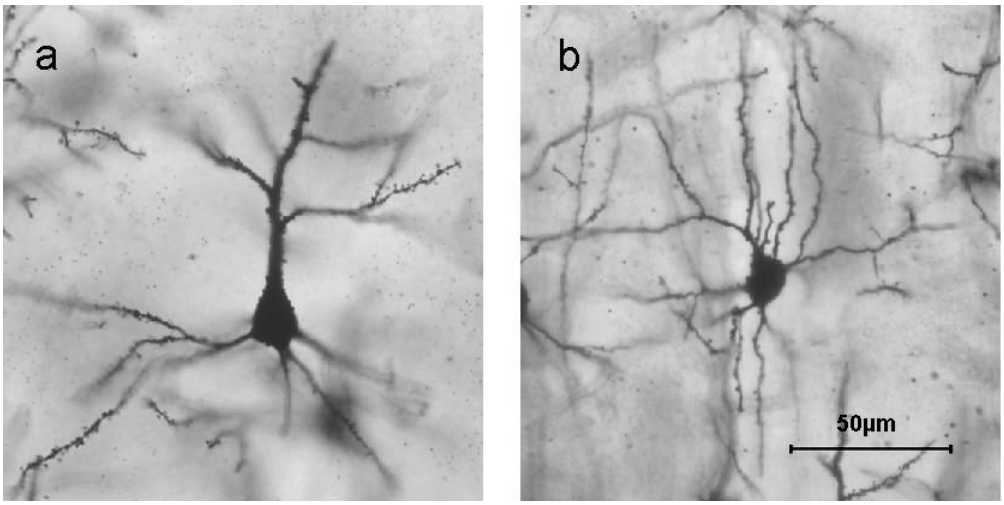


Figure 5.1.12: Golgi stained neurons in different layers of cerebral cortex: a) Layer II/III pyramidal cell; b) layer IV spiny stellate cell. Dendritic spines are visible on branching dendrites as clusters of tiny thorn-like bristles. (Image and one sentence caption from Wikipedia, Stellate Cell, retrieved 8/30/21; description of dendritic spines by Kenneth A. Koenigshofer, Ph.D.).

Brown et al. (2019) found different processing roles for basket neurons (see Figure 5.1.13) and stellate neurons in the cerebellum. Stellate cells and basket cells both make inhibitory synapses directly onto Purkinje neurons in the cerebellum. Stellate neurons influence pattern of firing and basket neurons affect rate of firing of the Purkinje neurons in the cerebellum. Basket cells are also found in the neocortex (cerebral cortex). Basket cells are inhibitory interneurons. They envelop pyramidal cell bodies in the neocortex in dense complexes resembling baskets (Kirkcaldie, 2012). Figure 5.1.13 shows a variety of interneurons classified by shape.

"Approximately 95% of the cortical neuronal activity is mediated by fast excitatory (glutamate, 80%) and fast inhibitory (GABA, 15%) neurons. The remaining 5% percent is associated with the slow modulatory action of monoaminergic (dopamine, serotonin, noradrenaline) and non-monoaminergic (acetylcholine, endorphins, etc.) neurons located in small subcortical nuclei of the mesencephalon and projecting to the cerebral cortex" (Marco Catani, in Encyclopedia of Behavioral Neuroscience, 2nd edition, 2022, Science Direct; https://www.sciencedirect.com/topics...pyramidal-cell; retrieved 4/25/22).

Though several basic types of neurons have been classified, the picture in the brain with regard to types of neurons present is quite complex, as expressed in this quote: "Whereas in the spinal cord we could easily distinguish neurons based on their function [sensory, interneuron, motor], that isn't the case in the brain. Certainly, there are brain neurons involved in sensory processing – like those in visual or auditory cortex – and others involved in motor processing – like those in the cerebellum or motor cortex. However, within any of these sensory or motor regions, there are tens or even hundreds of different types of neurons. In fact, researchers are still trying to devise a way to neatly classify the huge variety of neurons that exist in the brain . . . part of what gives the brain its complexity is the huge number of specialized neuron types. Researchers are still trying to agree on what these are" (University of Queensland, n.d.).

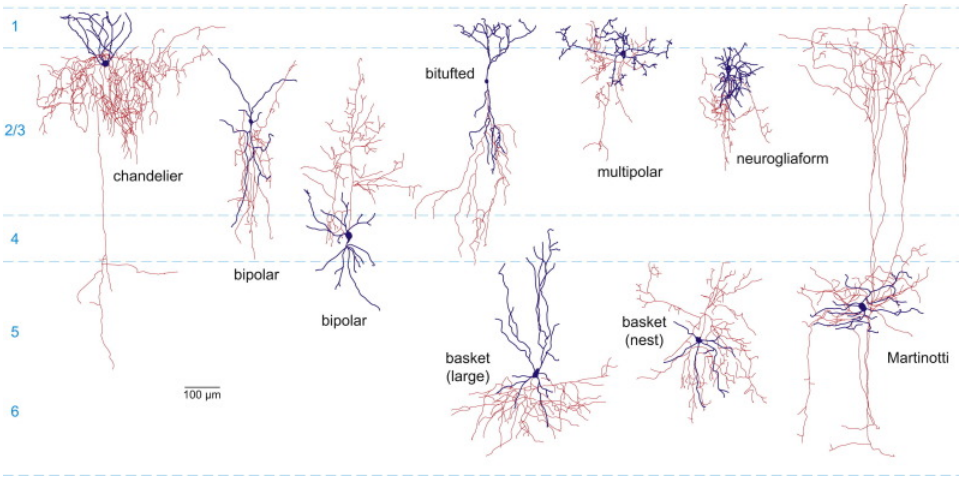


Figure **5.1.13**: (above). Variety of types of interneurons classified by shape. Representative morphologies of mouse interneuron types. Note that the axon (red) and dendrite (blue) arbors are typically less elaborate than in primate cortex. Two variants of bipolar cells and basket cells (large and nest) are shown (Kirkcaldie, 2012).

Yet another way to classify neurons is by the neurotransmitter they use (see Synaptic Transmission below). Perhaps equally important is to attempt to trace the connections that neurons make with other neurons to try to understand their inputs and outputs and what roles they play in processing. This has been done for many stellate and basket cells, inhibitory interneurons synapsing on Purkinje cells in the cerebellum. However, because of the complexity of brain circuitry this is very difficult to accomplish, especially in the cerebral cortex, although possible in cases where neurons belong to large neural tracts or pathways with straightforward connections which are easier to trace, such as from the pyramidal neurons in motor cortex to spinal cord motor neurons or from thalamus to sensory cortex. Understanding the complex circuitry of the brain in its entirety may never be fully accomplished. With 80-100 billion neurons in the human brain and most making perhaps many thousands of connections, the number of connections in the human brain is astronomical!

References

Bekkers, J. M. (2011). Pyramidal neurons. *Current biology*, *21*(24), R975.

Brown, A. M., Arancillo, M., Lin, T., Catt, D. R., Zhou, J., Lackey, E. P., ... & Sillitoe, R. V. (2019). Molecular layer interneurons shape the spike activity of cerebellar Purkinje cells. *Scientific reports*, *9* (1), 1-19.

Churchland, P. M. (2013). *Matter and consciousness*. MIT press.

Fields, R. D., & Stevens-Graham, B. (2002). New insights into neuron-glia communication. *Science*, *298* (5593), 556-562.

Kirkcaldie, M. T. (2012). Neocortex. In *The mouse nervous system* (pp. 52-111). Academic Press.

Koenigshofer, K.A. (2011). *Mind Design: The Adaptive Organization of Human Nature, Minds, and Behavior*. Pearson Education, Boston.

Mishra A, Singh S, Shukla S. (2018). Physiological and Functional Basis of Dopamine Receptors and Their Role in Neurogenesis: Possible Implication for Parkinson's disease. *J Exp Neurosci. 12:*1179069518779829. [PMC free article] [PubMed]

Nadim, F., & Bucher, D. (2014). Neuromodulation of neurons and synapses. *Current opinion in neurobiology*, *29*, 48-56.

Pasternak, J. F., & Woolsey, T. A. (1975). On the "selectivity" of the Golgi-Cox method. *J Comp Neurol*, 160(3), 307-312. doi: 10.1002/cne.901600304

Perea, G., Sur, M., & Araque, A. (2014). Neuron-glia networks: integral gear of brain function. *Frontiers in cellular neuroscience*, *8*, 378.

Smit, G. J., & Colon, E. J. (1969). Quantitative analysis of the cerebral cortex. I. Aselectivity of the Golgi-Cox staining technique. *Brain Res*, 13(3), 485-510.

University of Queensland, Queensland Brain Institute, Types of Neurons, n.d. https://qbi.uq.edu.au/brain/brain-an.../types-neurons; Retrieved 8/31/21

Van Essen, D., & Kelly, J. (1973). Correlation of cell shape and function in the visual cortex of the cat. *Nature*, *241*(5389), 403-405.

5.1.5 Attributions

1. Chapter 5, Communication within the Nervous System, section 5.1., "Neurons and their Basic Functions," written by Kenneth A. Koenigshofer, Ph.D. (except material listed in attributions below), is licensed under CC BY 4.0

2. Figure 5.1.2, Vocabulary, Discussion Questions, Outside Resources, and "Nissl and Golgi Stains" adapted by Kenneth A. Koenigshofer from: Furtak, S. (2021). Neurons. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology.* Champaign, IL: DEF publishers. Retrieved from http://noba.to/s678why4

5.1.6 Creative Commons License

Neurons by Sharon Furtak is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Permissions beyond the scope of this license may be available in our Licensing Agreement.

3. Figures 5.1.4., 5.1.5, and some text adapted from: General Biology (Boundless), Chapter 35, The Nervous System;

https://bio.libretexts.org/Bookshelv...gy_(Boundless); LibreTexts content is licensed by CC BY-NC-SA 3.0. Legal.

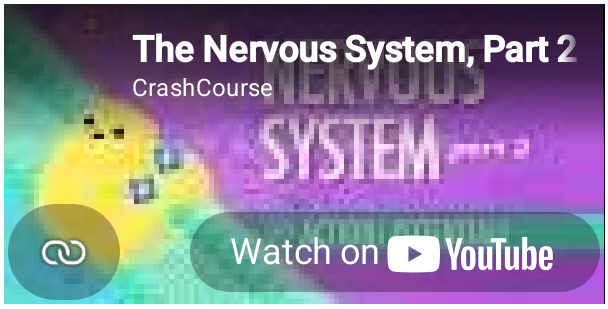
4. Figures 5.1.9 and 5.1.10 adapted from Chapter 11.3 (Neurons and Glia Cells), 11.4 (Nerve Impulses) in Book: Human Biology (Wakim & Grewal) - Biology LibreTexts by Suzanne Wakim & Mandeep Grewal, under license CC BY-NC

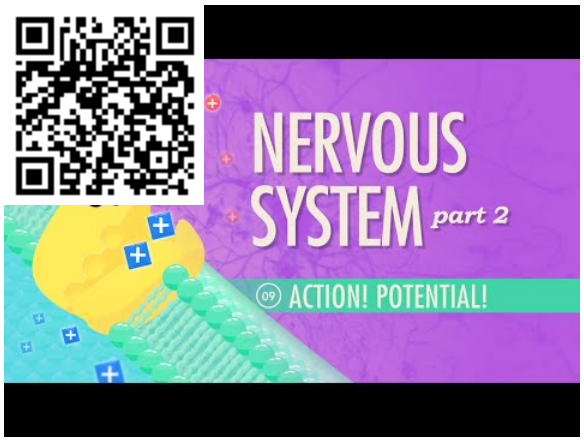
5.1.6.1 **Outside Resources (click on each to reveal the link)**

Video Series: Neurobiology/Biopsychology - Tutorial animations of action potentials, resting membrane potentials, and synaptic transmission.

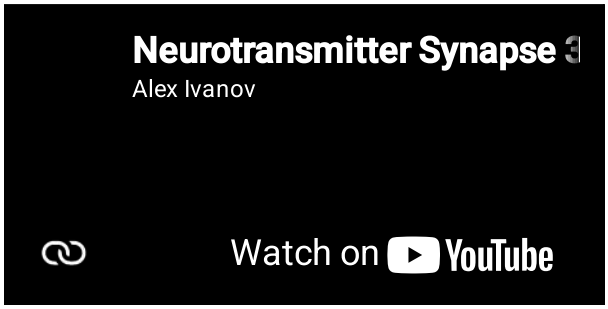
http://www.sumanasinc.com/webcontent/animations/neurobiology.html

Video: An animation and an explanation of an action potential





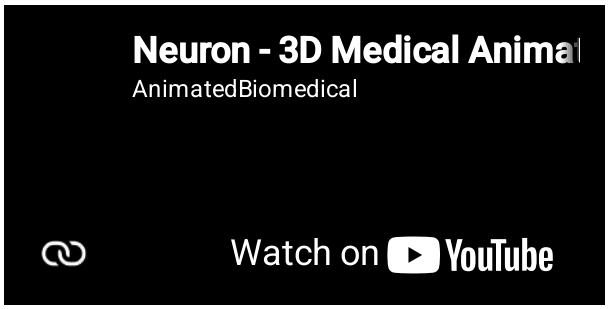
Video: An animation of neurotransmitter actions at the synapse



Vid animation that allows students to observe the results of manipulations to excitatory and inhibitory pos ıls. Also includes animations and explanations of transmission and neural circuits.

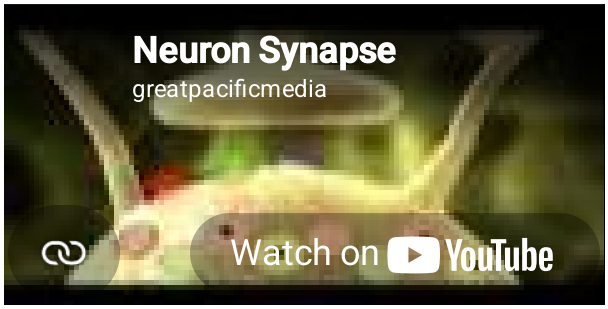
https://apps.childrenshospital.org/clinical/animation/neuron/

Video: Another animation of an action potential



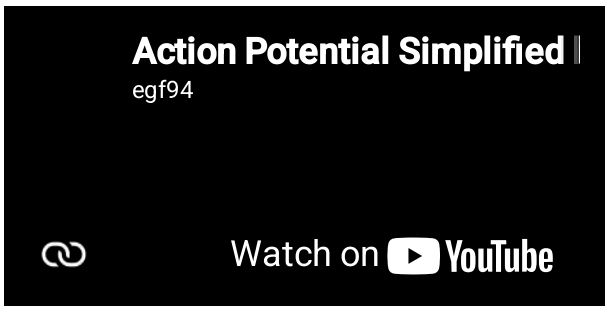


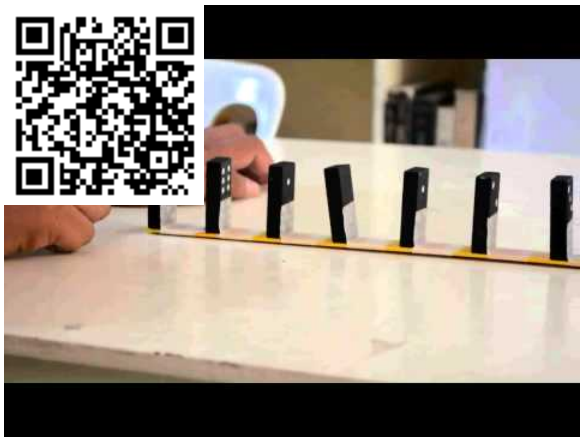
Video: Another animation of neurotransmitter actions at the synapse



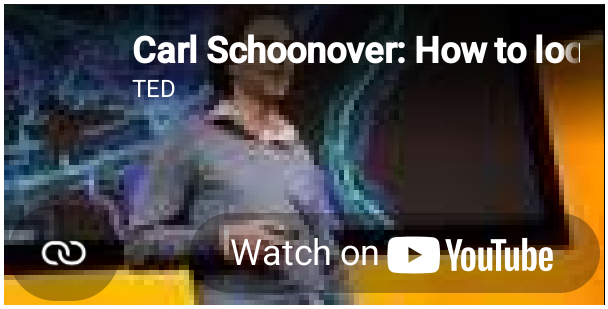


Video: Domino Action Potential: This hands-on activity helps students grasp the complex process of the action potential, as well as become familiar with the characteristics of transmission (e.g., all-or-none response, refractory period).





Video: For perspective on techniques in neuroscience to look inside the brain

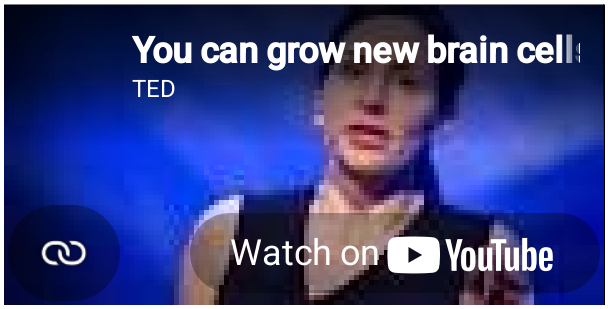




Video: The Behaving Brain is the third program in the DISCOVERING PSYCHOLOGY series. This program looks at the structure and composition of the human brain: how neurons function, how information is collected and transmitted, and how chemical reactions relate to thought and behavior.

http://www.learner.org/series/discoveringpsychology/03/e03expand.html

Video: You can grow new brain cells. Here\\'s how. -Can we, as adults, grow new neurons? Neuroscientist Sandrine Thuret says that we can, and she offers research and practical advice on how we can help our brains better perform neurogenesis—improving mood, increasing memory formation and preventing the decline associated with aging along the way.





Web: For more information on the Nobel Prize shared by Ramón y Cajal and Golgi

http://www.nobelprize.org/nobel_prizes/medicine/laureates/1906/

This page titled 5.1: Neurons and their Basic Functions is shared under a mixed license and was authored, remixed, and/or curated by Keneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

5.2: Neurons Generate Voltage Changes to Code Information

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Overview

According to materialists (those who believe that everything in the universe is physical), all of the mental activities of our minds and all of our behaviors are products of the physical activities of the brain and nervous system. These information processing operations that create our minds and control our behavior depend upon neurons and their electrical and chemical interactions. Neurons produce electrical potentials (voltages) that act as signals in the information processing activities of the brain. As previously mentioned, neurons communicate with one another across synaptic spaces using chemicals known as neurotransmitters. In this module, we examine how neurons create electrical potentials including the graded potentials (post-synaptic potentials) and the nerve impulse (the action potential), and the processes of synaptic transmission. The graded potentials can vary in voltage, while the action potential is fixed in voltage for any particular neuron and is said to be all or none--if it occurs, it occurs at its full strength or not at all. This is analagous to the firing of a gun. It has a trigger threshold equal to a particular amount of pressure applied to its trigger and then that threshold pressure is reached, the gun fires, and it fires at its full "strength." In a like manner, the action potential will only be generated if its trigger threshold, a particular voltage within the neuron, is reached. When this happens, the action potential is generated in the neuron at its full "strength," measured in voltage, and this nerve impulse is conducted along the axon to the axon endings which then release neurotransmitter onto receptor sites located on the receiving neuron. We now examine these processes in greater detail in this module and the next.

How Do Neurons Produce Electrical Potentials?

As described in Module 5.1, neurons, regardless of type, use voltage changes, known as electrical potentials, to code and process information. But how do they do it? How do neurons make voltage which they then use as electrical signals in the brain's electrochemical code? More specifically, the question is, how do neurons produce electrical potentials such as the resting potential, the action potential, and the post-synaptic potentials?

Neurons produce electrical potentials or voltages by the unequal distribution and the movement of electrically charged atoms called ions across the neuron's cell membrane. These ions come mainly from dissolved salts in the body fluids inside and outside neurons. The main ions used by neurons to produce their voltages are sodium (Na+), potassium (K+), and chloride (Cl-). Notice that the first two are positive ions and the last, chloride ions, are negatively charged. A fourth ion, organic anions (A-), which are large (on the molecular scale) negatively charged proteins, are manufactured inside the neuron. They are too large to cross the cell membrane, and therefore give the neuron's resting voltage a negative bias. The distribution or concentrations of these ions inside and outside a neuron determine its voltage (voltage is just the physical separation of charged particles, like in a car battery with its positive and negative poles around which are concentrated positively and negatively charged particles floating in battery acid).

By these means, four main types of neuron voltages or neuron potentials are produced by neurons: 1) the resting potential (often equal to about negative 70 thousandths of a volt, -70 millivolts) 2) two types of post-synaptic potentials (PSPs)--excitatory post-synaptic potentials (EPSPs) and inhibitory post-synaptic potentials (IPSPs) and 3) the action potential (AP, the nerve impulse). Let's examine these potentials in more detail and see how they are generated inside neurons. These electrical potentials in large populations of interacting neurons encode the information that guides our behavior and produces our psychological experience of the world.

The Neuron Cell Membrane

The **cell membrane**, which is composed of a lipid bilayer of fat molecules, separates the fluid inside of the cell from the surrounding extracellular fluid. There are proteins that span the membrane, forming **ion channels** that allow, when open, particular ions to pass between the intracellular and extracellular fluid (see Figure 8). These ions are in different concentrations inside the cell relative to outside the cell, and the ions have different electrical charges. Due to this difference in ion concentration and charge, in part enforced by the physical barrier of the cell membrane when ion channels to specific ions are closed, a voltage is produced, the resting potential.

The Resting Membrane Potential

The resting potential can be thought of as a baseline voltage from which the other neuron potentials are generated. It is the voltage inside a nerve cell when it is at rest, that is, it is neither receiving inputs (PSPs) at the moment nor generating any outputs (action potentials) at the moment. In this state of "rest", the voltage inside the nerve cell is approximately -70 mv. It is negative inside the neuron at "rest" because there are more negative ions inside the neuron (the intracellular fluid) and more positive ions on the outside (the extracellular fluid).

Specifically, there are large numbers of Na+ ions (sodium ions) outside the neuron and very few of these on the inside, when the neuron is "at rest" (when it is at resting potential). And there are more negatively charged ions inside the cell than on the outside during resting potential. This unequal distribution of ions across the cell membrane sets up the electrical "resting potential," making it equal to approximately -70 mv in the typical neuron. During resting potential (resting membrane potential) ion channels to Na+ are closed.

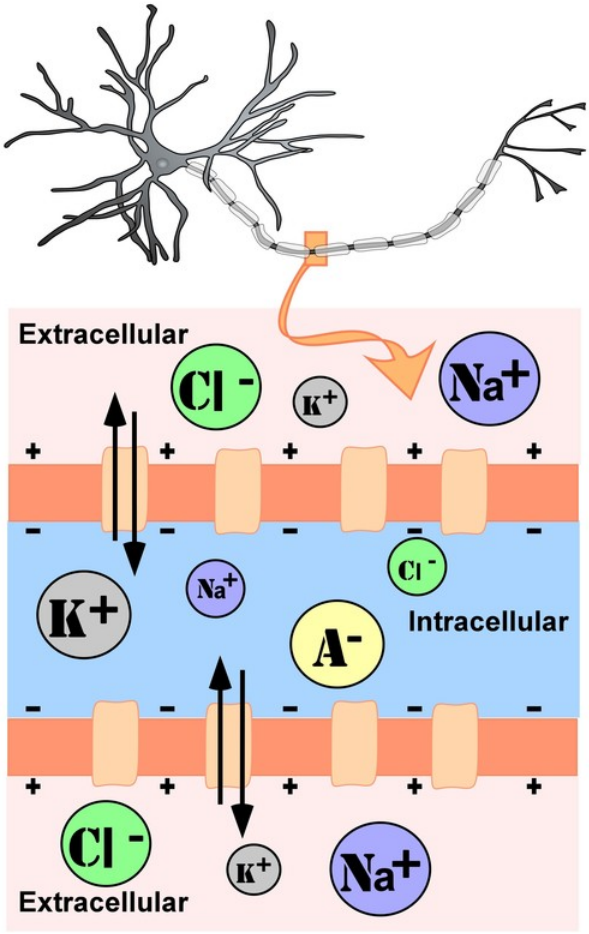


Figure 5.2.1: Representation of ion concentrations inside (intracellular) and outside (extracellular) a neuron in the unmyelinated segment of the axon. Size of the circles represents relative concentrations of each ion inside and outside the neuron; note that when the neuron is "at rest" it has a net negative charge inside, the resting potential, equal to about -70 millivolts in most neurons. Also note the higher concentrations of sodium (Na+) and chloride (Cl-) ions outside the cell. Sodium on the outside of the neuron "would like" to move into the cell along its charge and concentration gradients, but it can't when the cell is at rest, because the ion channels for sodium ions are closed, creating a physical barrier keeping most of the extracellular sodium ions from crossing through the cell membrane. But what happens when the sodium channels open? And, test yourself, what opens the ion channels?

5.2.0.1 The Sodium-Potassium Pump

During the resting potential, the sodium-potassium pump maintains a difference in charge across the cell membrane of the neuron. The sodium-potassium pump is a mechanism of active transport that moves sodium ions out of cells and potassium ions into cells.

The sodium-potassium pump moves both ions from areas of lower to higher concentration, using energy in ATP and carrier proteins in the cell membrane (see Figure 5.2.2).

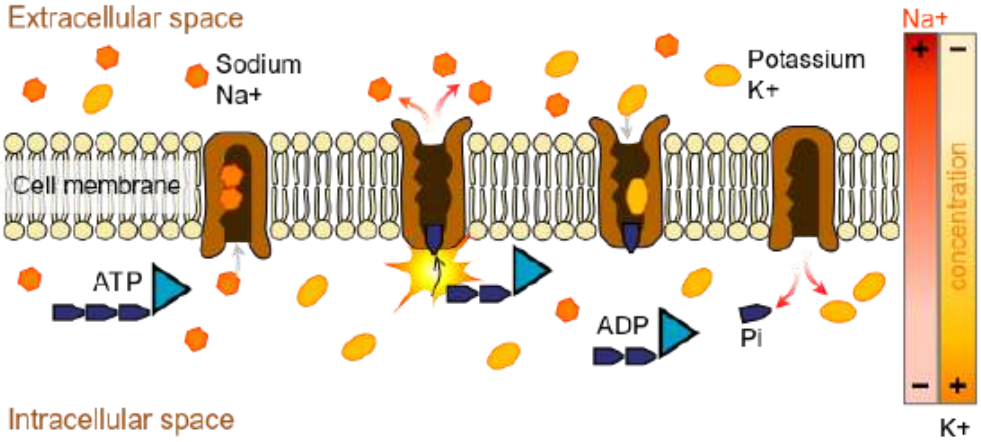


Figure 5.2.2: The sodium-potassium pump helps maintain the resting potential of a neuron. During resting potential, there is more negative charge inside than outside the cell creating a resting potential of -70mv. During resting potential, some Na+ leaks into the neuron and some K+ leaks out. ATP (Adenosine triphosphate) provides energy to pump sodium out and potassium into the cell. There is more concentration of sodium outside the membrane and more concentration of potassium inside the cell due in part to the unequal movement of these ions by the pump. The presence of negatively charged organic anions (A-) contributes to the net negative charge (-70mv) inside the neuron at rest. (Image from Wikimedia Commons; File:Scheme sodium-potassium pump-en.svg; https://commons.wikimedia.org/wiki/F...um_pump-en.svg; by LadyofHats Mariana Ruiz Villarreal; released into the **public domain** by its author, **LadyofHats**. This applies worldwide).

Producing Other Membrane Potentials: Post-synaptic potentials and the Action potential

How are the other neuron potentials produced? To generate the other neuron potentials, which are just voltage shifts away from resting potential, there must occur a redistribution of ions (and their electrical charges) across the cell membrane. In short, ions must move.

There are two main forces (called gradients) that can cause these ions to move.

First, opposite charges attract one another ("opposites attract"), and like charges repel one another. When ions of opposite charge are unequally distributed across the cell membrane (as is the case during the resting potential), this sets up what is called a **charge gradient** (an unequal distribution of charged particles) that creates an **electrostatic force.** If these ions are allowed to move freely, they will move along the charge gradient. Positive charges will move toward negative ones and vice versa.

Secondly, when ions of any particular type (for example, sodium ions) are unequally distributed across the cell membrane (like more Na+ outside the neuron than inside during resting potential), this sets up what is called a **concentration gradient** which can cause **diffusion,** the net movement of ions from a region of higher concentration to a region of lower concentration. Diffusion is driven by a gradient in concentration. Such concentration gradients for several ions (sodium, potassium, and chloride ions) exist when the neuron is at resting potential (see Figure 10). Ions, if allowed to move freely, will move along their concentration gradients, such that ions of a particular type (like Na+ ions) will move from a region of high concentration (of Na+) to a region of lower concentration (of Na+).

In short, ions, if allowed to move freely, will move along the charge gradient and along their concentration gradients. Ions "want" to move to equalize their concentrations across the cell membrane and also "want" to move to equalize the charges across the cell membrane (by moving toward opposite charges and away from like charges).

However, it is important to note that **these two forces created by the charge gradient and concentration gradient for each ion can oppose one another** in the case of some ions, **or can act in concert** (as is the case with Na+ ions during the resting potential).

Let us see how these two forces, diffusion (due to concentration gradients) and electrostatic pressure (due to a charge gradient), act on the four groups of ions mentioned above.

1. ***Anions (A-)****:* Anions are highly concentrated inside the cell and contribute to the negative charge of the resting membrane potential. Diffusion and electrostatic pressure are not forces that determine A- concentration because A- is impermeable to the cell membrane. There are no ion channels that allow for A- to move between the intracellular and extracellular fluid.
2. ***Potassium (K+)****:* The cell membrane is very permeable to potassium at rest, but potassium remains in high concentrations inside the cell. Diffusion created by the concentration gradient pushes K+ to the outside of the cell because it is in high concentration inside the cell. However, electrostatic pressure created by the charge gradient pushes K+ into the cell because the positive charge of K+ is attracted to the negative charge inside the cell. In combination, these forces oppose one another with respect to K+ *with the charge gradient (opposite charges attract) overpowering the concentration gradient so that the net effect is to push and keep K+ inside the neuron in higher concentrations*.
3. ***Chloride (Cl-)****:* The cell membrane is also very permeable to chloride at rest, but chloride remains in high concentration outside the cell. Diffusion created by a concentration gradient for Cl- pushes Cl- toward the inside of the cell because it is in high concentration outside the cell. However, electrostatic pressure created by the charge gradient for Cl- pushes Cl- toward the outside of the cell because the negative charge of Cl- is attracted to the positive charge outside the cell created primarily by the high concentration of Na+ there. Similar to K+, these forces oppose one another with respect to Cl- *and again, in this case, like Na+, the more powerful charge gradient (like charges repel) overcomes the weaker concentration gradient for Cl-.*
4. ***Sodium (Na+)****:* The cell membrane is not very permeable to sodium when the neuron is at rest. Diffusion created by a concentration gradient pushes Na+ toward the inside of the cell because it is in high concentration outside the cell. Electrostatic pressure created by the charge gradient for Na+ also pushes Na+ toward the inside of the cell because the positive charge of Na+ is attracted to the negative charge inside the cell. Both of these forces push Na+ inside the cell; however, as discussed above, Na+ cannot permeate the cell membrane because the channels for Na+ are closed and so Na+ remains in high concentration outside the cell. The small amounts of Na+ inside the cell are removed by the **sodium-potassium pump** (Figure 5.2.2), which uses the neuron's energy (adenosine triphosphate, ATP) to pump three Na+ ions out of the cell in exchange for bringing two K+ ions inside the cell.

Though ions "want" to move in these ways dictated by their charge gradients and concentration gradients, Na+ can't move freely when the neuron is at resting membrane potential because Na+ ion channels are closed and the cell membrane acts as a physical barrier to Na+ ions, preserving the unequal distributions of Na+ ions inside and outside the neuron. To get ion movement, pores or ion channels in the cell membrane of the neuron "at rest" must be opened, overcoming the physical barrier created by the cell membrane when the neuron is "at rest" (not receiving any inputs via its dendrites and not generating any outputs, action potentials, via its axon) and its sodium channels are closed.

What is it that causes the ion channels to open, allowing the ions to move along their gradients? You can test yourself here to see how much you have retained from your reading thus far. What is it that opens the ion channels?

The arrival of transmitter molecules from the axon ending of a pre-synaptic neuron, and their attachment to post-synaptic receptor sites, is the key event. This "lock and key" interaction between molecules of transmitter and the post-synaptic receptor sites causes "doors", specific ion channels, to open.

When specific ion channels get opened in this way (Na+ channels, for example), those specific ions (Na+, in this case) move through the cell membrane along their specific concentration and charge gradients. This will cause a voltage shift away from resting potential. This voltage shift is the post-synaptic potential (PSP), in this case it is an excitatory post-synaptic potential (EPSP), a positive shift in voltage away from resting potential because when Na+ channels in the cell membrane are opened by neurotransmitter, then Na+ ions will move along their charge gradient and concentration gradient into the cell, making the inside of the neuron more positively charged.

Excitatory and Inhibitory Post-Synaptic Potentials

1. **Excitatory postsynaptic potentials** (EPSPs): a *depolarizing* current that causes the membrane potential to become more positive and closer to the threshold of excitation (the "trigger threshold" of minus 55 millivolts) which lead to generation of an action potential (see Figure 5.2.4); caused by excitatory transmitter.
2. **Inhibitory postsynaptic potentials** (IPSPs): a *hyperpolarizing* current that causes the membrane potential to become more negative (see Figure 5.2.4) and further away from the threshold of excitation (the "trigger threshold" of minus 55 millivolts), caused by inhibitory transmitter.

The post-synaptic potentials occur in a post-synaptic neuron (thus the name), a receiver neuron. When such a neuron receives an input from a pre-synaptic neuron, in the form of transmitter molecules which attach to the **post-synaptic receptor sites (which are proteins that are embedded in the membrane of the post-synaptic cell)**, ion channels then open, ions move across the cell membrane, carrying with them their electrical charges. These events change the voltage inside the post-synaptic neuron. It is this voltage shift away from resting potential that constitutes the post-synaptic potential (PSP).

An **EPSP (Excitatory Post-synaptic Potential)** is a positive shift in voltage (depolarization) from the resting potential, say from -70 mv to -60 mv; as described above, an EPSP occurs when Na+ channels open and Na+ ions flow into the neuron along their concentration and charge gradients, carrying with them their positive charges, making the interior of the post-synaptic neuron more positively charged. This positive shift in voltage is also known as *depolarization* of the neuron and is caused by excitatory transmitter molecules binding to post-synaptic receptor sites of the receiving neuron.

An **IPSP (Inhibitory Post-synaptic Potential)** is a negative shift in voltage (hyperpolarization) from the resting potential, say from -70 mv to -80 mv; an IPSP occurs when Cl- and K+ channels open. These ions then flow along charge and concentration gradients, which cause negatively charged Cl- ions to move in and positively charged K+ ions to move out of the post-synaptic neuron, making it more negatively charged. This increase in net negative charge inside the cell is the IPSP (inhibitory post-synaptic potential). Cl- ions move in against their charge gradient (the negative voltage of the resting potential repels Cl- ions) because the concentration gradient pushing them inward into the neuron is stronger than the opposing charge gradient that tries to push them out of the neuron (like charges repel). Similarly, K+ ions move out in opposition to the charge gradient pulling them in, again, because the concentration gradient pulling them outward is stronger than the charge gradient which tries to pull them inward. These movements are prevented during the resting potential by the added physical barrier, the cell membrane, but when these ion channels are opened more than they are during the resting potential, the gradients move the ions causing the IPSP. As noted above, in the case of Na+ both charge and concentration gradients pull Na+ inward when Na+ ion channels in the neuron's cell membrane open.

Note, again, that in both the EPSP and the IPSP, it is the attachment of molecules of transmitter to post-synaptic receptor sites (like keys going into locks of a specific shape) that opens the "doors", the ion channels, allowing ions to move through the cell membrane.

But there is one more key issue here. What is it that determines *which* ion channels open, and therefore, whether an EPSP or an IPSP occurs? The answer was implied above. The answer is: it is the type of neurotransmitter and type of receptor site receiving the transmitter molecules. There are **two basic types of neurotransmitter, excitatory and inhibitory.**

Excitatory transmitters (such as glutamate, acetylcholine--ACh, norepinephrine--NE, dopamine--DA) are those which **open sodium (Na+) channels** in the post-synaptic membrane, allowing sodium ions, carrying their positive charge, into the cell, making an EPSP.

Inhibitory transmitters (such as gamma-amino-butyric acid--GABA, serotonin--5-HT, or dopamine--DA) are those which affect **chloride and potassium channels i**n the post-synaptic membrane, allowing chloride to follow its concentration gradient (overcoming the opposing charge gradient) into the post-synaptic neuron and allowing potassium to follow its concentration gradient (overcoming its opposing charge gradient) and moving out. These ion movements make the inside of the neuron more negative, making an IPSP.

Notice that DA is listed as both excitatory and inhibitory. That's because some post-synaptic DA receptors are inhibitory, leading to IPSPs when activated by DA, and other DA receptors are excitatory, producing EPSPs when they bind with DA.

5.2.0.1 Spatial and Temporal Summation

There is one additional factor in this process. Each neuron connects with numerous other neurons, often receiving multiple impulses from them. Sometimes, a single excitatory postsynaptic potential (EPSP) is strong enough to induce an action potential in the postsynaptic neuron, but often multiple presynaptic inputs must create EPSPs around the same time for the postsynaptic neuron to be sufficiently depolarized to fire an action potential. Summation, either spatial or temporal, is the addition of these impulses at the axon hillock. Together, synaptic summation and the threshold for excitation act as a filter so that random “noise” in the system is not transmitted as important information.

At any moment in time, each neuron can be receiving mixed messages, both EPSPs and IPSPs. Thus, receiver neurons (post-synaptic neurons) can receive multiple inputs--simultaneously or over time or space. The multiple inputs can add up--this is called summation and there are two types. **Spatial summation** refers to two or more PSPs arriving at different locations (i.e. different spaces on the receiving neuron, thus spatial summation) on the post-synaptic (receiver) neuron simultaneously or close enough in time so that their voltages add together. For example, an IPSP of negative 25 millivolts (thousandths of a volt) might add to an EPSP of 50 millivolts occurring at the same time = 50 millivolts positive - 25 millivolts = net positive 25 millivolts. That would be spatial summation. The other kind of summation is **temporal summation** when the PSPs from a single pre-synaptic source, arriving to the post-synaptic neuron in "rapid fire," add together. For example, three EPSPs of 10 millivolts each occurring in rapid sequence would add together. Spatial and temporal summation are important in determining whether "trigger threshold" will be reached in the receiving neuron, triggering an action potential in that neuron. Summation permits activity from many input neurons to be integrated in the neuron receiving the inputs. If membrane depolarization does not reach the threshold level, an action potential will not happen.

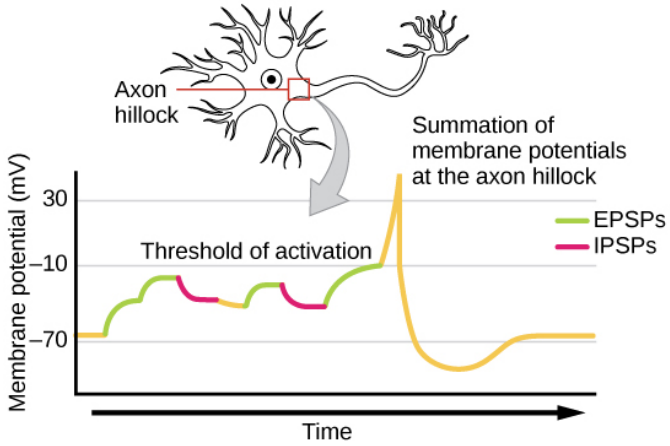


Figure **5.2.3**: **Signal summation at the axon hillock**: A single neuron can receive both excitatory and inhibitory inputs from multiple neurons. All these inputs are added together at the axon hillock. If the EPSPs are strong enough to overcome the IPSPs and reach the threshold of excitation, the neuron will fire an action potential. (Image from Biology, Signal Summation; Libretexts; https://bio.libretexts.org/Bookshelv...%20will%20fire. Unless otherwise noted, LibreTexts content is licensed by CC BY-NC-SA 3.0).

One neuron often has input from many presynaptic neurons, whether excitatory or inhibitory; therefore, inhibitory postsynaptic potentials (IPSPs) can cancel out EPSPs and vice versa. The net change in postsynaptic membrane voltage determines whether the postsynaptic cell has reached its threshold of excitation ("trigger threshold") needed to fire an action potential. If the neuron only receives excitatory impulses, it will also generate an action potential. However, if the neuron receives both inhibitory and excitatory inputs, the inhibition may cancel out the excitation and the nerve impulse will stop there. To review, spatial summation means that the effects of impulses received at different places on the neuron add up so that the neuron may fire when such impulses are received simultaneously, even if each input on its own would not be sufficient to cause firing. Temporal summation means that the effects of impulses received at the same place can add up if the impulses are received in close temporal succession. Thus, the neuron may fire when multiple inputs are received, even if each input on its own would not be sufficient to cause firing.

5.2.0.1 Key Points

- Simultaneous impulses may add together from different places on the neuron to reach the threshold of excitation during spatial summation.
- When individual impulses cannot reach the threshold of excitation on their own, they can can add up at the same location on the neuron over a short time; this is known as temporal summation.
- The action potential of a neuron is fired only when the net change of excitatory and inhibitory impulses is non-zero.

5.2.0.1 Key Terms

- **temporal summation**: the additive effect when successive impulses (and the resulting PSPs) received at the same place on the neuron add up

- **spatial summation**: the additive effect when simultaneous impulses (and the resulting PSPs) received at different places on the neuron add up
- **axon hillock**: the specialized part of the soma of a neuron at the root of the axon where impulses are added together

The Action Potential (the "nerve impulse")

The nerve impulse, or action potential, is generated in the post-synaptic neuron only if a **"trigger threshold" ("threshold of excitation") of -55 millivolts** (minus 55 mv) is reached (the trigger threshold varies from neuron to neuron and may be anywhere from minus 65 to minus 55, but will always be the same for any particular neuron. For purposes of our discussion, we will continue to refer to the trigger threshold of neurons as minus 55 millivolts, not minus 65 as is sometimes indicated in some textbooks). When that trigger voltage is attained (for example, as a result of an EPSP of at least 15 millivolts), then **all the voltage-gated sodium ion channels suddenly open**, allowing a massive inflow of sodium ions, Na+, into the cell along both concentration and electrical gradients for sodium. This produces a rapid, **large positive shift or "spike"** in the voltage of the post-synaptic neuron. This is the **nerve impulse or action potential**. In most neurons, it is a positive shift of about 100 to 130 millivolts, if we measure from the -70 millivolts of the resting potential, up to about a positive 30 to 60 millivolts, depending on the neuron. This value for any particular neuron is always the same for that neuron. Typically neurons with larger diameters (such as A fibers which are typically myelinated as well as large diameter) produce the largest action potentials (up to about positive 60mv, i.e. 130 mv above resting potential), while the neurons with smaller diameter axons (classified as C fibers, typically unmyelinated) produce action potentials in the lower range.

As fast as the voltage of the action potential rises, it starts to fall after reaching its peak (corresponding to peak Na+ concentration inside the neuron). It quickly falls back to the resting potential and even a bit below resting potential (the so-called refractory period), as Na+ and K+ ions move out, before the return of the neuron's potential back to -70 mv, the resting potential. This rapid rise to the action potential's peak and then its rapid fall gives the action potential, when graphed, a spike appearance. For this reason, action potentials are often called "spikes" by neuroscientists. See the diagram of the action potential (Figures 11, 12) to get a clearer picture of these events. Note that an **EPSP**, an excitatory post-synaptic potential, **moves the neuron's voltage closer to "trigger threshold"**, increasing the chances that the neuron will be sufficiently "excited" to generate an action potential. By contrast, the **IPSP**, the inhibitory post-synaptic potential, **moves the neuron's voltage further away from "trigger threshold"** inhibiting the neuron from firing an action potential.

Once **trigger threshold ("threshold of excitation")** is reached (dotted line in Figure 9) and an action potential is generated, it is then conducted down the length of this neuron's axon (saltatory conduction in a myelinated axon; see above). **Once the AP reaches this neuron's axon ending** or terminal button, its arrival causes the **release of neurotransmitter** molecules from the synaptic vesicles located there. Now this neuron is no longer called a post-synaptic neuron, but becomes a pre-synaptic neuron (a sender neuron) with respect to the next cell in line. It's the release of transmitter (step 3 in the list of 8 steps in synaptic transmission shown in Synaptic Transmission below) leads to steps (4 and 5 in the list of 8 steps) and an EPSP or IPSP is generated in the next cell in line. Remember that whether an EPSP or an IPSP is caused in the next cell in line is determined by whether an excitatory or an inhibitory transmitter has been released from the pre-synaptic neuron (see the 8 steps below).

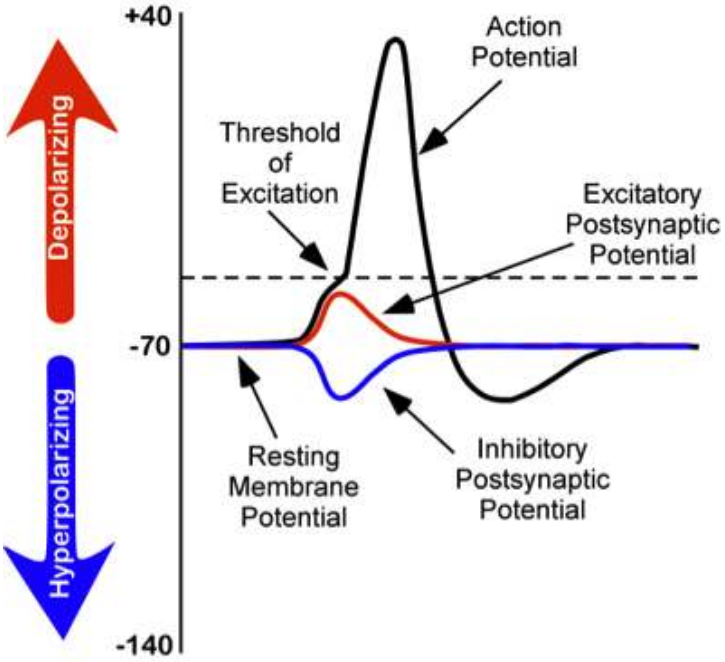


Figure 5.2.4: Changes in membrane potentials of neurons. (left) Dotted line represents trigger threshold ("threshold of excitation"), about -55 millivolts, an action potential is generated once trigger threshold voltage is reached.

Note that an EPSP (depolarization) moves the neuron's voltage more positive and thus closer to trigger threshold, making it more likely that the voltage reaches trigger threshold, "firing" an action potential down the axon of the neuron; thus it is excitatory. An IPSP does the opposite. It moves the voltage in the negative direction, further from trigger threshold, and thus inhibiting the neuron from producing an output (an action potential) in its axon. Also note that the peak of the action potential (top of the black line shaped like a spike) is the peak of its positive voltage and corresponds to the maximum concentration of Na+ ions inside the cell as a result of Na+ ion channels opening after trigger threshold has been reached. After this peak concentration of Na+ making the peak of the voltage of the action potential, positively charged ions (including K+ ions) begin to leave the interior of the cell and as they do so, the positive voltage inside the neuron progressively drops, corresponding to the downward slope of the spike. Note in addition that the spike goes further negative than the resting potential. In this state the neuron is inhibited by this **refractory period** and cannot fire another action potential for a brief time. This keeps the action potential as a discrete digital event. This is important for source coding, an important feature of neural coding, discussed in the chapter on learning and memory in this text. The entire process from the triggering of the action potential (which starts in the root of the axon, called the **axon hillock**, nearest the cell body) to the end of the refractory period takes about 1 millisecond, making the maximum rate at which a neuron can generate and "fire" action potentials (nerve "impulses") about 1,000 per second, although most neurons when active fire at a much lower frequency. Frequency of action potentials is one code that the nervous system uses to represent information. For example, the brighter a light source is, the higher the frequency of action potentials in the optic nerve in response to it.

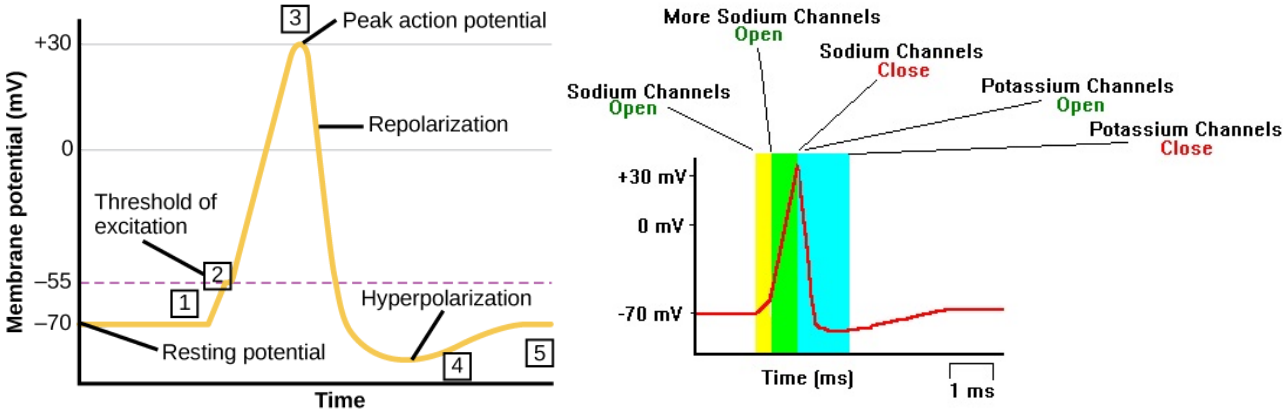


Figure 5.2.5: Formation of an action potential. (left) The formation of an action potential can be divided into five steps. (1) A stimulus from a sensory cell or another neuron causes the target cell to depolarize toward the threshold potential. (2) If the threshold of excitation is reached, all Na+ channels open and the membrane depolarizes. (3) At the peak action potential, K+ channels open and K+ begins to leave the cell. At the same time, Na+ channels close. (4) The membrane becomes hyperpolarized as K+ ions continue to leave the cell. The hyperpolarized membrane is in a refractory period and cannot fire. (5) The K+ channels close and the Na+/K+ transporter (requiring energy expenditure) restores the resting potential. (right) Sequence of opening and closing of Sodium (Na+) and Potassium (K+) ion channels producing the rising and falling phases of the action potential. (Image on left and caption from Lumen Boundless Biology; How Neurons Communicate; https://courses.lumenlearning.com/bo...s-communicate/. Unless otherwise noted, content is licensed under the Creative Commons Attribution 4.0 License. Image on right from Wikimedia Commons; File.مراحل ارسال سیگنال عصبی.jpg; https://commons.wikimedia.org/wiki/F...8%A8%DB%8C.jpg; by Vidakarimnia; licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license).

A single neuron by itself can't generate psychological states like feelings, perceptions, or thoughts. Neurons must interact with other neurons. The generation of a thought, or other complex psychological/mental experience, requires that enormous numbers of neurons interact with one another. To interact, they must communicate with other neurons.

As described in module 5.1, neurons communicate with one another, across spaces between them known as "synaptic gaps" by the release of chemicals (called neurotransmitters or simply transmitters) from one neuron's axon ending onto receptor sites on the dendrites or soma of the target neuron (a post-synaptic neuron). The neurotransmitters involved in this communication between neurons are manufactured in the soma of the neuron and are then transported down a long axon where they are stored in the synaptic vesicles until released from the axon ending into the synaptic space; the release is triggered by the arrival of an action potential (the nerve impulse) at the axon ending.

As discussed in module 5.1, the synapse includes the pre-synaptic neuron (the sender cell) and its axon ending with its synaptic vesicles, along with the synaptic gap, and the post-synaptic neuron (the receiving cell) with its receptor sites. There are an enormous number of neurons in the human brain, but the number of *possible* different combinations of synaptic connections among those 80-100 billion neurons is unimaginable--one neuroscientist estimated that the number of possible patterns of interconnect between neurons in a human brain exceeds the number of atoms in the entire universe! This neural complexity appears to be sufficient to code all the information contained in a human brain including all the thoughts, feelings, perceptions and memories of a human lifetime.

5.2.1 Key Points

- The resting potential, typically equal to -70 millivolts, is the voltage inside a neuron when it is receiving no inputs from other neurons and producing no outputs (no action potentials)
- Post-synaptic potentials (PSPs), EPSPs (excitatory; depolarizations) and IPSPs (inhibitory; hyperpolarizations) are positive and negative voltage shifts, respectively, away from the neuron's resting potential. These voltage shifts, PSPs, in the post-synaptic neuron result from release of transmitter from one or more pre-synaptic neurons.
- Action potentials are formed when inputs (summed EPSPs and IPSPs) cause the cell membrane to depolarize past the threshold of excitation ("trigger threshold"), causing all sodium ion channels to open, leading to a large positive shift in the neuron's

voltage (Figures 5.2.4 and 5.2.5).

- When the potassium ion channels are opened and sodium ion channels are closed, the cell membrane becomes hyperpolarized as potassium ions leave the cell; the cell cannot fire during this refractory period (Figures 5.2.4 and 5.2.5).
- The action potential travels down the axon as the membrane of the axon depolarizes and repolarizes (see 5.1, Figures 5.1.4 and 5.1.5) .
- Myelin insulates many axons to prevent leakage of the current as it "leaps" from node to node down the axon.
- Nodes of Ranvier are gaps in the myelin along the axons; they contain sodium and potassium ion channels, allowing the action potential to travel quickly down the axon by jumping from one node to the next (saltatory conduction; see module 5.1).

5.2.2 Key Terms

- **action potential**: a short term change in the electrical potential that travels along a cell
- **depolarization**: a decrease in the difference in voltage between the inside and outside of the neuron
- **hyperpolarize**: to increase the polarity of something, especially the polarity across a biological membrane
- **node of Ranvier**: a small constriction in the myelin sheath of axons
- **saltatory conduction**: the process of regenerating the action potential at each node of Ranvier

5.2.3 Attributions

1. Chapter 5, Communication within the Nervous System, 5.2. "Neurons Generate Voltage Changes to Code Information" by Kenneth A. Koenigshofer, PhD, Chaffey College, is licensed under CC BY 4.0

2. Figures 5.2.1, 5.2.4, Vocabulary, Discussion Questions, Outside Resources, and some text adapted from: Furtak, S. (2021). Neurons. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology.* Champaign, IL: DEF publishers. Retrieved from http://noba.to/s678why4; Neurons by Sharon Furtak at NOBA is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Figure 5.2.4 caption by Kenneth A. Koenigshofer, PhD, Chaffey College.

3. "Key Points" and "Key Terms" adapted from: General Biology (Boundless), Chapter 35, The Nervous System;

https://bio.libretexts.org/Bookshelv...gy_(Boundless); LibreTexts content is licensed by CC BY-NC-SA 3.0. Legal.

4. Figures 5.2.3, 5.2.5, and some text adapted from: General Biology (Boundless), Chapter 35, The Nervous System;

https://bio.libretexts.org/Bookshelv...gy_(Boundless); LibreTexts content is licensed by CC BY-NC-SA 3.0. Legal.

5. Figures 5.2.2 adapted from Chapter 11.3 (Neurons and Glia Cells), 11.4 (Nerve Impulses) in Book: Human Biology (Wakim & Grewal) - Biology LibreTexts by Suzanne Wakim & Mandeep Grewal, under license CC BY-NC

5.2.4 Creative Commons License

Neurons by Sharon Furtak is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Permissions beyond the scope of this license may be available in our Licensing Agreement.

This page titled 5.2: Neurons Generate Voltage Changes to Code Information is shared under a mixed license and was authored, remixed, and/or curated by Kenenth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

5.3: Synaptic Transmission

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

5.3.0.0.1 Learning Objectives

1. Describe ion channels, and what changes they undergo when neuron potentials are produced; what causes ion channels to change during synaptic transmission?
2. Define ionotropic and metabotropic receptors and discuss in what ways they differ from one another in their effects during synaptic transmission
3. Explain the steps in synaptic transmission from pre-synaptic neuron to post-synaptic neuron
4. Describe how excitatory and inhibitory transmitters differ in their effects on post-synaptic neurons during synaptic transmission
5. Discuss how layers of neurons, simulated by processing units in artificial neural networks, might produce psychological processes such as learning

5.3.1 Overview

After an action potential is generated in the presynaptic neuron, this all or none impulse is conducted along the axon to the axon ending (the terminal button). In the presynaptic terminal button, the arrival of the action potential triggers the release of neurotransmitters (see Figure 5.3.2). Neurotransmitters cross the synaptic gap and open subtypes of receptors in a lock-and-key fashion (see Figure 5.3.2). Depending on the type of neurotransmitter, an EPSP or IPSP occurs in the dendrite of the post-synaptic cell. Neurotransmitters that open Na+ or calcium (Ca+) channels cause an EPSP; an example is the NMDA receptors, which are activated by glutamate (the main excitatory neurotransmitter in the brain). In contrast, neurotransmitters that open Cl- or K+ channels cause an IPSP; an example is gamma-aminobutryric acid (GABA) receptors, which are activated by GABA, the main inhibitory neurotransmitter in the brain. Once the EPSPs and IPSPs occur in the postsynaptic site, the process of communication within and between neurons cycles on (see Figure 5.3.3). A neurotransmitter that does not bind to receptors is broken down and inactivated by enzymes or glial cells, or it is taken back into the presynaptic terminal button in a process called *reuptake* (Figure 5.3.2, Step 6).

5.3.2 SYNAPTIC TRANSMISSION

The transmission of neural messages across the synaptic gap, via the release of transmitter from a pre-synaptic neuron onto **post-synaptic receptor sites** on a post-synaptic neuron, is called **synaptic transmission or neurotransmission**. Synaptic transmission is central to the brain's capacity to process information, to generate mental states, and to generate adaptive behavior. Some synapses are purely electrical and make direct electrical connections between neurons (for example, between basket cells in the cerebellum). However, most synapses are chemical synapses, involving neurotransmitters. The transmission of neural signals across **chemical synapses** is more complex than at **electrical synapses** and involves many steps.

5.3.3 **Electrical Synapses**

Electrical synapses are much less common than chemical synapses, but they are nevertheless distributed throughout the brain. In chemical synapses neurotransmitter is needed for communication between neurons, but for electrical synapses this is not the case. In electrical synapses current, ions, and molecules can flow between two neurons through direct physical connections that allow cytoplasm to flow between them. The physical connection between neurons with electrical synapses is in the form of large pore structures, called connexons, at **gap junctions** between such neurons. Communication between neurons with electrical synapses is faster than at chemical synapses which must go through more steps to transmit signals to another neuron. Therefore, electrical synapses are often found in neural systems that require rapid response such as defensive reflexes. Electrical synapses can communicate signals in both directions between neurons in contrast to chemical synapses which transmit messages in one direction, from pre-synaptic by transmitter release to post-synaptic neuron.

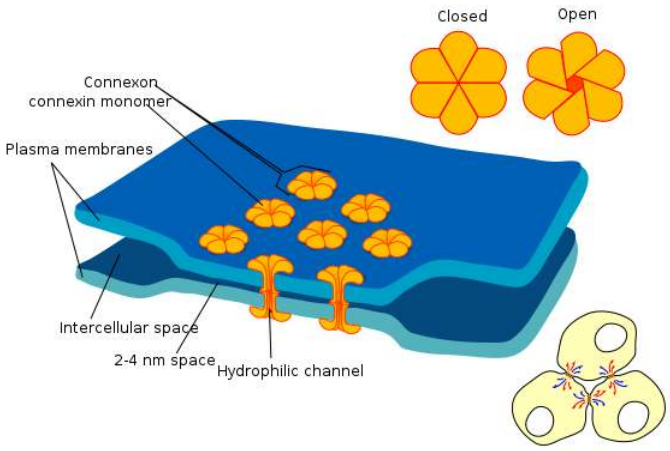


Figure **5.3.1**: Gap junction at an electrical synapse. Two adjacent neurons with electrical synapse between them can communicate through hydrophilic channels Note that the gap between cell membranes of pre and post-synaptic neurons at electrical synapses is much smaller than the synaptic gap at chemical synapses which is about 10 times larger. (Image from Wikimedia Commons; File:Gap cell junction-en.svg; by Mariana Ruiz LadyofHats.https://commons.wikimedia.org/wiki/F...unction-en.svg; by Mariana Ruiz LadyofHats; **public domain** by its author, **LadyofHats**).

5.3.4 Chemical Synapses

We discussed previously which ions are involved in maintaining the resting membrane potential. Not surprisingly, some of these same ions are involved in the action potential. When the cell becomes depolarized (more positively charged) and reaches the threshold of excitation, this causes a voltage-dependent Na+ channel to open. A voltage-dependent ion channel is a channel that opens, allowing some ions to enter or exit the cell, depending upon when the cell reaches a particular membrane potential (i.e. a particular voltage).

When the cell is at resting membrane potential, these voltage-dependent Na+ channels are closed. As we learned earlier, both diffusion from concentration gradients and electrostatic pressure from charge gradients are pushing Na+ ions toward the inside of the cell. However, Na+ cannot permeate the membrane when the cell is at rest.

Once these channels are opened when trigger threshold has been reached, Na+ rushes inside the cell, causing the cell to become very positively charged relative to the outside of the cell. This is responsible for the rising or depolarizing phase of the action potential (see Module 5.2). The inside of the cell becomes very positively charged, from about +30mV to about +60mv, depending upon the particular neuron. At this point, the Na+ channels close and become *refractory.* This means the Na+ channels cannot reopen again until after the cell returns to the resting membrane potential. Thus, a new action potential cannot occur during the refractory period. The refractory period also ensures the action potential can only move in one direction down the axon, away from the soma.

As the cell becomes more depolarized, a second type of voltage-dependent channel opens; this channel is permeable to K+. With the inside of the cell very positive relative to the outside of the cell (depolarized) and the high concentration of K+ within the cell, both the force of diffusion, along concentration gradients, and the force of electrostatic pressure, along charge gradients, drive K+ outside of the cell. The movement of K+ out of the cell causes the cell potential to return back to the resting membrane potential; this is the falling or hyperpolarizing phase of the action potential (see Module 5.2).

A short hyperpolarization occurs partially due to the gradual closing of the K+ channels. With the Na+ channels closed, electrostatic pressure from a K+ charge gradient continues to push K+ out of the cell. In addition, the sodium-potassium pump is pushing Na+ out of the cell. The cell returns to the resting membrane potential, and the excess extracellular K+ diffuses away. This exchange of Na+ and K+ ions happens very rapidly, in less than one millisecond. The action potential occurs in a wave-like motion down the axon until it reaches the terminal button. Only the ion channels in very close proximity to the action potential are affected, causing the action potential to be regenerated along the axon--this creates the movement of the action potential down the axon.

The binding of a neurotransmitter to its receptor is reversible, and for good reason. As long as it is bound to a post synaptic receptor, a neurotransmitter continues to affect membrane potential, so must be removed from the synapse. The effects of the neurotransmitter generally lasts a few milliseconds before being terminated. If the used transmitter is not removed or inactivated it can cause over-activation of neurons, potentially leading to pathological mental states and behavior if enough synapses are affected. Neurotransmitter termination can occur in three ways: First, reuptake by astrocytes or by the presynaptic terminal where the used transmitter can be destroyed by enzymes (Figure 5.3.2, step 6). Second, degradation by enzymes in the synaptic cleft such as the enzyme, acetylcholinesterase, which destroys used acetylcholine transmitter. Third, diffusion of the neurotransmitter as it moves away from the synapse. Again, destruction or removal of used transmitter from the synapse after it has done its work is essential to normal functioning of the nervous system.

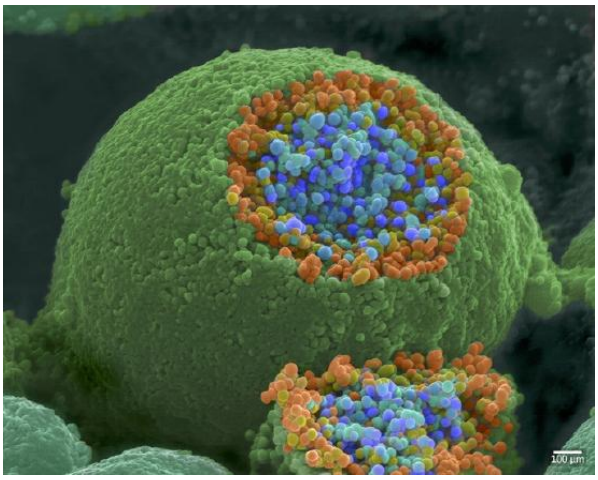


Figure **5.3.2**: **Synaptic vesicles inside a neuron**: This pseudocolored image taken with a scanning electron microscope shows an axon terminal that was broken open to reveal synaptic vesicles (blue and orange) inside the neuron. (Image and caption adapted from: General Biology (Boundless), Chapter 35, The Nervous System; https://bio.libretexts.org/Bookshelv...gy_(Boundless); LibreTexts content is licensed by CC BY-NC-SA 3.0. Legal).

5.3.5 Ionotropic and Metabotropic Synapses

There are two major categories of post-synaptic receptors: ionotropic receptors and metabotropic receptors.

Ionotropic receptors, when activated by transmitter from a pre-synaptic neuron, cause ion channels to open, allowing ions, with their electrical charges, to move across the cell membrane of the receiving (post-synaptic) neuron, causing an EPSP or an IPSP (see above). These are fast acting receptors and are involved in the kind of neural transmission we have been describing above. Ionotropic receptors are receptors on ion channels that open, allowing some ions to enter or exit the cell, depending upon the presence of a particular neurotransmitter. The type of neurotransmitter and the permeability of the ion channel it activates will determine if an EPSP or IPSP occurs in the dendrite of the post-synaptic cell. These EPSPs and IPSPs summate and determine whether an action potential will be generated. For a video summary, copy and paste this web address into your browser: https://nobaproject.com/modules/neurons.

By contrast, **metabotropic receptors** (usually coupled with G-proteins; i.e. guanine nucleotide-binding proteins), when activated by transmitter from a pre-synaptic neuron, act indirectly and more slowly, using second messengers to produce a variety of metabolic effects to modulate cell activity. These effects include changes in gene transcription, regulation of proteins in the cell, release of Ca+ (calcium ions) within the cell, and effects on ion channels on the neuron's cell membrane (Sterling & Laughlin, 2015). Such modulation of neurons and synapses can be more long-lasting than effects of the activation of ionotropic receptors and may play an important role in cellular level mechanisms of learning and memory (Nadim and Bucher, 2014).

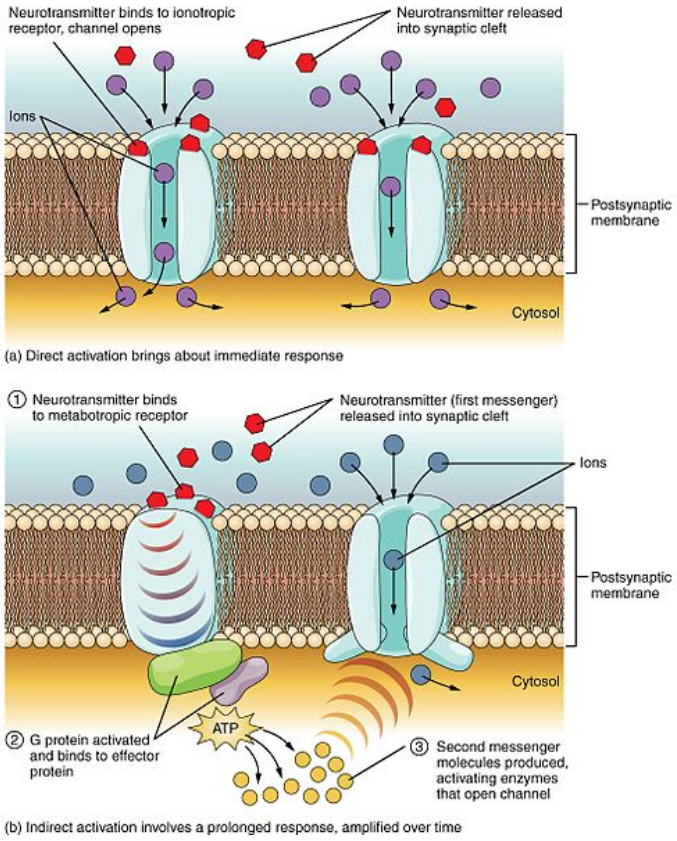


Figure **5.3.3**: Comparison of Ionotropic and Metabotropic Post-Synaptic Receptors. The image at the top (a) shows ionotropic receptors which when activated by transmitter open ion channels immediately resulting in ion movement and an immediate response, a post-synaptic potential. The image at the bottom (b) shows metabotropic receptors which when activated by transmitter initiate a second messenger system. Second messengers can have a variety of effects including indirectly opening ion channels (Image from Wikimedia Commons; File:1226 Receptor Types.jpg; https://commons.wikimedia.org/wiki/F...ptor_Types.jpg; by OpenStax; licensed under the Creative Commons Attribution 4.0 International license).

5.3.6 8-Step Summary of Synaptic Transmission

Here's a somewhat simplified but useful 8-step summary of the **steps in synaptic transmission** at synapses with **ionotropic receptors**:

2) increased intracellular Ca+2 at the axon ending **binds synaptic vesicles to pre-synaptic membrane triggering release of transmitter** from synaptic vesicles in the axon ending of this pre-synaptic neuron

3) **molecules of neurotransmitter cross the fluid-filled synaptic space** between pre- and post-synaptic neurons

4) **molecules of transmitter attach to post-synaptic** ionotropic **receptor sites on** the dendrite or soma of the **post-synaptic neuron** on the other side of the synaptic space

Generic Neurotransmitter System

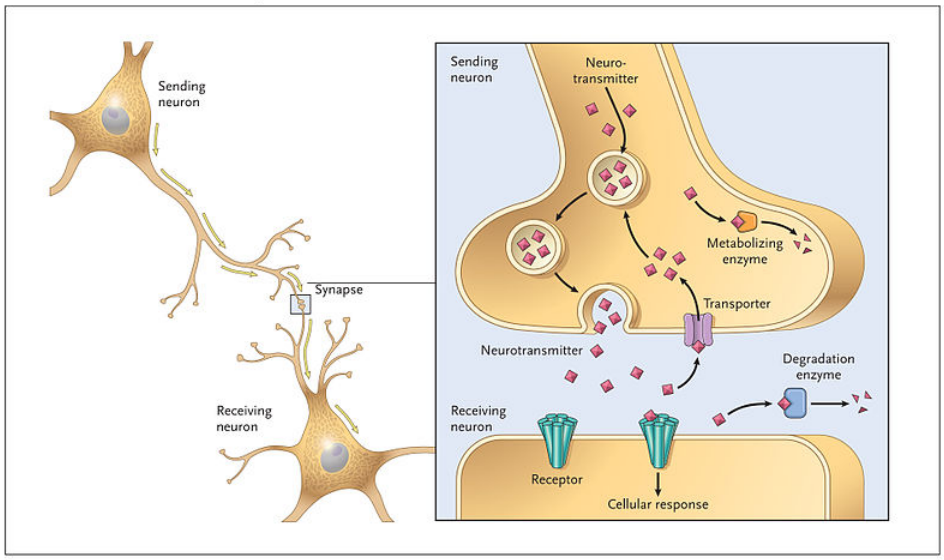


Figure 5.3.4: Image on left shows two neurons communicating with one another (synaptic transmission). The neuron at the top is the sender or pre-synaptic neuron. The neuron on the bottom is the receiver or post-synaptic neuron. In the image on the left of the figure, notice the small box at the synapse. This box is enlarged in the image on the right side of the figure to reveal details of the synapse and events there during synaptic transmission. The axon ending or bouton/button of the pre-synaptic neuron is at the top. We see transmitter molecules (small red squares) transported down the axon (arrow) from the cell body to the axon ending where they are stored in synaptic vesicles (yellow circles containing small red squares, i.e. transmitter molecules). When a nerve impulse (action potential) reaches the axon ending, calcium ions enter the axon ending/bouton/button causing synaptic vesicles to move to the axon ending's cell membrane (steps 1 and 2 referred to above) causing release of transmitter molecules into and across the synaptic gap (steps 2 and 3). Transmitter molecules bind to post-synaptic receptors (green) causing specific ion channels ("doors") to open to specific ions (steps 4 and 5). Ions (not shown in this figure) move through the ion channels in the post-synaptic neuron's cell membrane (step 6) with their electrical charges changing the voltage inside the post-synaptic neuron (either an EPSP or an IPSP). If trigger threshold (about negative 55 millivolts) is reached, an action potential is generated in the receiver neuron, which now becomes a sender neuron for the next cell in line (not shown in this figure). Inactivation of the used transmitter is the final step, step 8, and is depicted in this figure by "enzyme degradation" (enzymatic destruction). (Image from Wikimedia Commons; File:Generic Neurotransmitter System.jpg; https://commons.wikimedia.org/wiki/F...ter_System.jpg; by NIDA(NIH); this work is in the **public domain** in the United States. Caption by Kenneth A. Koenigshofer, Ph.D.).

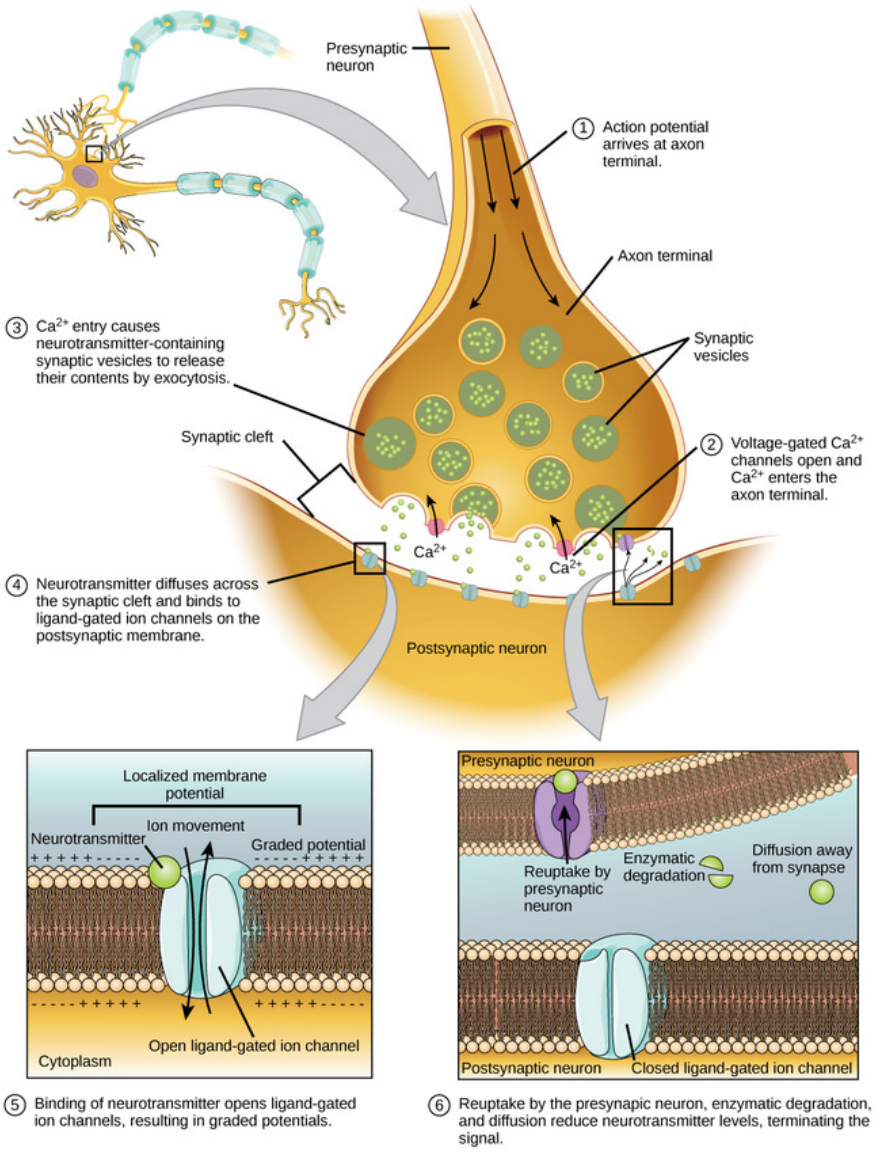
5) **this "lock and key" interaction, between transmitter molecules and** the **receptor sites** that they attach to, "opens doors," that is, causes holes or pores (**ion channels) in the post-synaptic neuron's cell membrane to open to specific ions** (electrically charged atoms; sodium, chloride, potassium)

6) **ions move (along charge and concentration gradients)** through specific ion channels across the cell membrane into or out of the post-synaptic neuron, carrying with them their electrical charges, **altering the electrical potential inside the post-synaptic neuron** (as noted above, this voltage shift is called a post-synaptic potential or **PSP**). **If the transmitter is excitatory**, then **sodium channels open** and **sodium ions (Na+) follow the electric and charge gradients into the post-synaptic neuron** making the **voltage inside the neuron more positive (an EPSP)**. **If the transmitter is inhibitory**, think of chloride ions (Cl -) moving in and potassium ions (K+) moving out, making the **inside of the post-synaptic neuron more negatively charged**, moving the neuron's voltage further away from trigger threshold, thus inhibiting the neuron from firing (this is **an IPSP**). EPSPs and IPSPs can summate.

7) **if the voltage shift** in the post-synaptic neuron is positive (an EPSP) and if it is large enough, or if summated inputs (summation) are **positive enough to reach "trigger threshold"** of the post-synaptic neuron (-55 millivolts), then **Na+ channels suddenly open (voltage gated channels), Na+ rushes inside the cell, and an action potential is generated** in the post-synaptic (receiver) neuron, which now becomes a pre-synaptic (sender) neuron for the next cell in line. The action potential is generated at

the axon hillock and then is **conducted down the length of the axon**, jumping from node to node if the axon is myelinated. As the Na+ rushes inside the cell, this is responsible for the rising or depolarizing phase of the action potential (see Figure 5.3.2).

8) **Inactivation of used transmitter** (by its reuptake into the pre-synaptic neuron or by its enzymatic destruction by specific enzymes)



5.3.6.0.1

Figure **5.3.5**: **Communication at a chemical synapse**: Communication at chemical synapses requires release of neurotransmitters (refer to 8-step summary above). When the pre-synaptic membrane is depolarized, voltage-gated Ca2+ channels open and allow Ca2+ to enter the cell. The calcium entry causes synaptic vesicles to fuse with the membrane and release neurotransmitter molecules into the synaptic cleft. The neurotransmitter diffuses across the synaptic cleft and binds to ligand-gated ion channels (channels opened by chemical transmitter) in the post-synaptic membrane, resulting in a localized depolarization (EPSP) or hyperpolarization (IPSP) of the post-synaptic neuron depending upon the type of transmitter, excitatory or inhibitory. (Image and caption adapted from: General Biology (Boundless), Chapter 35, The Nervous System; https://bio.libretexts.org/Bookshelv...gy_(Boundless); LibreTexts content is licensed by CC BY-NC-SA 3.0. Legal).

5.3.6.1 Types of Neurotransmitter

There are at least 60-100 neurotransmitters and probably many others yet to be discovered. The best known can be grouped into types based on their chemical structure.

Today, the majority of neuroscientists will tell you that most neurons release the same neurotransmitter from their axons—which is why you may see some neurons referred to as “dopaminergic” or “serotonergic,” releasing dopamine or serotonin, respectively. But new work in the field has uncovered that neurons are not fixed when it comes to the chemicals they release.

Some cells change the type of neurotransmitters they release depending on the circumstances, sometimes releasing up to five different kinds. Scientists call this phenomenon “neurotransmitter switching.”

Neurotransmitters, at the highest level, can be sorted into two types: small-molecule transmitters and neuropeptides. Small-molecule transmitters, like dopamine and glutamate, typically act directly on neighboring cells. The neuropeptides, small molecules like insulin and oxytocin, work more subtly, modulating, or adjusting, how cells communicate at the synapse. These powerful neurochemicals are at the center of neurotransmission, and, as such, are critical to human cognition and behavior.

Often, neurotransmitters are talked about as if they have a single role or function. Dopamine is a “pleasure chemical” and GABA is a “learning” neurotransmitter. But neuroscientists are discovering they are multi-faceted and complex, working with and against each other to facilitate neural signaling across the cortex. Here is a list of some of the most common neurotransmitters discussed in neuroscience.

5.3.6.1.1 Amino acids neurotransmitters

- **Glutamate (GLU).** This is the most common and most abundant excitatory neurotransmitter. Glutamate has an important role in cognitive functions like thinking, learning and memory. Too much glutamate results in excitotoxicity, or the death of neurons due to stroke, traumatic brain injury, or amyotrophic lateral sclerosis (ALS), the debilitating neurodegenerative disorder better known Lou Gehrig’s disease. GLU is also important to learning and memory: long term potentiation (LTP), occurs in glutamatergic neurons in the hippocampus and cortex.
- **Gamma-aminobutryic acid (GABA).** GABA works to inhibit neural signaling. GABA is the most common inhibitory neurotransmitter in the nervous system, particularly in the brain. New research suggests that GABA helps lay down important brain circuits in early development. GABA also has a nickname: the “learning chemical.” Studies have found a link between the levels of GABA in the brain and whether or not learning is successful.
- **Glycine.** Glycine is the most common inhibitory neurotransmitter in the spinal cord and is involved in auditory processing, pain and metabolism.

5.3.6.1.2 Monoamines neurotransmitters

Monoamines neurotransmitters are involved in consciousness, cognition, attention and emotion.

- **Serotonin (5HT).** Serotonin is an inhibitory neurotransmitter. Serotonin helps regulate mood, sleep patterns, sexuality, anxiety, appetite and pain. Diseases associated with serotonin imbalance include seasonal affective disorder, anxiety, depression, fibromyalgia and chronic pain. Medications that regulate serotonin and treat these disorders include selective serotonin reuptake inhibitors (SSRIs) and serotonin-norepinephrine reuptake inhibitors (SNRIs), which increase the levels of transmitter by inhibiting their reuptake after they have done their work at the postsynaptic receptor sites. Serotonin (5HT), sometimes called the “calming chemical,” is best known for its mood modulating effects. A lack of 5HT has been linked to depression and related neuropsychiatric disorders. 5HT has also been implicated in facilitating memory, and, most recently, in decision-making behavior
- **Histamine.** Histamine regulates body functions including wakefulness, feeding behavior and motivation.
- **Dopamine (DA).** Dopamine is involved in the brain's reward system, feelings of pleasure, cognitive arousal, learning, focus of attention, concentration, memory, sleep, mood and motivation. Dopamine is often referred to as the “pleasure chemical” because it is released when mammals receive a reward in response to their behavior; that reward could be food, drugs, or sex. Diseases associated with dysfunctions of the dopamine system include Parkinson’s disease, schizophrenia, bipolar disease, restless legs syndrome and attention deficit hyperactivity disorder (ADHD). Many highly addictive drugs (cocaine, methamphetamines, amphetamines) act directly on the brain's dopamine circuits.
- **Epinephrine.** Epinephrine (also called adrenaline) and norepinephrine (see below) are responsible for the “fight-or-flight response” to fear and stress, activating the sympathetic nervous system (see Chapter 4).
- **Norepinephrine (NE).** Norepinephrine is both a hormone and a neurotransmitter. Some refer to it as noradrenalin. It has been linked to mood, arousal, vigilance, memory, and stress. Newer research has focused on its role in both post-traumatic stress disorder (PTSD) and Parkinson’s disease. Norepinephrine (noradrenaline) increases blood pressure, heart rate, alertness,

arousal, attention and focus. Many medications (stimulants and depression medications) increase norepinephrine to improve focus or concentration to treat ADHD or to reduce symptoms of depression.

5.3.6.1.3 Peptide neurotransmitters

Peptides are short chains of amino acids.

- **Endorphins.** Endorphins are natural pain killers. They are natural opiate-like substances, similar in molecular properties to morphine. Release of endorphins reduces pain, and elevates mood. They are endogenous opiates released by the hypothalamus and pituitary gland during stress or pain.

5.3.6.1.4 Acetylcholine (ACh)

This excitatory neurotransmitter is found in both the central and peripheral nervous systems, specifically in the autonomic nervous system and in spinal motor neurons releasing acetylcholine onto skeletal muscles producing movements. It is also involved in memory, motivation, sexual desire, sleep and learning. Abnormalities in acetylcholine levels are associated with Alzheimer's disease. But it also has other roles in the brain, including helping direct attention and playing a key role in facilitating neuroplasticity across the cortex.

5.3.6.1.5 **Other Neurotransmitters**

Neurochemicals like oxytocin and vasopressin are also classified as neurotransmitters. Made and released from the hypothalamus, they act directly on neurons and have been linked to pair-bond formation, monogamous behaviors, and drug addiction. Hormones like estrogen and testosterone can also work as neurotransmitters and influence synaptic activity.

Other neurotransmitter types include corticotropin-releasing factor (CRF), galanin, enkephalin, dynorphin, and neuropeptide Y. CRH, dynorphin, and neuropeptide Y have been implicated in the brain's response to stress. Galanin, encephalin, and neuropeptide Y are often referred to as "co-transmitters," because they are released and then work in partnership with other neurotransmitters. Enkephalin, for example, is released with glutamate to signal the desire to eat and process rewards.

As neuroscientists are learning more about the complexity of neurotransmission, it's clear that the brain needs these different molecules so it can have a greater range of flexibility and function.

5.3.6.1.6 **Glia Release Neurotransmitters, Too**

It was once believed that only neurons released neurotransmitters. New research, however, has demonstrated glia, the cells that make up the "glue" that fills the space between neurons to help support and maintain those cells, also have the power to release neurotransmitters into synapses. In 2004, researchers found that glial cells release glutamate into synapses in the hippocampus, helping synchronize signaling activity.

Astrocytes, a star-shaped glial cell, are known to release a variety of different neurotransmitters into the synapse to help foster synaptic plasticity, when required. Researchers are working diligently to understand the contributions of these different cell types–and the neurotransmitter molecules they release—on how humans think, feel, and behave.

5.3.6.2 Different Types of Receptors Activated by the Same Transmitter

There are different types of receptor for transmitters of a specific type. For example, for dopamine there are at least 5 types, D1 through D5 receptors, all for dopamine transmitter. Different receptor types for a specific transmitter may be localized in different parts of the brain and therefore may produce different effects and have different functions.

The function of each dopamine receptor type (Mishra, et al., 2018):

- D1: memory, attention, impulse control, regulation of renal (kidney) function, locomotion (movement)
- D2: locomotion, attention, sleep, memory, learning
- D3: cognition, impulse control, attention, sleep
- D4: cognition, impulse control, attention, sleep
- D5: decision making, cognition, attention, renin secretion (by the kidney)

Of particular interest, the mental disorder schizophrenia, characterized by disordered thought, hallucinations, and delusions, is associated with excess dopamine neuron activity in the brain. Some drug treatments for schizophrenia decrease activity primarily at

D2 receptors.

Another example is acetylcholine transmitter. There are two distinct types of acetylcholine receptors affected by two different substances, either muscarine or nicotine. Those postsynaptic acetylcholine receptors that respond to muscarine are called muscarinic receptors. Those that respond to nicotine (in tobacco products, for example) are called nicotinic. Nicotinic receptors cause sympathetic postganglionic neurons and parasympathetic postganglionic neurons to fire and release their chemicals and cause skeletal muscle to contract. Muscarinic receptors are associated mainly with parasympathetic functions and stimulate receptors located in peripheral tissues (e.g., glands, smooth muscle). Acetylcholine transmitter activates all of these sites.

5.3.6.3 Synaptic Transmission in Review

Let's summarize the sequence of events in synaptic transmission in synapses with ionotropic receptors:

Action potential in pre-synaptic neuron --- transmitter release --- transmitter crosses synaptic gap --- transmitter attaches to post-synaptic receptor sites --- ion channels open --- ions move --- PSP results --- if the PSP is an EPSP and if the EPSP is big enough (or if summation results in voltage change big enough), then trigger threshold of -55mv is reached in post-synaptic neuron ---post-synaptic neuron produces its own action potential. These same steps are summarized in the first 7 of the 8 steps in synaptic transmission listed above. The last step is transmitter inactivation.

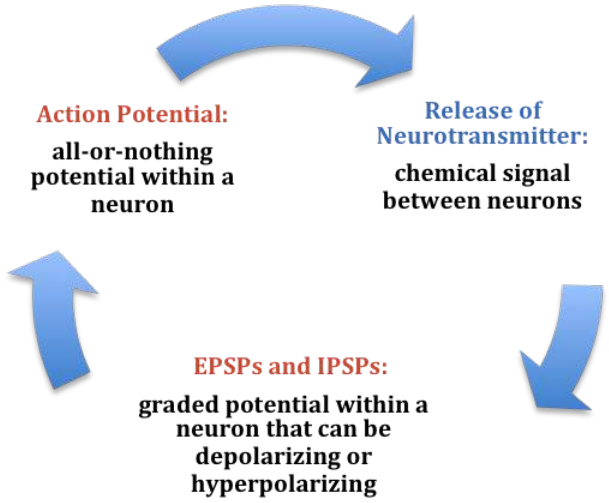


Figure **5.3.6**: Summary of the electrochemical communication within and between neurons. (Image from Furtak, S. (2021). Neurons. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology.* Champaign, IL: DEF publishers. Retrieved from http://noba.to/s678why4).

Step 8 in the list of 8 steps above simply refers to the fact that after the released transmitter molecules have done their job, the used transmitter molecules must be cleaned out of the synapse. If not, the receiving neuron will be affected for too long and this may result in dysfunction at the synapse, resulting in pathological effects, and if occurring at a large number of synapses can lead to abnormalities such as seizures or hallucinations or other maladaptive effects.

As noted above, removal of used transmitter is accomplished by two mechanisms, reuptake and enzymatic destruction. Reuptake means that the used transmitter is reabsorbed back into the pre-synaptic ending from which it was released. Enzymatic destruction means that the used transmitter is chemically destroyed by an enzyme, and thereby inactivated.

If step 8, transmitter inactivation, didn't occur, in the case of excitatory transmitters, then the post-synaptic neuron would be over activated. If this occurs on a large scale, at many synapses, as noted above, behavioral and mental abnormalities will result. For example, there are some drugs (Soman, Sarin, Malathion) that block enzymatic destruction of the neurotransmitter Acetylcholine, ACh. As noted above, ACh is a transmitter involved in various functions in the brain and peripheral nervous system, including stimulation of skeletal muscles, which are responsible for movement. Soman and Sarin are nerve agents. They block the enzyme acetylcholinesterase, AChE, and thereby prevent the enzymatic destruction of ACh at motor synapses which stimulate the muscles. As a result, the excess ACh at the muscles causes their over activation. The result? Epileptic seizures so intense that death may

occur. The treatment for poisoning by these agents is a drug that blocks acetylcholine neurotransmitter at the acetylcholine receptor sites, counteracting effects of the excess acetylcholine neurotransmitter, thereby preventing or reversing the seizures otherwise caused by these nerve agents.

5.3.7 Neurons, The Mind, and Artifical Neural Networks

How do these neural processes relate to the real world of our everyday conscious experience and behavior? As mentioned in module 5.2, in complex vertebrate species, single nerve cells and their activity do not control a behavior or create a thought or a feeling. Instead, information is processed, mental states are created, and behaviors are organized by neural events in complex circuits involving very large numbers of neurons. It is the interaction among huge numbers of neurons, interconnected by enormous numbers of synapses, that create our perceptions, thoughts, emotions, or a complex behavior such as human speech or the creation of a work of art.

How groups of neurons in complex circuits might produce complex things like a perception or a memory has been studied using computer modeling (see Module 10.9). Artificial neural networks are computer-based models of neural circuits and their functioning. In a computer program, artificial information processing units, sometimes called "neurodes", are programmed to code and process information similar to how neurons process information. These neuron-like processing units are interconnected with one another and organized into layers, and then these layers of processing units are connected.

A three-layered artificial neural network (see Module 10.9) is capable of performing some very complex information processing tasks, producing responses similar to those that a real brain might produce. Furthermore, because the artificial neurodes or processing units are given (by programmers) the capability of altering the strengths or "weights" of their "synaptic connections", artificial neural networks are capable of learning when given feedback about their performance. Given these properties, it has been shown by researchers (many of whom are at UC San Diego) that artificial neural networks can learn patterns and regularities in inputs to make generalizations, to form categories, and to solve complex problems (Churchland, 2013).

For example, one neural network, trained (electronically) in the mathematical principles used to find mathematical proofs, discovered a proof that had eluded human mathematicians for many decades. Another, known as NETtalk, equipped with a voice synthesizer and photocell "eyes" learned from scratch to read and correctly pronounce written English text. It accomplished this astounding feat after only 10 hours of trial and error learning, during which it was given feedback about its attempts at pronunciation. It went through a babbling stage initially, like a child would, but soon was articulating comprehensible speech. When tested on new material, it generalized what it had learned to the new text including many new words that it had not "seen" before. An analysis by researchers of how it did this showed that it had formed categories of letters, learning to divide them into vowels and consonants, with rules of pronunciation for each category. It accomplished this without direct programming by humans. It extracted regularities from the words presented to it, and made generalizations, rules which it applied to new cases (Churchland, 2013).

Artificial neural networks use a form of information processing called parallel distributed processing (PDP). Though these networks are not brains, they have come closer than any other computer models to simulating brain function and psychological capabilities of real brains. This suggests that biological brains may use PDP, carried out by PDP networks made of real neurons, to produce the kinds of psychological and behavioral capacities which real brains like ours demonstrate. This is a very promising area of research being carried out in the psychology, computer science, and neuroscience departments at many universities worldwide.

Keep in mind that the circuitry in biological brains, such as the human brain, though modifiable by experience, is also hard-wired to a significant extent by natural selection and other forces of evolution (see Chapter 3). This accounts for the fact that human beings have an innate or inborn human nature just like other animals have their own innate natures. Humans relate to the world in a distinctively human way whereas wolves, tigers, birds, wildebeests, and other animal species understand and relate to the world in their own particular species-typical ways. Each species' distinctive psychological nature is a result of the brain circuitry inherited from its predecessors--circuitry molded by eons of evolution (Koenigshofer, 2011).

5.3.7.1 References

Bekkers, J. M. (2011). Pyramidal neurons. *Current biology*, *21*(24), R975.

Brown, A. M., Arancillo, M., Lin, T., Catt, D. R., Zhou, J., Lackey, E. P., ... & Sillitoe, R. V. (2019). Molecular layer interneurons shape the spike activity of cerebellar Purkinje cells. *Scientific reports*, *9* (1), 1-19.

Churchland, P. M. (2013). *Matter and consciousness*. MIT press.

Fields, R. D., & Stevens-Graham, B. (2002). New insights into neuron-glia communication. *Science*, *298* (5593), 556-562.

Kirkcaldie, M. T. (2012). Neocortex. In *The mouse nervous system* (pp. 52-111). Academic Press.

Koenigshofer, K.A. (2011). *Mind Design: The Adaptive Organization of Human Nature, Minds, and Behavior*. Pearson Education, Boston.

Mishra A, Singh S, Shukla S. (2018). Physiological and Functional Basis of Dopamine Receptors and Their Role in Neurogenesis: Possible Implication for Parkinson's disease. *J Exp Neurosci. 12:*1179069518779829. [PMC free article] [PubMed]

Nadim, F., & Bucher, D. (2014). Neuromodulation of neurons and synapses. *Current opinion in neurobiology*, *29*, 48-56.

Pasternak, J. F., & Woolsey, T. A. (1975). On the "selectivity" of the Golgi-Cox method. *J Comp Neurol*, 160(3), 307-312. doi: 10.1002/cne.901600304

Perea, G., Sur, M., & Araque, A. (2014). Neuron-glia networks: integral gear of brain function. *Frontiers in cellular neuroscience*, *8*, 378.

Smit, G. J., & Colon, E. J. (1969). Quantitative analysis of the cerebral cortex. I. Aselectivity of the Golgi-Cox staining technique. *Brain Res*, 13(3), 485-510.

Sterling, P., & Laughlin, S. (2015). *Principles of neural design*. MIT press.

University of Queensland, Queensland Brain Institute, Types of Neurons, n.d. https://qbi.uq.edu.au/brain/brain-an.../types-neurons; Retrieved 8/31/21

Van Essen, D., & Kelly, J. (1973). Correlation of cell shape and function in the visual cortex of the cat. *Nature*, *241*(5389), 403-405.

5.3.8 Chapter 5 Review

1. Define nerve impulse.
2. What is the resting potential of a neuron, and how is it maintained?
3. Explain how and why an action potential occurs.
4. Outline how a signal is transmitted from a presynaptic cell to a postsynaptic cell at a chemical synapse.
5. What generally determines the effects of a neurotransmitter on a postsynaptic cell?
6. Identify three general types of effects neurotransmitters may have on postsynaptic cells.
7. Explain how an electrical signal in a presynaptic neuron causes the transmission of a chemical signal at the synapse.
8. The flow of which type of ion into the neuron results in an action potential?
    1. How do these ions get into the cell?
    2. What does this flow of ions do to the relative charge inside the neuron compared to the outside?
9. The sodium-potassium pump:
    1. is activated by an action potential
    2. requires energy
    3. does not require energy
    4. pumps potassium ions out of cells
10. *True or False.* Some action potentials are larger than others, depending on the amount of stimulation.
11. *True or False.* Synaptic vesicles from the presynaptic cell enter the postsynaptic cell.
12. *True or False.* An action potential in a presynaptic cell can ultimately cause the postsynaptic cell to become inhibited.
13. Name three neurotransmitters.

5.3.9 Chapter 5 Discussion Questions

1. What structures of a neuron are the main input and output channels of that neuron?
2. What does the statement mean that communication within and between cells is an electrochemical process?
3. How does myelin increase speed and efficiency of the action potential?

4. How does diffusion (concentration gradients) and electrostatic pressure (electrical gradients) contribute to the resting membrane potential and the action potential?
5. Describe the cycle of communication within and between neurons.

5.3.10 Chapter 5 Vocabulary

Action potential (nerve impulse)

A transient all-or-nothing positive electrical current (depolarization) that is conducted down the axon when the membrane potential reaches the threshold of excitation (trigger threshold equal to about negative 55 millivolts).

Axon

Part of the neuron that extends off the soma, splitting several times to connect with other neurons; main output of the neuron.

Cell membrane

A bi-lipid layer of molecules that separates the cell from the surrounding extracellular fluid.

Dendrite

Part of a neuron that extends away from the cell body and is the main input to the neuron.

Diffusion (along a concentration gradient or inequality)

The force on molecules to move from areas of high concentration to areas of low concentration.

Electrostatic pressure (charge gradient)

The force on two ions with similar charge to repel each other; the force of two ions with opposite charge to attract to one another.

Excitatory postsynaptic potentials (EPSPs)

A graded depolarizing postsynaptic current that causes the membrane potential to become more positive and move towards the threshold of excitation (trigger threshold). EPSPs and IPSPs can summate with one another.

Inhibitory postsynaptic potentials (IPSPs)

A graded hyperpolarizing postsynaptic current that causes the membrane potential to become more negative and move away from the threshold of excitation (trigger threshold). EPSPs and IPSPs can summate with one another.

Ion channels

Proteins that span the cell membrane, forming channels that specific ions can flow through between the intracellular and extracellular space (the "doors" in step 5 in the list of 8 steps of synaptic transmission above; see Figure 5.3.2).

Ionotropic receptor

Receptor (receives transmitter molecules and binds to them) and its associated ion channel that opens to allow ions to permeate the cell membrane under specific conditions, such as the presence of a neurotransmitter (chemically-gated channel) or a specific membrane potential (voltage-gated channel).

Myelin sheath

Substance around the axon of a neuron that serves as insulation to allow the action potential to conduct rapidly toward the terminal buttons.

Neurotransmitters or transmitters

Chemical substance released by the presynaptic terminal button that acts on the postsynaptic cell.

Nucleus

Collection of nerve cells found in the brain which typically serve a specific function.

Resting membrane potential

The voltage inside the cell relative to the voltage outside the cell while the cell is a rest (approximately -70 mV).

Sodium-potassium pump

An ion channel that uses the neuron's energy (adenosine triphosphate, ATP) to pump three Na+ ions outside the cell in exchange for bringing two K+ ions inside the cell.

Soma

Cell body of a neuron that contains the nucleus and genetic information, and directs protein synthesis.

Spines

Protrusions on the dendrite of a neuron that form synapses with terminal buttons of the presynaptic axon.

Synapse

Junction between the presynaptic terminal button of one neuron and the dendrite, axon, or soma of another postsynaptic neuron.

Synaptic gap or synaptic space

Also known as the synaptic cleft; the small space between the presynaptic terminal button (axon ending) and the postsynaptic dendritic spine, axon, or soma.

Synaptic vesicles

Groups of neurotransmitters packaged together and located within the terminal button.

Terminal button or axon ending

The part of the end of the axon that form synapses with postsynaptic dendrite, axon, or soma.

Trigger threshold (Threshold of excitation)

Specific membrane potential that the neuron must reach to initiate an action potential.

5.3.11 Outside Resources

5.3.11.0.1 Video Summary: https://www.stem.org.uk/system/files...he%20brain.mp4

5.3.12 **Attributions**

1. Chapter 5, Communication within the Nervous System, Module 5.3. Neurons and Synaptic Transmission by Kenneth A. Koenigshofer, PhD, Chaffey College, is licensed under CC BY 4.0

2. "Types of Neurotransmitters" and "Glia Release Transmitters Too" adapted by Kenneth A. Koenigshofer from Sukel, K (2019) Neurotransmitters from https://dana.org/article/neurotransmitters/

2. Figure 5.3.3, Vocabulary, Discussion Questions, Outside Resources, and some text adapted by Kenneth A. Koenigshofer from Furtak, S. (2021). Neurons. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology.* Champaign, IL: DEF publishers. Retrieved from http://noba.to/s678why4

Creative Commons License

Neurons by Sharon Furtak is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Permissions beyond the scope of this license may be available in our Licensing Agreement.

3. Figures 5.3.1, 5.3.2 and some text adapted by Kenneth A. Koenigshofer from: General Biology (Boundless), Chapter 35.2C, The Nervous System; Synaptic Transmission, https://bio.libretexts.org/Bookshelv...gy_(Boundless); LibreTexts content is licensed by CC BY-NC-SA 3.0. Legal.

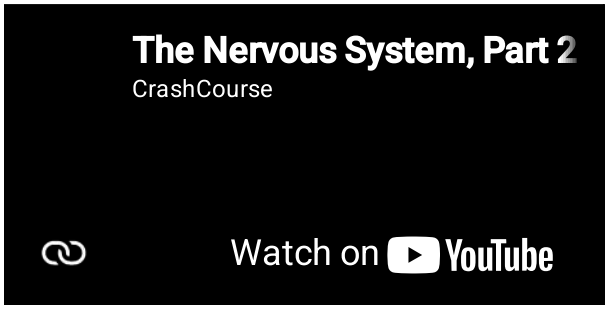
4. Figures 5.2.2 and "Chapter 5 Review" adapted by Kenneth A. Koenigshofer from Chapter 11.3 (Neurons and Glia Cells), 11.4 (Nerve Impulses) in Book: Human Biology (Wakim & Grewal) - Biology LibreTexts by Suzanne Wakim & Mandeep Grewal, under license CC BY-NC

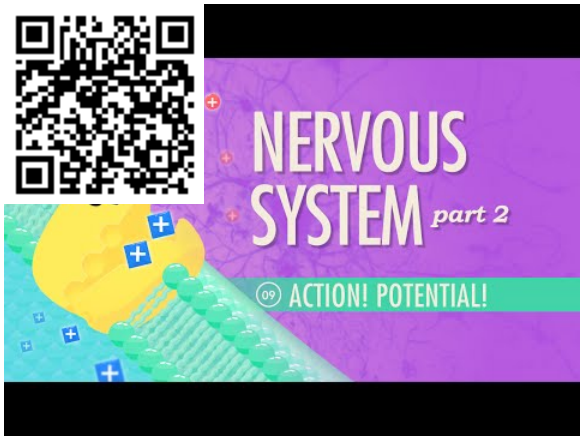
5.3.13 Outside Resources for Chapter 5

Video Series: Neurobiology/Biopsychology - Tutorial animations of action potentials, resting membrane potentials, and synaptic transmission.

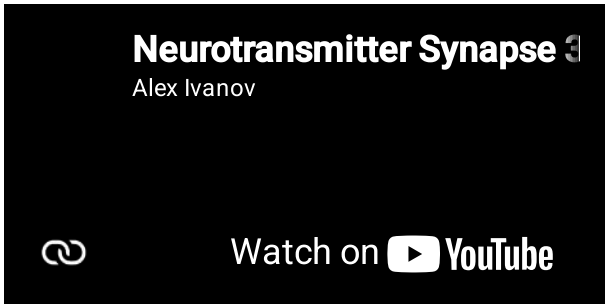
http://www.sumanasinc.com/webcontent/animations/neurobiology.html

Video: An animation and an explanation of an action potential





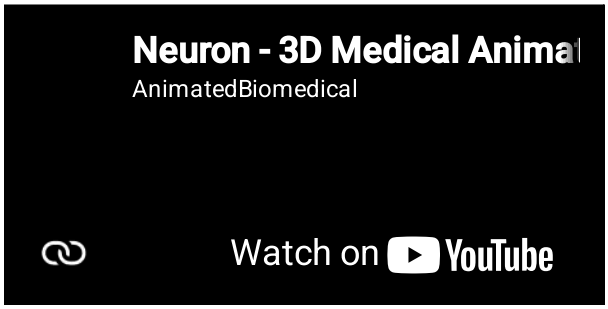
Video: An animation of neurotransmitter actions at the synapse



Vid animation that allows students to observe the results of manipulations to excitatory and inhibitory pos ıls. Also includes animations and explanations of transmission and neural circuits.

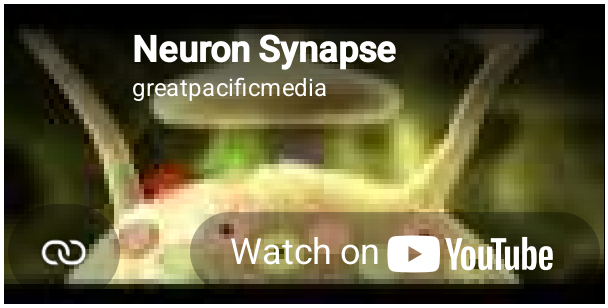
https://apps.childrenshospital.org/clinical/animation/neuron/

Video: Another animation of an action potential



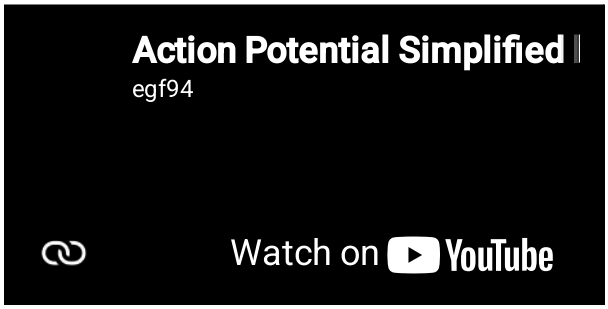


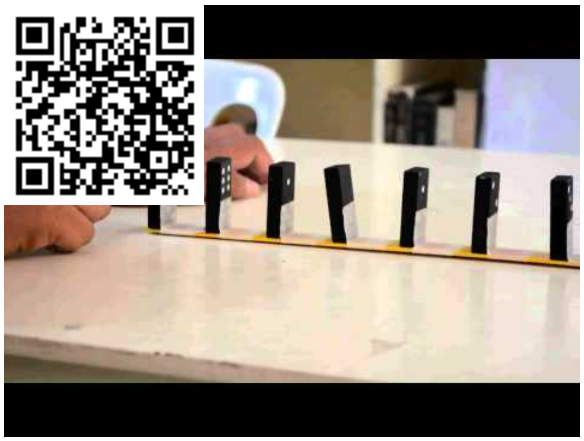
Video: Another animation of neurotransmitter actions at the synapse



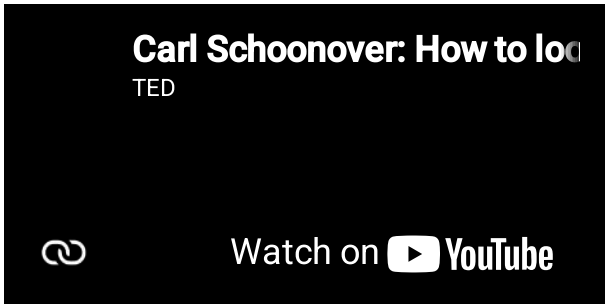


Video: Domino Action Potential: This hands-on activity helps students grasp the complex process of the action potential, as well as become familiar with the characteristics of transmission (e.g., all-or-none response, refractory period).





Video: For perspective on techniques in neuroscience to look inside the brain



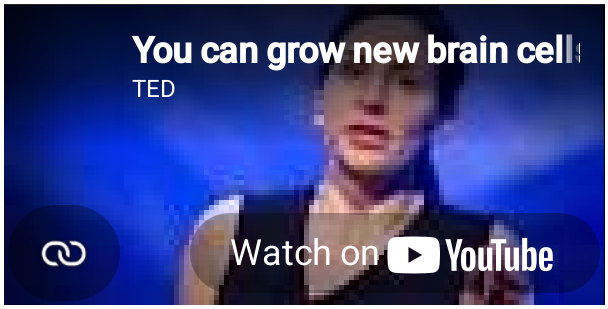


Video: The Behaving Brain is the third program in the DISCOVERING PSYCHOLOGY series. This program looks at the structure and composition of the human brain: how neurons function, how information is collected and transmitted, and how chemical reactions relate to thought and behavior.

http://www.learner.org/series/discoveringpsychology/03/e03expand.html

Video: You can grow new brain cells. Here\\'s how. -Can we, as adults, grow new neurons? Neuroscientist Sandrine Thuret says that we can, and she offers research and practical advice on how we can help our brains better perform neurogenesis—

improving mood, increasing memory formation and preventing the decline associated with aging along the way.





Web: For more information on the Nobel Prize shared by Ramón y Cajal and Golgi

http://www.nobelprize.org/nobel_prizes/medicine/laureates/1906/

This page titled 5.3: Synaptic Transmission is shared under a mixed license and was authored, remixed, and/or curated by Kenenth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

CHAPTER OVERVIEW

6: The Effects of Psychoactive Drugs

> This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

6.1: Psychopharmacology and Psychoactive Drug Classification
6.2: Mechanisms of Psychoactive Drug Action
6.3: Substance Use and Addiction - Theories, Brain Mechanisms, and Treatment
6.4: Specific Drug Effects, Misuse, and Risks
6.5: Psychotherapeutics

This page titled 6: The Effects of Psychoactive Drugs is shared under a mixed license and was authored, remixed, and/or curated by ASCCC OERI & Alan Keys (ASCCC Open Educational Resources Initiative (OERI)) .

6.1: Psychopharmacology and Psychoactive Drug Classification

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe the general principles of psychopharmacology.
2. Explain the criteria for psychoactive drugs and the various types of classification systems.
3. Differentiate drug types based on their neurochemical effects on neurotransmission and behavioral impacts.

6.1.1 Overview

Psychopharmacology is the study of how drugs affect how we think, feel, or behave most often through their actions on the nervous system. Understanding some of the basics about psychopharmacology can help us better explain a wide range of issues that interest psychologists and others. For example, the pharmacological treatment of certain neurodegenerative diseases such as Parkinson's and Alzheimer's disease tells us something about the disease itself. The pharmacological treatments used to treat psychiatric conditions such as schizophrenia or depression have undergone amazing development since the 1950s, and the drugs used to treat these disorders provide clues to what is happening in the brain of individuals with these conditions. Finally, understanding something about the actions of drugs of abuse on the nervous system and the way they are processed by the body can help us understand why some psychoactive drugs are so likely to be abused. In this chapter, we will provide an overview of some of these topics as well as discuss some current controversial areas in the field of psychopharmacology.

6.1.2 Psychopharmacology

Psychopharmacology, the study of how drugs affect the brain and behavior, is a relatively new science, although people have probably been taking drugs to change how they feel from early in human history (consider the eating of fermented fruit, ancient beer recipes, chewing on the leaves of the cocaine plant for stimulant properties as just some examples). The word *psychopharmacology* itself tells us that this is a field that bridges our understanding of behavior, brain function, and pharmacology. Additionally, the topics included within this field are extremely broad including the influence of expectations of drug effects, how the drug is processed by the body, and ultimately, to how the drug impacts various systems, particularly the central nervous system.

6.1.2.1 Psychoactive Drugs

If a drug changes the way you feel, think, or behave it is often doing so by acting on your brain and other parts of your nervous system. We call these ***psychoactive* drugs,** and almost everyone has used them at some point (yes, caffeine counts). Virtually all psychoactive drugs cause psychological or behavioral changes by altering how neurons communicate with each other. It is important to recall that neurons communicate with each other by releasing **neurotransmitters** across the **synapse**. When the neurotransmitter crosses the synapse, it binds to a postsynaptic receptor (protein) on the receiving neuron and the message may then be transmitted onward. Neurotransmission is far more complicated than this, but the first step is understanding that virtually all **psychoactive drugs** alter how neurons communicate with each other in one way or another. While some drugs, such as most antidepressants and many stimulants, have well-defined effects at the level of the neuron, others, like alcohol, have more widespread and less-clear effects.

Some of the most important in terms of psychopharmacological treatment and drugs of abuse are outlined in Table 1. The neurons that release these neurotransmitters, for the most part, are localized within specific circuits of the brain that mediate specific types of behaviors.

Table 1: Key Neurotransmitters and Highlights of Their Neural, Behavioral, and Psychological Effects

Neurotransmitter	Abbreviation	Behaviors or Diseases Related to These Neurotransmitters
Acetylcholine	ACh	Learning and memory; Alzheimer's disease; Voluntary muscle movement in the peripheral nervous system
Dopamine	DA	Reward circuits; Motor control; Parkinson's disease; Schizophrenia
Norepinephrine	NE	Arousal; Depression
Serotonin	5-HT	Depression; Aggression; Schizophrenia
Glutamate	GLU	Learning; Major excitatory neurotransmitter in the brain
GABA	GABA	Anxiety disorders; Epilepsy; Major inhibitory neurotransmitter in the bran
Endogenous Opioids	Endorphins, Enkephalins	Pain; Analgesia; Reward

Psychoactive drugs can either increase the typical effect of neurotransmitters at the synapse (these are called **agonists**) or decrease it (**antagonists**). It is important to remember that the "typical" effect of a neurotransmitter can be either excitatory or inhibitory depending on the type of receptor it binds to. For example, stimulant drugs, such as amphetamine, methamphetamine, or ADHD medications like Adderall are agonists for dopamine and norepinephrine systems and can increase the "typical" *excitation* that occurs at certain dopamine or norepinephrine receptors. Other agonistic drugs, such as morphine and fentanyl will increase the typical *inhibitory* effect at many of our endogenous opioid (endorphin) synapses. In other words, acting as an agonist at an inhibitory receptor will cause even more inhibition than without the drug present. Antagonists for these same systems will have the opposite effects, such as the effect of the opioid receptor blocker, naloxone (Narcan) which results in less inhibition at endogenous opioid synapses. Yes, this is complicated, but you just need to remember that agonist drugs increase and antagonist drugs decrease whatever the typical effect is at a particular synapse, be it excitatory or inhibitory.

There are various ways that drugs can effect the steps of synaptic communication and, thus produce their agonistic and antagonistic effects. Some of these are more intuitive than others. For example, antidepressants such as Prozac or Celexa are selective serotonin reuptake inhibitors (commonly referred to as SSRIs) are agonists for the serotonin system. However, they achieve increased serotonin activity at the synapse by *blocking* the reuptake of serotonin back into the presynaptic axon terminal from which it was released. In other words, they increase serotonin activity by interfering with the mechanism that normally turns its activity off (i.e., reuptake). As you can see, the basic distinction of whether a drug is an agonist or antagonist can get quite complicated very quickly. More examples of agonists and antagonists for various neurotransmitter systems are presented in Table 2. For each example, the drug’s trade name, which is the name of the drug provided by the drug company, and generic name (in parentheses) are provided.

Table 2 provides examples of drugs and their primary mechanism of action, but it is very important to realize that drugs can also have effects on other neurotransmitters or even hormones. Sometimes these varied effects can change the potency and/or effectiveness of a drug. This could prove useful when trying to find just the right combination of transmitter action to treat an individual's psychological disorder. Or it could be problematic if it leads to unwanted side effects. The reality is that most drugs currently available work only where we would like them to in the brain or only on a specific neurotransmitter. In many cases, individuals are sometimes prescribed one drug but then may also have to take additional drugs to reduce the side effects. Sometimes individuals stop taking medication because the side effects are so profound.

Table 2: Drug function and mechanism

Drug	Mechanism	Use	Agonist/Antagonist
Seroquel (quetiapine)	Blocks DA and 5-HT receptors	Schizophrenia, bipolar disorder	Antagonist for DA, 5-HT
L-dopa	Increase synthesis of DA	Parkinson's disease	Agonist for DA
Prozac (fluoxetine)	Blocks removal of 5-HT from synapse (prevents reuptake)	Depression, obsessive compulsive disorder	Agonist 5-HT
Aricept (donepezil)	Blocks removal of ACh from synapse	Alzheimer's disease	Agonist for ACh
Revia (naltrexone)	Blocks opioid post-synaptic receptors	Alcoholism, opioid addiction	Antagonist (for opioids)
Adderall (mixed salts amphetamine)	Increase release of DA, NE	ADHD	Agonist for DA, NE
Ritalin (methylphenidate)	Blocks removal of DA, NE, and lesser (5-HT) from synapse	ADHD	Agonist for DA, NE mostly

6.1.3 Drug Classifications

Art in a Cup

Who knew that a cup of coffee could also be a work of art? A talented barista can make coffee look as good as it tastes. If you are a coffee drinker, you probably know that coffee can also affect your mental state. It can make you more alert and may improve your concentration. That's because the caffeine in coffee is a psychoactive drug. In fact, caffeine is the most widely consumed psychoactive substance in the world. In North America, for example, 90 percent of adults consume caffeine daily.

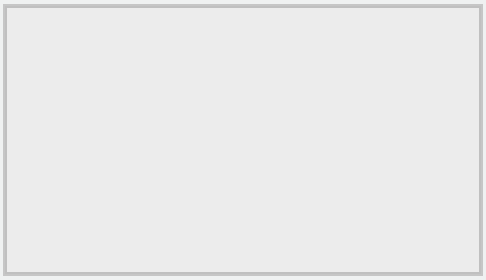


Figure **6.1.1**: Murano Coffee with decoration on the surface.

Besides caffeine, other examples of psychoactive drugs include anti-anxiety medications (Xanax, Valium), antidepressants (Prozac, Celexa), alcohol, tobacco, marijuana, cocaine, psilocybin, oxycodone, and morphine. Organizing the wide range of different drugs into distinct categories is no easy task. One of the challenges is their various and, sometimes overlapping, pharmacological effects and mechanisms of action. Another challenge is that psychoactive drugs may be used for a variety of purposes, including therapeutic, ritual, and/or recreational.

In addition, many of these drugs may be legal prescription medications (e.g., oxycodone and morphine), legal nonprescription drugs (e.g., alcohol and tobacco), or illegal drugs (cocaine and psilocybin). To further complicate matters, some of these drugs, legal or not can be both therapeutic and abused. The status of cannabis (or marijuana), for example, is in flux, at least in the United States. Depending on where you are, cannabis may be used recreationally and/or medically, and it may be either legal or illegal. Legal prescription medications, such as opioids are also used illegally by alarmingly large numbers of people, resulting in a tragically high number of overdose deaths.

Although classification of psychoactive drugs is clearly complicated, there are some general categorizations that can prove useful.

6.1.3.1 Classification Based on Psychopharmacological Effects

Psychoactive drugs can be divided into different classes according to their psychopharmacological effects. Several classes are listed below, along with examples of commonly used drugs in each class.

- ***Stimulants****:* Stimulate the brain and increase alertness and wakefulness. Examples of stimulants include caffeine, nicotine, cocaine, and amphetamines.
- ***Depressants****:* Calm the brain, reduce anxious feelings, and induce sleepiness. Examples of depressants include ethanol (in alcoholic beverages) and opioids such as oxycodone and heroin.
- ***Anxiolytics****:* Have a tranquilizing (calming) effect and inhibit anxiety. Examples of anxiolytic drugs include benzodiazepines such as diazepam (Valium), barbiturates such as phenobarbital, opioids, cannabis, and antidepressant drugs such as sertraline (Zoloft).
- ***Euphoriants****:* Cause a state of euphoria, or intense feelings of well-being and happiness. Examples of euphoriants include the so-called club drug MDMA (ecstasy), amphetamines, ethanol, and opioids such as morphine.
- ***Hallucinogens****:* Cause hallucinations and other perceptual anomalies. They also cause subjective changes in thoughts, emotions, and consciousness. Examples of hallucinogens include LSD, mescaline, nitrous oxide, and psilocybin.
- ***Empathogens****:* Produce feelings of empathy, or sympathy with other people. Examples of empathogens include MDMA (Ectasy).



Figure **6.1.2**: Ecstasy (MDMA) is most commonly taken in tablet form, like the colorful and decorated tablets shown

Many psychoactive drugs have multiple effects so they may be placed in more than one class. For example, many stimulants also have euphoriant properties, such as MDMA (Ectasy). Furthermore, MDMA may also act as an empathogen or hallucinogen. As of 2016, MDMA had no accepted medical uses, but is undergoing testing for use in the treatment of post-traumatic stress disorder and certain other types of anxiety disorders (Mitchell et al., 2021). As you can tell, drug classification can get a bit complicated.

6.1.3.2 Classification Based on Synaptic Mechanisms of Action

As previously stated, psychoactive drugs generally produce their effects by affecting brain chemistry, which in turn may cause changes in a person's mood, thinking, perception, and/or behavior. Each drug tends to have a specific action on one or more neurotransmitters or neurotransmitter receptors in the brain. Generally, they act as either agonists or antagonists.

- **Agonists** are drugs that mimic or increase the activity of particular neurotransmitters. They might act by promoting the synthesis of the neurotransmitters, reducing their reuptake from synapses, or mimicking their action by binding to receptors for the neurotransmitters.
- **Antagonists** are drugs that decrease the activity of particular neurotransmitters. They might act by interfering with the synthesis of the neurotransmitters or by blocking their receptors so the neurotransmitters cannot bind to them.

Consider the example of the neurotransmitter GABA. This is one of the most common neurotransmitters in the brain, and it normally has an inhibitory effect on cells. GABA *agonists*, which increase its effect at the synapse, include ethanol, depakote (anti-convulsant/bipolar medication), benzodiazepines (anti-anxiety medication), and other psychoactive drugs. All of these drugs work in different ways at the synapse, but ultimately increase the postsynaptic effect of GABA in the brain.

6.1.3.3 Classification Based on Type of Use

You may have been prescribed psychoactive drugs by your doctor. For example, you may have been prescribed a drug to treat anxiety or depression or an opioid, drug such as codeine for pain (most likely in the form of Tylenol with added codeine). You may

also use nonprescription psychoactive drugs, such as caffeine for mental alertness or cannabis (CBD or Marijuana) to treat pain or anxiety. These are just some of the many possible uses of psychoactive drugs.

6.1.3.4 Medical Uses



Figure **6.1.3**: This child is being prepared to receive a general anesthetic prior to surgery

Medical uses of psychoactive drugs include general anesthesia, in which pain is blocked and unconsciousness is induced. General anesthetics are most often used during surgical procedures and may be administered in gaseous form. General anesthetics include the drugs halothane and ketamine. Other psychoactive drugs are used to manage pain without affecting consciousness. They may be prescribed either for acute pain in cases of trauma such as broken bones; or for chronic pain such as pain caused by arthritis, cancer, or fibromyalgia. Most often, the drugs used for pain control are opioids, such as morphine and codeine. Their pain inhibitory actions rest with their ability to enhance our endogenous opioid activity in the brain, which ultimately will lead to a reduction of incoming pain signals from the body.

Many psychiatric disorders are also managed with psychoactive drugs. For example, antidepressants such as Sertraline or Celexa are used to treat depression, anxiety, and eating disorders. These drugs act as agonists for serotonin systems in key circuits of the brain which play an important role in mood regulation. Anxiety disorders may also be treated with anxiolytic drugs, such as buspirone and diazepam. Diazepam is from the benzodiazepine class of drugs which are agonists of the GABA system and can inhibit limbic areas of the brain involved with the anxiety and fear response.

Stimulants such as amphetamines are agonists for monamine transmitters and can be effective treatments for attention deficit disorder and certain sleep disorders. Antipsychotics such as clozapine and risperidone, as well as mood stabilizers such as lithium, are used to treat schizophrenia and bipolar disorder. These drugs act on a variety of transmitter systems in order to treat a variety of symptoms. Although controversial, relatively recent studies on the therapeutic, controlled use of hallucinogens such as MDMA and Psilocybin to treat disorders such as PTSD have been conducted.

6.1.3.5 Ritual Uses

Certain psychoactive drugs, particularly hallucinogens, have been used for ritual purposes since prehistoric times. For example, Native Americans have used the mescaline-containing peyote cactus (pictured below) for religious ceremonies for as long as 5,700 years. In prehistoric Europe, the mushroom *Amanita muscaria*, which contains a hallucinogenic drug called muscimol, was used for similar purposes. Various other psychoactive drugs — including jimsonweed, psilocybin mushrooms, and cannabis — have also been used by various peoples for ritual purposes for millennia.



Figure **6.1.4**: The peyote cactus contains a hallucinogenic drug that is still used by some Native Americans for religious rituals.

6.1.3.6 Recreational Uses

The most typical recreational uses of psychoactive drugs have the purpose of altering one's consciousness and creating a feeling of euphoria commonly called a "high." Some of the drugs used most commonly for these purposes include cannabis, ethanol, opioids, and stimulants such as nicotine, amphetamine, or cocaine. Hallucinogens are also used recreationally, primarily for the alterations in thinking and perception that they cause.

Some investigators have suggested that the urge to alter one's state of consciousness is a universal human drive, similar to the drive to satiate thirst, hunger, or sexual desire. They think that the drive to alter one's state of mind is even present in children, who may attain an altered state by repetitive motions such as spinning or swinging. Some nonhuman animals also exhibit a drive to experience altered states. For example, they may consume fermented berries or fruit and become intoxicated. The way cats respond to catnip (Figure **6.1.6**) is another example.

A variety of information on specific drugs used for recreational purposes is provided in section 6.3: Drugs of Abuse.



Figure **6.1.5**: This cat is holding a catnip toy and apparently enjoying its psychoactive effects.

6.1.4 Summary

It is probably clear now that psychoactive drugs can have profound and quite varied effects on how we think, feel, perceive the world, and behave. This range of effects depends on a multitude of factors including the particular neurotransmitter systems affected, the general pharmacology of the drug, and it's specific synaptic effects. As described, the ability for drugs to act as either agonists or antagonists at synapses and the role of receptor subtypes (i.e. excitatory or inhibitory) have a major impact on the ultimate outcome to the individual. These synaptic effects and the influence of physiological effects that occur prior to drugs reaching the brain will be explored further in the next section of this chapter.

6.1.5 Supplemental Resources

Learn more about psychiatric drugs that are being researched to treat mental health disorders. In this inspiring TED talk, neurobiologist David Anderson explains how modern psychiatric drugs treat the chemistry of the whole brain and why a more nuanced view of how the brain functions could lead to targeted psychiatric drugs that work better and avoid side effects.

6.1.6 Attribution

- Psychopharmacology and Psychoactive Drugs sections adapted from "NOBA - Psychology - Section 9.12, Psychopharmacology"; **License:** CC BY SA-NC 4.0 International
- Drug Classifications adapted from "Psychoactive Drugs" by Wakim and Grewal. **License:** CC BY-NC 4.0

6.1.7 References

- Mitchell, J.M., Bogenschutz, M., Lilienstein, A. *et al.* (2021) MDMA-assisted therapy for severe PTSD: a randomized, double-blind, placebo-controlled phase 3 study. *Nat Med* **27,** 1025–1033. https://doi.org/10.1038/s41591-021-01336-3

This page titled 6.1: Psychopharmacology and Psychoactive Drug Classification is shared under a mixed license and was authored, remixed, and/or curated by ASCCC OERI & Alan Keys (ASCCC Open Educational Resources Initiative (OERI)) .

- **11.8: Psychoactive Drugs** by Suzanne Wakim & Mandeep Grewal is licensed CK-12. Original source: https://www.ck12.org/book/ck-12-human-biology/.

6.2: Mechanisms of Psychoactive Drug Action

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe the key elements of pharmacokinetics and their relationship to drug action.
2. Compare and contrast the various drug administration methods in terms of potency, latency of action, and abuse potential.
3. Describe the key elements of pharmacodynamics, particularly drug effects on steps of synaptic transmission and major neural circuitry
4. Differentiate the major structures of the mesolimbic, "reward" pathway and the role of dopamine in its function.

6.2.1 Overview

Do drugs affect the body or does the body affect drugs? This may seem like either a very easy or a very odd question. Of course, the answer is drugs affect the body, right? For example, you take a certain drug and then your heart rate changes or you feel energized to move around a lot. In terms of psychoactive drugs, the brain, which is part of the body, is definitely affected by drugs. As we will see, neurotransmitters can be become more active and the receptors that these neurotransmitters bind to may trigger either increased or decreased communication with other neurons, glands, or muscles. These examples should make it clear that drugs definitely affect the body. However, even before these effects on the body occur, the *body* is able to affect incoming drugs in profound ways. Thus, the answer to the earlier question is that the body affects drugs and drugs affect the body.

This section will review the basics on how both of these processes occur with a general review of the concepts of **pharmacokinetics** (how the body affects drugs) and **pharmacodynamics** (how drugs affect the body, or more specifically the brain). In covering these topics we will address various physiological processes, such as drug metabolism and elimination and the influence of drug administration methods on a drug's impact. In addition, we will go into the specifics of how drugs interact with the process of synaptic communication and then zoom out to see how they alter the activity of major circuits and structures of the brain, such as the so-called, "reward pathway."

6.2.2 Pharmacokinetics: What Is It – Why Is It Important?

Pharmacokinetics refers to how the body processes drugs as they enter the body. While this section may sound more like pharmacology, it is important to realize how important pharmacokinetics can be when considering the ultimate effects of psychoactive drugs. As mentioned previously, psychoactive drugs exert their effects on behavior by altering neuronal communication in the brain, and the majority of drugs reach the brain by traveling in the blood.

The acronym **ADME** is often used to specify the processes of**:**

- **A - A**dministration **and A**bsorption (how the drug gets into the blood),
- **D - D**istribution (how the drug gets to the organ of interest – in this case, the brain),
- **M - M**etabolism (how the drug is broken down so it no longer exerts its psychoactive effects), and
- **E - E**limination (how the drug leaves the body).

We will focus on some of these processes to show their importance in determining the effects of psychoactive drugs.

6.2.2.1 Drug Administration and Absorption



Figure **6.2.1**: A drug delivered by IV reaches the brain more quickly than if the drug is taken orally. While rapid delivery has advantages, there are also risks involved with IV administration. (CC BY-SA 3.0; Calleamanecer via Wikimedia Commons)

Before a drug can be absorbed by the body it needs to be administered in some way (i.e., the drug has to get into the body). There are many ways to take drugs, and these routes of drug administration can have a significant impact on how quickly that drug reaches brain.

The most common route of administration is **oral administration,** which begins in the mouth and continues in the digestive system. This route is relatively **slow** and – perhaps surprisingly – often the most variable and complex route of administration. Drugs enter the stomach and then get absorbed by the blood supply and capillaries that line the small intestine. The rate of absorption can be affected by a variety of factors including the quantity and the type of food in the stomach (e.g., fats vs. proteins). This is why the medicine label for some drugs (like antibiotics) may specifically state foods that you should or should not consume within an hour of taking the drug because they can affect the rate of absorption.

Two of the most **rapid** routes of administration include **inhalation** (i.e., smoking or gaseous anesthesia) and **intravenous (IV)** in which the drug is injected directly into the vein and hence the blood supply. Both of these routes of administration can get the drug to brain in less than 10 seconds. IV administration also has the distinction of being the most dangerous because if there is an adverse drug reaction, there is very little time to administer any antidote, as in the case of an IV heroin overdose.

Key Elements of Drug Routes of Administration (Keys)

Route of Administration	Typical Methods Used	Speed of Absorption in Blood	Advantages / Disadvantages
Oral (PO)	Pill, Liquid, Edibles	**SLOWEST**	Convenient, Gradual onset, Relatively safe / Variable blood levels due to individual differences in absorption and metabolism
Subcutaneous	Injection into skin		Gradual onset, Avoids 1st pass metabolism / Only small volumes can be given, Possible localized irritation, tissue damage, infection
Intramuscular (IM)	Injection into muscle		Steady absorption rates, Avoids 1st pass metabolism, Extended release possible / Only small volumes can be given, Possible localized irritation, tissue damage, infection
Rectal	Rectal Suppository		Useful in those with digestive issues or unconscious, Mostly avoids 1st pass metabolism, Relatively high blood levels / Invasive, Unpredictable absorption rates
Sublingual	Under tongue		Direct & fast absorption, Avoids 1st pass metabolism, Quick termination possible
Inhalation	Smoke, Vape, Huff, Inhaler, Intranasal		Rapid absorption, Useful for emergency applications or localized effect (i.e. asthma), Reaches brain quickly / Variable dose & absorption, Relatively high abuse potential w/ certain drugs
Intravenous	Injected into vein	**FASTEST**	Very rapid absorption (~instantly), Avoids 1st pass metabolism, Useful for emergency applications, Accurate dosing / High abuse potential and risk for overdose with certain drugs

Why might knowing how quickly a drug gets to the brain be important? If we are considering therapeutic drugs, such as those used to treat anxiety, depression, or psychotic disorders, the time it takes to begin to relieve symptoms and how long this lasts can be critical information. For example, if treating an acute panic attack you would want a drug to act fast, whereas, if you want to prevent non-specific anxiety from occurring on a daily basis, you may use a different drug that acts more gradually but sticks around longer.

When considering potential drugs of abuse, administration and absorption rates are important for other reasons. Although there are multiple non-pharmacological risk factors for drug craving and substance use disorders, the phamacokinetics of a drug can play a

significant role. For example, if a drug activates the reward circuits in the brain and reaches these areas quickly, the drug may have a relatively high risk for abuse and psychological dependence. Psychostimulants like amphetamine or cocaine are examples of drugs that have high risk for abuse in part because they are agonists at dopamine neurons involved in reward and because these drugs exist in forms that can be either smoked or injected intravenously. Some argue that cigarette smoking is one of the hardest addictions to quit, and although part of the reason for this may be that smoking gets the nicotine into the brain very quickly (and indirectly acts on dopamine neurons), it is a more complicated story.

For drugs that reach the brain very quickly, not only is the drug potentially very addictive, but so are the cues associated with the drug (see Rohsenow, Niaura, Childress, Abrams, and Monti, 1990). A learned connection between drug use and the environmental elements that accompany it can complicate quitting by triggering symptoms of withdrawal and/or cravings. For an IV opioid user, this could be the sight of a container in which the drug and syringes are typically stored or other aspects of their typical drug-taking environment. For a cigarette smoker, it could be the smell of another person's smoke or even something as typical as finishing dinner or waking up in the morning (if that is when the smoker usually has a cigarette). For both the opioid user and the cigarette smoker, the cues associated with the drug may actually cause craving that is alleviated by (you guessed it) – lighting a cigarette or injecting the opioid (i.e., relapse). This is one of the reasons individuals that enroll in drug treatment programs, especially out-of-town programs, are at significant risk of relapse if they later find themselves in proximity to old "environments", including friends and situations associated with drug use. But this is much *more* difficult for a cigarette smoker. How can someone avoid eating? Or avoid waking up in the morning, etc. These examples help you begin to understand how important the route of administration can be for psychoactive drugs.

6.2.2.2 Drug Metabolism

Metabolism involves the breakdown of psychoactive drugs, and this occurs primarily in the liver. The liver produces **enzymes** (proteins that speed up a chemical reaction), and these enzymes help catalyze a chemical reaction that breaks down psychoactive drugs. Enzymes exist in “families,” and many psychoactive drugs are broken down by the same family of enzymes, the Cytochrome P450 super family. Usually, there is not a unique enzyme for each drug; rather, certain enzymes can break down a wide variety of drugs. **Tolerance** to the effects of many drugs can occur with repeated exposure; that is, the drug produces less of an effect over time, so more of the drug is needed to get the same effect. This is particularly true for sedative drugs like alcohol or opiate-based painkillers (e.g., fentanyl, codeine). *Metabolic tolerance* is one kind of tolerance and it takes place primarily in the liver. Some drugs, such as alcohol cause **enzyme induction** – an increase in the enzymes produced by the liver. For example, chronic drinking results in alcohol being broken down more quickly, so the alcoholic needs to drink more to get the same effect – of course, until so much alcohol is consumed that it damages the liver (i.e. fatty liver or cirrhosis).

There are a variety of other ways that enzyme action can be altered by both physiological and environmental aspects of drug taking. For example, just the expectation of drug taking or the cues from the environment where drug taking regularly occurs can trigger either enzyme induction or inhibition and, thus altering the ultimate physical and/or psychological effects of the drug. These differences in response can be particularly dangerous if they involve an unexpected increase in life-threatening side effects, such as the respiratory depression seen in response to opioids.

After "surviving" the processes of absorption and initial metabolism but before being eliminated, drugs are distributed throughout the body until they reach their target. In the case of psychoactive drugs, the main target is the brain. At this point, the process shifts to *pharmacodynamics* where the roles reverse and drugs are now able to affect the body.

6.2.3 Pharmacodynamics

If you have a headache or muscle pain, you might take a drug like ibuprofen (Advil, Motrin) or acetaminophen (Tylenol). Part of how these drugs work is to block the production of chemicals that cause the inflammation that is responsible for the pain. Thus, their target is the location of action is where the inflammation or pain is occurring. For psychoactive drugs, including those which treat pain the target of action for their effects is the brain or nervous system. The pharmacodynamics of their effects can be divided into two general categories: Their influence on specific steps of communication at the level of the synapse and the more holistic effect on the activity of certain circuits or major structures of the nervous system.

Just about all psychoactive drugs have their initial effects on one or more of the **steps of synaptic communication: neurotransmitter synthesis, storage, release, binding, reuptake, and breakdown (See Figure 6.2.2)**. Remember, drugs can act

as either agonists or antagonists of a neurotransmitter's typical effects. The drug L-dopa, for example acts as an agonist to increase the synthesis of dopamine and treat the dopamine dysfunction present in Parkinson's disease. Other agonist drugs like amphetamine and some other stimulants can increase the amount of dopamine activity by enhancing the release step. And yet others, such as cocaine can increase the synaptic activity of dopamine by blocking its reuptake. All these drugs are dopamine agonists (increase dopamine synaptic activity), but they can have very different psychological and behavioral effects due to the difference in the synaptic step targeted.

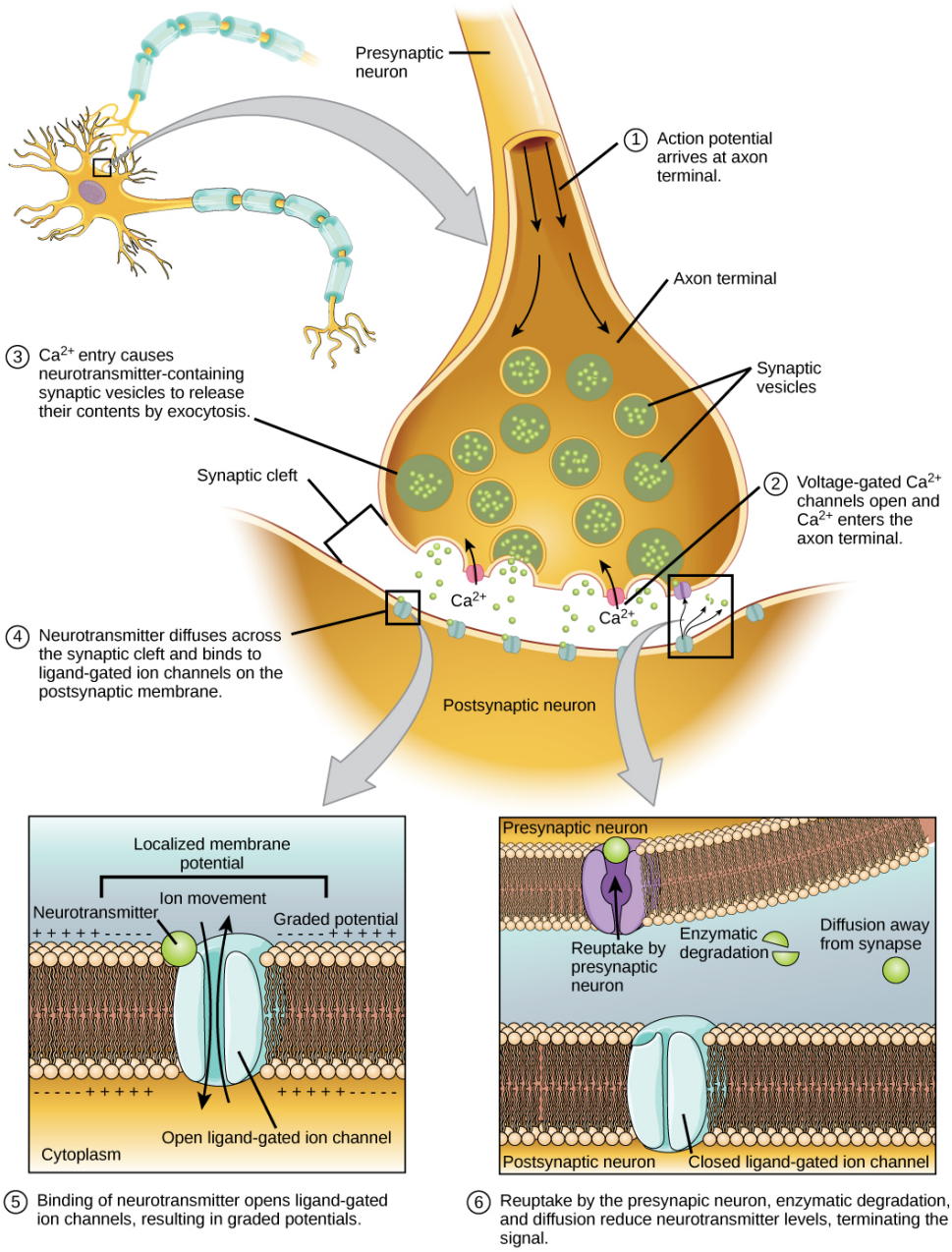


Figure **6.2.2**: Steps in Synaptic Transmission starting with the arrival of the action potential and culminating in transmitter release into the synapse. Details of neurotransmitter binding and termination mechanisms in the synapse are included in insets. Psychoactive drugs can affect one or more of these steps, depending on the drug. (CC BY 4.0 by OpenStax)

Knowing these initial actions of drugs at the level of the synapse is important in order to understand how they ultimately affect our thinking, feeling, and behavior. However, the thousands or even millions of synaptic changes that occur every split second due to drug action first need to be integrated in order to have noticeable effects. This integration is reflected in alterations of the activity of key structures and circuits of the nervous system.

6.2.4 Drug effects on Neural Structures and Circuits

Every distinct brain area or neural circuit can be affected in some way by the action of psychoactive drugs due to the integration process discussed above. For example, drugs that increase endorphin activity (i.e. Opioids) in certain areas of the brain can trigger changes in a major pain-inhibition pathway. Circuits involved with attention are affected by drugs that enhance norepinephrine systems and can be effective treatments for ADHD. Still other drugs, such as anti-anxiety medications (i.e. Xanax, Valium) are GABA agonists and can enhance the action of GABA throughout the brain, including the amygdala, which plays a significant role in fear and anxiety responses.

6.2.4.1 Mesolimbic "Reward" Circuitry

One of the most interesting and significant circuits that has been studied for its role in drug taking and substance use disorder is the "reward circuit" or "pleasure pathway" of the brain. This circuitry is technically called the mesolimbic circuitry and contains major dopamine pathways in addition to other transmitters which play a role in the positive feelings associated with both natural and drug rewards (See Figure 6.2.3).

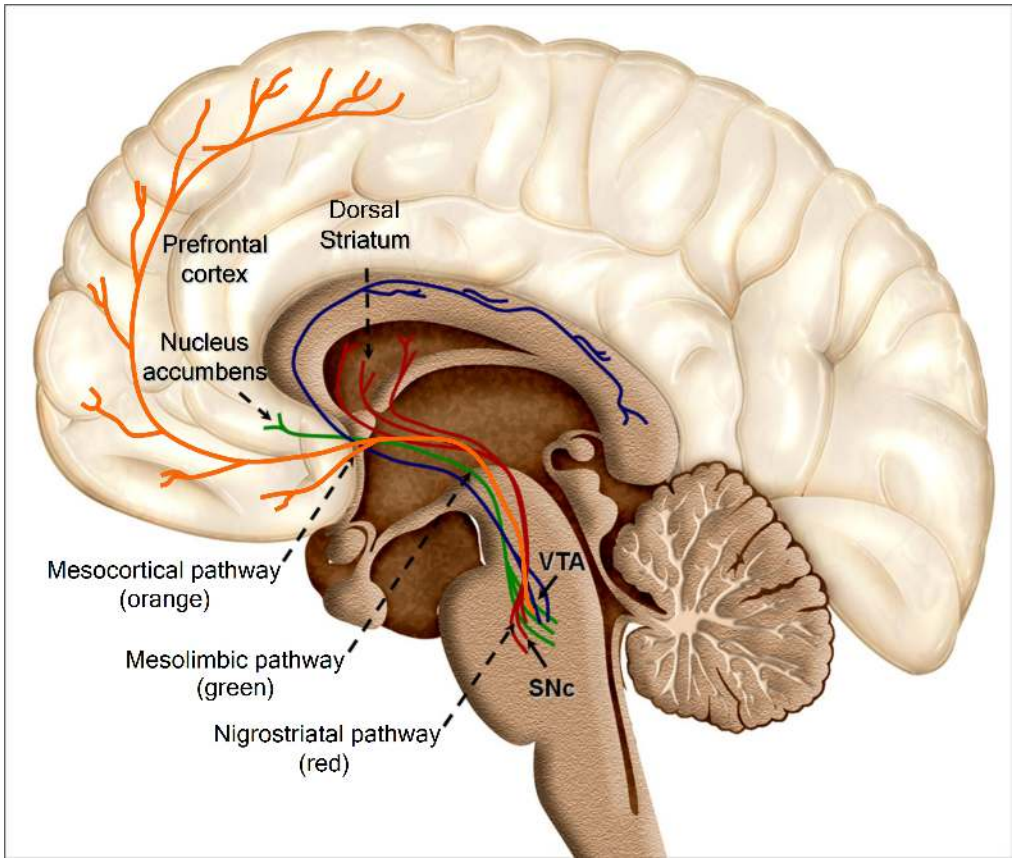


Figure **6.2.3**: Mesolimbic, "Reward" Circuit and other Dopamine Pathways. These include pathways from both the Substantia Nigra (SNc) and Ventral Tegmental Area (VTA) to forebrain targets such as the Nucleus Accumbens and Prefrontal Cortex. From: Oscar Arias-Carrión1, Maria Stamelou, Eric Murillo-Rodríguez, Manuel Menéndez-González and Ernst Pöppel.Substantially modified by Seppi333, CC BY 2.0 , via Wikimedia Commons

The nucleus accumbens definitely plays a central role in the **reward circuit.** Its operation is based chiefly on two essential neurotransmitters: dopamine, which promotes desire, and serotonin, whose effects include satiety and inhibition. Many animal studies have shown that many psychoactive drugs and natural rewards increase the production of dopamine in the nucleus accumbens, while reducing that of serotonin.

But the nucleus accumbens does not work in isolation. It maintains close relations with other areas involved in pleasure and reward. One area in particular is the **ventral tegmental area (VTA)**. Located in the midbrain, at the top of the brainstem, the VTA is one of the most primitive parts of the brain. Neurons of the VTA synthesize dopamine and then send it via their axons primarily to the

nucleus accumbens. The VTA is also influenced by endorphins whose receptors are targeted by opiate drugs such as heroin and morphine.

Another structure involved in pleasure mechanisms is the **prefrontal cortex**, whose role in planning and motivating action is well established. The prefrontal cortex is a significant relay in the reward circuit and also is modulated by dopamine.

The locus coeruleus, an alarm area of the brain and packed with norepinephrine, is another brain structure that plays an important role in drug abuse. When stimulated by a lack of the drug in question, the locus coeruleus drives the user to do anything necessary to obtain more drug.

Three other structures in the limbic system also play an active part in the pleasure circuit and, consequently, in drug dependency. The first is the **amygdala**, which can provide affective or mood information in response to drugs or environmental stimuli.

The second is the **hippocampus**, which preserves the agreeable memories associated with drug taking or other non-drug behaviors and, by association, all of the details of the environment in which these behaviors occur. The memory of these details may trigger positive feelings and, in the case of drugs reawaken the desire to take the drug again contributing to the possibility of relapse.

The third structure, the most anterior portion of the **insular cortex**, or insula, is regarded as part of the limbic system that is thought to play a role in the active pleasure-seeking associated both with food and with psychoactive substances. It has been proposed that this part of the cortex tells us about the bodily states associated with our emotional experiences and then relates these feeling states to cognitive processes such as decision-making (Damasio et al., 2013).

6.2.5 Summary

Psychoactive drugs clearly undergo quite a journey through the body and, ultimately the brain prior to causing changes in how we think, feel, perceive the world, and behave. The pharmacokinetic steps of administration, distribution, metabolism, and elimination have a major impact on the timing and amount of drug action prior to their arrival at their targets in the nervous system. Upon arrival, pharmacodynamic actions primarily at the synapse contribute to the diverse and dynamic action of drugs on neural communication and subsequent psychological and behavioral changes. Psychoactive drugs have the potential to affect structures in all parts of the nervous system as long as the appropriate receptors are there to receive the. The structures of the Mesolimbic Reward Circuitry play a particularly critical role in the action of many psychoactive drugs, both therapeutic and recreational.

6.2.6 Attributions

- Pharmacokinetics adapted from ""NOBA - Psychology - Section 9.12, Psychopharmacology." **License:** CC BY SA-NC 4.0 International

6.2.7 References

- Damasio A, Damasio H, Tranel D. (2013) Persistence of feelings and sentience after bilateral damage of the insula. *Cereb Cortex.* 23(4):833-46. doi: 10.1093/cercor/bhs077. PMID: 22473895.
- Rohsenow DJ, Niaura RS, Childress AR, Abrams DB, Monti PM. (1990-1991) Cue reactivity in addictive behaviors: theoretical and treatment implications. *Int J Addict.* 25(7A-8A):957-93. DOI: 10.3109/10826089109071030. PMID: 2131326.

This page titled 6.2: Mechanisms of Psychoactive Drug Action is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

6.3: Substance Use and Addiction - Theories, Brain Mechanisms, and Treatment

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Define addiction (substance use disorder) in terms of changes in behavior, emotion, and cognition.
2. Describe the neural mechanisms of addiction, including the role of key neurotransmitters (dopamine & endorphins), brain structures & circuits, and genetics.
3. Differentiate the actions of drugs of abuse in terms of their neurotransmitter effects, neuroanatomy of action, and abuse potential

6.3.1 Overview

Many people don't understand why or how other people become addicted to drugs. They may mistakenly think that those who use drugs lack moral principles or willpower and that they could stop their drug use simply by choosing to. In reality, drug addiction is a complex disease, and quitting usually takes more than good intentions or a strong will. Drugs change the brain in ways that make quitting very difficult, even for those who want to. Fortunately, researchers know more than ever about how drugs affect the brain and have found treatments that can help people recover from drug addiction and lead productive lives.

6.3.2 What Is Addiction?

According to the National Institute of Drug Abuse (NIDA), drug addiction is defined as the continued compulsive use of drugs despite adverse health or social consequences. People addicted to drugs have lost control of their drug use. Because of this they can become isolated from family or friends, have difficulty at work or school, engage in unhealthy behaviors, and may become involved with the criminal justice system. For a person addicted to drugs, continuing to take them becomes the primary focus in life.

The Diagnostic and Statistical Manual, 5th edition (DSM V) from the American Psychological Association has reframed addiction as **substance use disorder** (Hasin, et al., 2013). This diagnostic category views substance use on a continuum with a range of severity from mild to severe based on the presence of 11 criteria. These criteria include some of the same factors noted by the NIDA definition stated above in addition to others, such as physiological effects (i.e. withdrawal & tolerance) and psychological problems related to drug use.

Disagreements about the nature of addiction remain: namely, whether it reflects voluntary or involuntary behavior and whether it should be punished or treated as a health issue. Even though the first time a person takes a drug, it is often by choice—to achieve a pleasurable sensation or desired emotional state—we now know from a large body of research that this ability to choose can be affected by drugs. And when addiction takes hold in the brain, it disrupts a person's ability to exert control over behavior—reflecting the compulsive nature of this disease.

The human brain is an extraordinarily complex and fine-tuned communications network made up of billions of cells that govern our thoughts, emotions, perceptions, and drives. Our brains reward certain behaviors such as eating or procreating—registering these as pleasurable activities that we want to repeat. Drug addiction taps into these vital mechanisms geared for our survival. And although not a life necessity, to an addicted person, drugs become life itself, driving the compulsive use of drugs—even in the face of dire life consequences—*that is the essence of addiction*.

6.3.3 How Does Addiction Take Hold in the Brain?

Decades of research supported by NIDA has proven that addiction is a complex brain disease characterized by compulsive, at times uncontrollable, drug craving, seeking, and use that persist despite potentially devastating consequences. Although not the only factor, the powerful rewarding effects of most drugs of abuse play a major role in their continued use despite negative effects.

The **rewarding effects** of most drugs of abuse come mainly from large and rapid upsurges in **dopamine**, a neurochemical critical to stimulating feelings of pleasure and to motivating behavior. Increased activity of **endorphins**, our natural opiate system can also play a role in these effects. The rapid dopamine "rush" from drugs of abuse mimics but greatly exceeds in intensity and duration the feelings that occur in response to such pleasurable stimuli as the sight or smell of food, for example. Repeated exposure to large, drug-induced dopamine surges has the insidious consequence of ultimately blunting the response of the dopamine system to everyday stimuli. Thus the drug disturbs a person's normal hierarchy of needs and desires and substitutes new priorities concerned with procuring and using the drug.

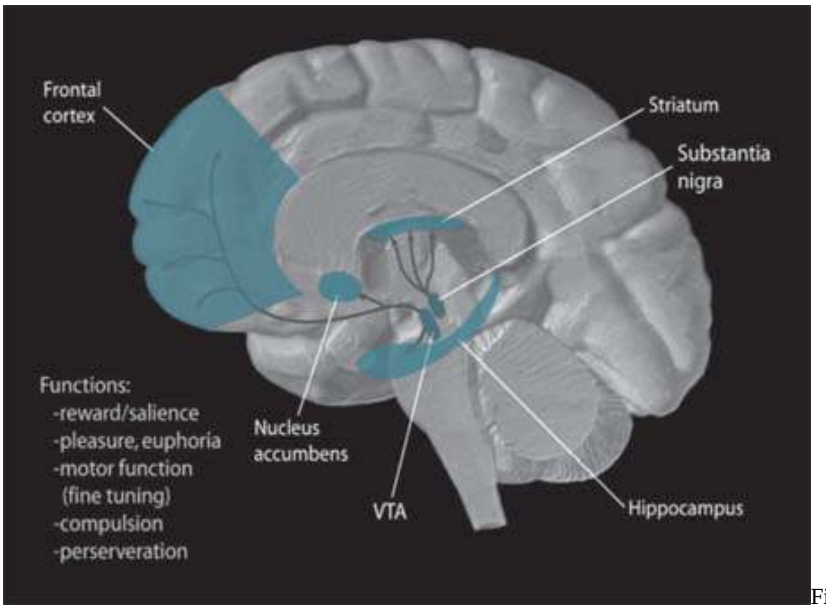


Figure **6.3.1**: Dopamine Pathways. In the brain, dopamine plays an important role in the regulation of reward and movement. As part of the reward pathway, dopamine is manufactured in nerve cell bodies located within the ventral tegmental area (VTA) and is released in the nucleus accumbens and the prefrontal cortex. Its motor functions are linked to a separate pathway, with cell bodies in the substantia nigra that manufacture and release dopamine into the striatum. From: NIDA Research Report Series - Methamphetamine Abuse and Addiction (2006), PD US NIH, via Wikimedia Commons

Drug abuse also disrupts the brain circuits involved in memory and control over behavior. Memories of the drug experience can trigger craving as can exposure to people, places, or things associated with former drug use. Stress is also a powerful trigger for craving. Control over behavior is compromised because the affected frontal brain regions are what a person needs to exert inhibitory control over desires and emotions. That is why addiction is a brain disease.

As a person's **reward circuitry** becomes increasingly dulled and **desensitized** by drugs, nothing else can compete with them—food, family, and friends lose their relative value, while the ability to curb the need to seek and use drugs evaporates. Ironically and cruelly, eventually even the drug loses its ability to reward, but the compromised brain leads addicted people to pursue it, anyway; the memory of the drug has become more powerful than the drug itself.

Certain drugs, including **opioids and alcohol**, cause **strong physical reactions** in the body when drug use stops. When a person addicted to heroin stops taking heroin, they can experience a variety of symptoms ranging from watery eyes and a runny nose to irritability and loss of appetite and then diarrhea, shivering, sweating, abdominal cramps, increased sensitivity to pain, and sleep problems. In general, **withdrawal** from heroin makes people feel miserable. Withdrawal from alcohol can cause serious effects such as seizures and even death. Withdrawal from other drugs, such as cocaine and amphetamines, does not typically lead to strong physical reactions, but it may make the person feel depressed or lethargic. For most drugs, physical withdrawal symptoms can usually be controlled effectively with medications.

Even though withdrawal from some drugs does not cause the person abusing them to have physical reactions, stopping drug use is difficult because of the changes the drugs have caused in the brain. Once the drugs stop, the person will have **cravings**, or intense desire for the drugs that may lead to **relapse** of drug taking. Craving arises from the brain's need to maintain a state of homeostasis (balance) that now relies on the presence of the drug. A person may experience cravings at any stage of drug abuse or addiction, even early in the experimentation phase of drug abuse. Cravings have been shown to have a physical basis in the brain. Using PET

imaging, scientists have shown that just seeing images of drug paraphernalia can stimulate the **amygdala** (part of the brain involved in emotional memory) in an addicted person.

Drugs of abuse do not merely cause short-term changes in an individual's cognitive skill and behavior. A drug "**high**" (i.e. euphoric state) typically lasts a short time, ranging from less than an hour to 12 hours, depending on the drug, dose, and route of administration. The changes in the brain that result from continued drug use, however, can last a long time. Scientists believe that some of these changes disappear when drug use stops; some disappear within a short time after drug use stops, and other changes are potentially permanent.

One of the first changes in the brain that may occur in response to repeated drug abuse is **tolerance**. Tolerance develops when a person needs increasing doses of a drug to achieve the same high or "rush" that previously resulted from a lower dose of the drug. Two primary mechanisms underlie the development of tolerance. First, the body may become more efficient at metabolizing the drug, thereby reducing the amount that enters the brain. Second, the cells of the body and brain may become more resistant to the effect of the drug. For example, after continued cocaine use, neurons decrease the number of dopamine receptors, which results in decreasing cocaine's stimulatory effect. Opioids, on the other hand, do not typically cause a change in the number of receptors. Instead the opioid receptors become less efficient in activating associated cellular processes, thus reducing the effects of the opioids.

6.3.4 Why do some people become addicted to drugs while others don't?

No one factor can predict if a person will become addicted to drugs. A combination of factors influences risk for addiction. The more risk factors a person has, the greater the chance that taking drugs can lead to addiction. For example:

- **Biology**. The genes that people are born with account for about half of a person's risk for addiction. Gender, ethnicity, and the presence of other mental disorders may also influence risk for drug use and addiction.
- **Environment**. A person's environment includes many different influences, from family and friends to economic status and general quality of life. Factors such as peer pressure, physical and sexual abuse, early exposure to drugs, stress, and socioeconomic pressures can greatly affect a person's likelihood of drug use and addiction.
- **Development**. Genetic and environmental factors interact with critical developmental stages in a person's life to affect addiction risk. Although taking drugs at any age can lead to addiction, the earlier that drug use begins, the more likely it will progress to addiction. This is particularly problematic for teens. Because areas in their brains that control decision-making, judgment, and self-control are still developing, teens may be especially prone to risky behaviors, including trying drugs.

6.3.5 Can drug addiction be treated?

As with most other chronic diseases, such as diabetes, asthma, or heart disease, treatment for drug addiction generally isn't a cure. However, addiction is treatable and can be successfully managed. People who are recovering from an addiction will most likely be at risk for relapse for years and possibly their whole lives. Research shows that combining addiction treatment medicines with behavioral therapy ensures the best chance of success for most patients. Treatment approaches tailored to each patient's drug use patterns and any co-occurring medical, mental, and social problems can lead to continued recovery.

More good news is that drug use and addiction are preventable. Results from NIDA-funded research have shown that prevention programs involving families, schools, communities, and the media are effective for preventing or reducing drug use and addiction. Although personal events and cultural factors affect drug use trends, when young people view drug use as harmful, they tend to decrease their drug taking. Therefore, education and outreach are key in helping people understand the possible risks of drug use. Teachers, parents, and health care providers have crucial roles in educating young people and preventing drug use and addiction.

6.3.5.1 Medications and Devices used in Drug Addiction Treatment

Medications and devices can be used to manage withdrawal symptoms, prevent relapse, and treat co-occurring conditions.

Withdrawal. Medications and devices can help suppress withdrawal symptoms during detoxification. Detoxification is not in itself "treatment," but only the first step in the process. Patients who do not receive any further treatment after detoxification usually resume their drug use. One study of treatment facilities found that medications were used in almost 80 percent of detoxifications (SAMHSA, 2014). In November 2017, the Food and Drug Administration (FDA) granted a new indication to an electronic stimulation device, NSS-2 Bridge, for use in helping reduce opioid withdrawal symptoms. This device is placed behind the ear and

sends electrical pulses to stimulate certain brain nerves. Also, in May 2018, the FDA approved lofexidine, a non-opioid medicine designed to reduce opioid withdrawal symptoms.

Relapse prevention. Patients can use medications to help re-establish normal brain function and decrease cravings. Medications are available for treatment of opioid (heroin, prescription pain relievers), tobacco (nicotine), and alcohol addiction. Scientists are developing other medications to treat stimulant (cocaine, methamphetamine) and cannabis (marijuana) addiction. People who use more than one drug, which is very common, need treatment for all of the substances they use.

- ***Opioids:*** Methadone (Dolophine®, Methadose®), buprenorphine (Suboxone®, Subutex®, Probuphine® , Sublocade™), and naltrexone (Vivitrol®) are used to treat opioid addiction. Acting on the same targets in the brain as heroin and morphine, methadone and buprenorphine suppress withdrawal symptoms and relieve cravings. Naltrexone blocks the effects of opioids at their receptor sites in the brain and should be used only in patients who have already been detoxified. All medications help patients reduce drug seeking and related criminal behavior and help them become more open to behavioral treatments. A NIDA study found that once treatment is initiated, both a buprenorphine/naloxone combination and an extended release naltrexone formulation are similarly effective in treating opioid addiction. Because full detoxification is necessary for treatment with naloxone, initiating treatment among active users was difficult, but once detoxification was complete, both medications had similar effectiveness.
- ***Tobacco:*** Nicotine replacement therapies have several forms, including the patch, spray, gum, and lozenges. These products are available over the counter. The U.S. Food and Drug Administration (FDA) has approved two prescription medications for nicotine addiction: bupropion (Zyban®) and varenicline (Chantix®). They work differently in the brain, but both help prevent relapse in people trying to quit. The medications are more effective when combined with behavioral treatments, such as group and individual therapy as well as telephone quitlines.
- ***Alcohol:*** Three medications have been FDA-approved for treating alcohol addiction and a fourth, topiramate, has shown promise in clinical trials (large-scale studies with people). The three approved medications are as follows:
  - **Naltrexone** blocks opioid receptors that are involved in the rewarding effects of drinking and in the craving for alcohol. It reduces relapse to heavy drinking and is highly effective in some patients. Genetic differences may affect how well the drug works in certain patients.
  - **Acamprosate (Campral®)** may reduce symptoms of long-lasting withdrawal, such as insomnia, anxiety, restlessness, and dysphoria (generally feeling unwell or unhappy). It may be more effective in patients with severe addiction.
  - **Disulfiram (Antabuse®)** interferes with the breakdown of alcohol. Acetaldehyde builds up in the body, leading to unpleasant reactions that include flushing (warmth and redness in the face), nausea, and irregular heartbeat if the patient drinks alcohol. Compliance (taking the drug as prescribed) can be a problem, but it may help patients who are highly motivated to quit drinking.
- ***Co-occuring conditions:*** Other medications are available to treat possible mental health conditions, such as depression or anxiety, that may be contributing to the person’s addiction.

More details on the addictive potential for some of the specific drugs mentioned above and others are included in the next section.

6.3.6 Summary

Drug addiction (i.e. Substance Use Disorder) is clearly a chronic disease characterized by drug seeking and use that is compulsive, or difficult to control, despite harmful consequences. Brain changes that occur over time with drug use challenge an individual's self-control and interfere with their ability to resist intense urges to take drugs, thus relapse is common. Most drugs affect the brain’s reward circuit by flooding it with the chemical messenger dopamine. This causes the “high” that leads people to take a drug again. Over time, the brain adjusts to excess dopamine activity, which ultimately results in the effect known as tolerance, leading the user to take more of the drug in order to achieve the same high. Typically, a combination of genetic, environmental, and developmental factors influences risk for drug misuse. The more risk factors a person has, the greater the chance that taking drugs can lead to a substance use disorder. However, there is hope in that substance use disorder is both preventable and treatable.

6.3.7 Attributions

- "Overview," "What Is Addiction?," "How Does Addiction Take Hold in the Brain," "Why do Some People Become Addicted to Drugs and Others Don't?," and "Can Addiction be Treated"- adapted from "Understanding Drug Use and Addiction." **Authored**

by: National Institute on Drug Abuse. **Provided by**: National Institutes of Health. **License**: *Public Domain: No Known Copyright*

- "Medications and Devices used in Drug Addiction Treatment" adapted from Drugs, Brain, and Behavior: The Science of Addiction. **Authored by**: National Institute on Drug Abuse. **Provided by**: National Institutes of Health. **License**: *Public Domain: No Known Copyright*
- Contemporary Health Issues - Section 8.1. **Authored by**: Lumen. **License**: *Public Domain: No Known Copyright*
- Mesolimbic, "Reward" pathway from Duboc, Brian, "The Brain from Top to Bottom," Canadian Institutes of Health Research: **Institute of Neurosciences, Mental Health and Addiction. License: CopyLeft**

6.3.8 References

- Hasin, D. S., O'Brien, C. P., Auriacombe, M., Borges, G., Bucholz, K., Budney, A., Compton, W. M., Crowley, T., Ling, W., Petry, N. M., Schuckit, M., & Grant, B. F. (2013). DSM-5 criteria for substance use disorders: recommendations and rationale. *The American journal of psychiatry*, *170*(8), 834–851. https://doi.org/10.1176/appi.ajp.2013.12060782

This page titled 6.3: Substance Use and Addiction - Theories, Brain Mechanisms, and Treatment is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

6.4: Specific Drug Effects, Misuse, and Risks

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Identify the specific types of drugs which are commonly misused and abused
- Describe the desired and adverse effects of commonly abused drugs
- Differentiate the potential contributing factors for drug misuse and abuse of commonly abused drugs.

6.4.1 Overview

A wide variety of drugs both legal and illegal can lead to misuse and addiction due to several possible causal factors. As noted previously, not everyone who takes these drugs will ultimately end up with a substance use disorder. Some, in fact are used therapeutically with success or recreationally without negative consequences. However, it is also true that some pose an unusually high risk for developing abusive patterns and significant adverse physical, psychological, and behavioral effects. This section will highlight some of the most prevalent and risky drugs in terms of their potential for being abused. This selection of potential drugs of abuse and their associated effects is just a sample of the wide array of chemical substances that can result in abusive patterns of intake. More information on the effects of the drugs presented here along with several other potential drugs of abuse can be found in the links to the National Institute of Drug Abuse sites provided at the end of this section.

6.4.2 Prescription Drugs

Prescription drugs are often strong medications, which is why they require a prescription from a doctor or dentist. Listed below are three common types of prescription drugs along with their primary neural mechanisms of action and clinical uses.

- **Opioids**— Acts on the **endorphin** neurotransmitter system and is used to relieve pain
- **Depressants**—Acts primarily on **inhibitory** mechanisms, such as the **GABA** system and is used to relieve anxiety or help a person sleep
- **Stimulants**— Acts primarily on **excitatory** mechanisms, such as **dopamine and norepinephrine** and is used to treat attention deficit hyperactivity disorder (ADHD), narcolepsy, and other disorders.

6.4.2.0.1 *How Prescription Drugs are Misused*

Along with their clinical uses, all three of the drug types stated above are commonly abused. In the initial stages of drug misuse, people tend to take these drugs to either get “high” or achieve a state of altered consciousness. Prescription drug misuse has become a very significant public health problem primarily due to the risks and adverse effects of prescription opioids which has contributed to a recent dramatic increase in overdose deaths.

Some of the most common ways that prescription drugs are misused include the following:

- Taking someone else’s prescription medication, even if it is for a medical reason (such as to relieve pain, to stay awake, or to fall asleep).
- Taking a prescription medication (yours or someone else's) in a way other than prescribed—for instance, taking more than the prescribed dose or taking it more often, or crushing pills into powder to snort or inject the drug.
- Mixing it with alcohol or certain other drugs.

All three of these methods have an inherent set of both psychological and physiological risks which contribute to their likelihood of being abused and adverse reactions.

6.4.2.0.1 *What Makes Prescription Drug Misuse Unsafe*

Every medication has some risk for harmful effects, sometimes serious ones. Doctors and dentists consider the potential benefits and risks to each patient before prescribing medications and take into account a lot of different factors, as described below:

- **Personal information.** Before prescribing a drug, health providers consider a person's weight, how long they've been prescribed the medication, other medical conditions, and what other medications they are taking. Someone misusing prescription drugs may overload their system or put themselves at risk for dangerous drug interactions.
- **Form and dose.** Doctors know how long it takes for a pill or capsule to dissolve in the stomach, release drugs to the blood, and reach the brain. When misused, prescription drugs are sometimes taken in larger amounts or in ways that change the way the drug works in the body and brain, putting the person at greater risk for an overdose. For example, when people who misuse OxyContin® crush and inhale the pills, a dose that normally works over the course of 12 hours hits the central nervous system all at once. This effect increases the risk for addiction and overdose.
- **Side effects.** Prescription drugs are designed to treat a specific illness or condition, but they often affect the body in other ways, some of which can be uncomfortable, and in some cases, dangerous. These are called side effects. Side effects can be worse when prescription drugs are not taken as prescribed or are used in combination with other substances. Side effects for the most commonly abused prescription drugs include the following:
    - Using **opioids** like oxycodone and codeine can cause you to feel sleepy, sick to your stomach, and constipated. At higher doses, opioids can make it hard to breathe properly and can cause overdose and death.
    - Using **stimulants** like Adderall or Ritalin can make you feel paranoid. It also can cause your body temperature to get dangerously high and make your heart beat too fast. This is especially likely if stimulants are taken in large doses or in ways other than swallowing a pill.
    - Using **depressants** like barbiturates can cause slurred speech, shallow breathing, sleepiness, disorientation, and lack of coordination. People who misuse depressants regularly and then stop suddenly may experience seizures. At higher doses depressants can also cause overdose and death, especially when combined with alcohol.

6.4.2.0.1 *Overdose Risk*

More than half of the drug overdose deaths in the United States each year are caused by prescription drug misuse. Overdose deaths involving prescription drugs—including pain relievers, benzodiazepines and antidepressants—increased steadily throughout the 1990s, peaked in 2017 and then decreased steadily in 2018 and 2019, and then increased again in 2020. Increases were linked to a rise in the misuse of prescription opioid pain relievers, as well as the presence of fentanyl in the drug supply. In 2013, only 9% of deaths (1,630 deaths) involving prescription drugs also involved fentanyl. In 2019, more than 46% of deaths (10,400 deaths) involving prescription drugs also involved fentanyl.

Mixing different types of prescription drugs can be particularly dangerous. For example, benzodiazepines interact with opioids (pain relievers) and increase the risk of overdose. Also, combining opioids with alcohol can make breathing problems worse and can lead to death.

6.4.2.1 **Alcohol**

Alcohol is among the most used drugs, plays a large role in many societies and cultures around the world, and greatly impacts public health. More people over age 12 in the United States have used alcohol in the past year than any other drug or tobacco product, and alcohol use disorder is the most common type of substance use disorder in the United States.

6.4.2.1.1 *Alcohol Effects on the Brain and Behavior*

Alcohol is a potent depressant whose main mechanism of action is to increase inhibitory action throughout the nervous system. It's primary mode of action to accomplish this is by acting as an agonist in the GABA neurotransmitter system. This can then act to release people from inhibitions, which gives the initial appearance of alcohol being excitatory.

When people drink alcohol, they may temporarily feel elated and happy, but they should not be fooled. As blood alcohol level rises, the effects on the body—and the potential risks—multiply.

- Inhibitions and memory become affected, so people may say and do things that they will regret later and possibly not remember doing at all.
- Decision-making skills are affected, so people may be at greater risk for driving under the influence or making unwise decisions.
- Aggression can increase, potentially leading to everything from verbal abuse to physical fights.

- Coordination and physical control are also impacted. When drinking leads to loss of balance, slurred speech, and blurred vision, even normal activities can become more dangerous.

Research also suggests that drinking during the teen years could interfere with normal brain development and change the brain in various ways, including negative effects on information processing and learning and increased risk of developing alcohol use disorder later in life.

6.4.2.1.2 *Adverse Effects and Overdose Risk*

Alcohol use disorder (AUD) is a chronic relapsing brain disorder characterized by an impaired ability to stop or control alcohol use despite adverse social, occupational, or health consequences. AUD ranges from mild to severe.

Consuming a dangerously high amount of alcohol can also lead to alcohol overdose and death. When people drink too much, they may eventually pass out (lose consciousness). Reflexes like gagging and breathing can be suppressed. That means people who have had too much alcohol could vomit and choke, or just stop breathing completely. Vulnerability to overdose increases if the person is already on a sedative-hypnotic (such as Valium, Xanax, or Benadryl) or pain medication.

An alcohol overdose occurs when there is so much alcohol in the bloodstream that areas of the brain controlling basic life-support functions—such as breathing, heart rate, and temperature control—begin to shut down. Alcohol overdose can lead to permanent brain damage or death.

Symptoms of alcohol overdose may include, mental confusion, difficulty remaining conscious, vomiting, seizures, slow (fewer than 8 breaths per minute) or irregular breathing (10 seconds or more between breaths), slow heart rate, clammy skin, dulled responses, such as no gag reflex (which prevents choking).

Know the danger signals and, if you suspect that someone has an alcohol overdose, call 911 for help immediately. Do not wait for the person to have all the symptoms, and be aware that a person who has passed out can die. Don’t play doctor—cold showers, hot coffee, and walking do not reverse the effects of alcohol overdose and could actually make things worse.

6.4.3 **Marijuana**

Marijuana is the dried leaves and flowers of the *Cannabis sativa* or *Cannabis indica* plant. Stronger forms of the drug include high potency strains - known as sinsemilla (sin-seh-me-yah), hashish (hash for short), and extracts.

Of the more than 500 chemicals in marijuana, *delta-9-tetrahydrocannabinol*, known as THC, is responsible for many of the drug’s psychotropic (mind-altering) effects. It’s this chemical that distorts how the mind perceives the world. In other words, it's what makes a person high.

The amount of THC in marijuana has increased over the past few decades. In the early 1990s, the average THC content in marijuana was less than 4 percent. It is now about 15 percent and much higher in some products such as oils and other extracts (see below). Some people adjust how they consume marijuana (by smoking or eating less) to compensate for the greater potency. There have been reports of people seeking help in emergency rooms with symptoms, including nervousness, shaking, and psychosis after consuming high concentrations of THC.

Smoking extracts and resins from the marijuana plant with high levels of THC is on the rise. These resins have 3 to 5 times more THC than the plant itself. Smoking or vaping it (also called **dabbing**) can deliver dangerous amounts of THC and has led some people to seek treatment in the emergency room. There have also been reports of people injured in fires and explosions caused by attempts to extract hash oil from marijuana leaves using butane (lighter fluid).

6.4.3.0.3 *Marijuana Effects on the Brain and Behavior*

When marijuana is smoked or vaporized, THC quickly passes from the lungs into the bloodstream, which carries it to organs throughout the body, including the brain. Its effects begin almost immediately and can last from 1 to 3 hours. This can affect decision making, concentration, and memory for days after use, especially in people who use marijuana regularly. If marijuana is consumed in "edibles," foods, or beverages the effects of THC appear later—usually in 30 minutes to 1 hour—and may last for many hours. Some people consume more and more waiting for the “high” and end up in the emergency room with uncomfortable symptoms from too much THC.

As it enters the brain, THC attaches to cells, or neurons, with specific kinds of receptors called *cannabinoid receptors*. Normally, these receptors are activated by chemicals similar to THC that occur naturally in the body. They are part of a communication network in the brain called the *endocannabinoid system*. This system is important in normal brain development and function.

Most of the cannabinoid receptors are found in parts of the brain that influence pleasure, memory, thinking, concentration, sensory and time perception, and coordinated movement. Marijuana activates the endocannabinoid system, which causes the "high" and stimulates the release of **dopamine** in the brain's reward centers, reinforcing the behavior. Other effects include changes in perceptions and mood, lack of coordination, difficulty with thinking and problem solving, and disrupted learning and memory.

6.4.3.0.4 *Potential Risks of Marijuana Use and Addiction*

Regular marijuana use has been associated with the following risks:

- **Reduced school performance.** Students who smoke marijuana tend to get lower grades and are more likely to drop out of high school than their peers who do not use. The effects of marijuana on attention, memory, and learning can last for days or weeks.
- **Reduced life satisfaction.** Research suggests that people who use marijuana regularly for a long time are less satisfied with their lives and have more problems with friends and family compared to people who do not use marijuana.
- **Impaired driving.** Marijuana affects a number of skills required for safe driving—alertness, concentration, coordination, and reaction time—so it's not safe to drive high or to ride with someone using marijuana. Marijuana makes it hard to judge distances and react to signals and sounds on the road. Combining marijuana with drinking even a small amount of alcohol greatly increases driving danger, more than either drug alone. Learn more about what happens when you mix **marijuana and driving**.
- **Use of other drugs.** Most young people who use marijuana do not go on to use other drugs. However, those who use are more likely to use other illegal drugs. It isn't clear why some people go on to try other drugs, but researchers have a few theories. The human brain continues to develop into the early 20s. Exposure to addictive substances, including marijuana, may cause changes to the developing brain that make other drugs more appealing.
- **Severe nausea and vomiting.** Studies have shown that in rare cases, regular, long-term marijuana use can lead some people to have cycles of severe nausea, vomiting, and dehydration, sometimes requiring visits to the emergency room.

Marijuana can also be addictive---meaning users continue to use it despite negative consequences. Approximately 10 percent of people who use marijuana may develop what is called a **marijuana use disorder**—problems with their health, school, friendships, family or other conflicts in their life. The person can't stop using marijuana even though it gets in the way of daily life. People who begin using marijuana before the age of 18 are 4–7 times more likely than adults to develop a marijuana use disorder.

What causes one person to become addicted to marijuana while another does not depends on many factors—including their family history (genetics), the age they start using, if they also use other drugs, their family and friend relationships, and if they take part in positive activities like school, clubs, or sports. More research needs to be done to determine if people who use marijuana for medical reasons are at the same risk for addiction as those who use it just to get high.

People who use marijuana may feel a mild **withdrawal** when they stop using the drug, but might not recognize their symptoms as drug withdrawal. These symptoms may include, irritability, sleeplessness, lack of appetite, anxiety, and drug cravings.

These effects can last for several days to a few weeks after drug use is stopped. Relapse (returning to the drug after you've quit) is common during this period because people may crave the drug to relieve these symptoms.

6.4.3.0.5 *Potential Therapeutic Effects*

The marijuana plant itself has not been approved as a medicine by the federal government, yet several states have made it legal for recreational and/or medical use. The plant contains chemicals—called cannabinoids—that may be useful for treating a range of illnesses or symptoms. Here are some samples of cannabinoids that have been approved or are being tested as medicines:

- **THC:** The cannabinoid that can make you "high"—THC—has some medicinal properties. Two laboratory-made versions of THC, nabilone and dronabinol, have been approved by the federal government to treat nausea, prevent sickness and vomiting from chemotherapy in cancer patients, and increase appetite in some patients with AIDS.
- **CBD:** Another chemical in marijuana with potential therapeutic effects is called cannabidiol, or CBD. CBD doesn't have mind-altering effects and is being studied for its possible uses as medicine. For example, CBD oil has been approved as a possible treatment for seizures in children with some severe forms of epilepsy.

- **THC and CBD:** A medication with a combination of THC and CBD is available in several countries outside the United States as a mouth spray for treating pain or the symptoms of multiple sclerosis.

It is important to remember that smoking marijuana can have side effects, making it difficult to develop as a medicine. Side effects might outweigh its value as a medical treatment, especially for people who are not very sick. Another problem with smoking or eating marijuana plant material is that the ingredients can vary a lot from plant to plant, so it is difficult to get an exact dose. Until a medicine can be proven safe and effective, it will not be approved by the federal government.

6.4.4 Tobacco, Nicotine, & Vaping (E-Cigarettes)

Tobacco is a leafy plant grown around the world, including in parts of the United States. There are many chemicals found in tobacco leaves but **nicotine** is the one that can lead to addiction. Other chemicals produced by smoking tobacco, such as tar, carbon monoxide, acetaldehyde, and nitrosamines, also can cause serious harm to the body. For example, tar causes lung cancer and other serious diseases that affect breathing, and carbon monoxide can cause heart problems.

6.4.4.0.6 *How Tobacco and Nicotine Products Are Used*

Tobacco and nicotine products come in many forms. People can smoke, chew, sniff them, or inhale their vapors.

- **Smoked tobacco products.**
  - *Cigarettes:* These are labeled as regular, light, or menthol, but no evidence exists that “lite” or menthol cigarettes are safer than regular cigarettes.
  - *Cigars and pipes:* Some small cigars are hollowed out to make room for marijuana, known as "blunts," often done to hide the fact that they are smoking marijuana.
  - *Hookahs or water pipes*: Hookah tobacco comes in many flavors, and the pipe is typically passed around in groups. A typical hookah session delivers approximately 125 times the smoke, 25 times the tar, 2.5 times the nicotine, and 10 times the carbon monoxide as smoking a cigarette.
- **Smokeless tobacco products.** The tobacco is not burned with these products:
  - *Chewing tobacco.* It is typically placed between the cheek and gums.
  - *Snuff:* Ground tobacco that can be sniffed if dried or placed between the cheek and gums.
  - *Dissolvable products:* These include lozenges, orbs, sticks, and strip.
- **Vaping/electronic cigarettes** (also called e-cigarettes, electronic nicotine delivery systems, vaping devices, e-cigs, or JUULing). Vaping products are battery-operated devices that deliver nicotine and flavorings without burning tobacco. In most products, puffing activates the battery-powered heating device, which vaporizes the liquid in the cartridge. The resulting vapor is then inhaled (called “vaping”).

6.4.4.1.7 *Tobacco Effects on the Brain*

Like many other drugs, nicotine increases levels of a neurotransmitter called **dopamine**. Dopamine is released naturally when you experience something pleasurable like good food, your favorite activity, or spending time with people you care about. When a person uses tobacco products, the release of dopamine causes similar “feel-good” effects. This effect wears off quickly, causing people who smoke to get the urge to light up again for more of that good feeling, which can lead to addiction.

A typical smoker will take 10 puffs on a cigarette over the period of about 5 minutes that the cigarette is lit. So, a person who smokes about a pack of 25 cigarettes a day gets 250 “hits” of nicotine.

When smokeless tobacco is used, nicotine is absorbed through the mouth tissues directly into the blood, where it goes to the brain. Even after the tobacco is removed from the mouth, nicotine continues to be absorbed into the bloodstream. Also, the nicotine stays in the blood longer for users of smokeless tobacco than for smokers.

6.4.4.1.8 *Is Vaping Worse Than Smoking?*

Regardless of how vaping compares to cigarette smoking, it is important to recognize that nicotine vaping has its own risks, which include addiction and other potentially harmful health effects. Research so far suggests that nicotine vaping might be less harmful than cigarettes when people who regularly smoke switch to them completely and no longer use tobacco cigarettes.

However, nicotine in any form is a highly addictive drug, and health experts have raised many questions about the safety of vaping devices, particularly for teens:

- Testing of some vaping products found the aerosol (vapor) to contain known cancer-causing and toxic chemicals. The health effects of repeated exposure to these chemicals are not yet clear.
- Some research suggests that nicotine vaping may increase the likelihood that teens will try other tobacco products, including cigarettes. A study showed that students who have vaped nicotine by the time they start 9th grade are more likely than others to start smoking traditional cigarettes and other smoked tobacco within the next year.
- Some research suggests that certain vaping products contain metals like nickel and chromium, possibly coming from the heating of coils, that may be harmful when inhaled.

6.4.4.1.9 *Potential Risks and Addiction*

It is clear that the nicotine in tobacco is addictive. Each cigarette contains about 10 milligrams of nicotine. A person inhales only some of the smoke from a cigarette, and not all of each puff is absorbed in the lungs. The average person gets about 1 to 2 milligrams of nicotine from each cigarette.

Studies of widely used brands of smokeless tobacco showed that the amount of nicotine per gram of tobacco ranges from 4.4 milligrams to 25.0 milligrams. Holding an average-size dip in your mouth for 30 minutes gives you as much nicotine as smoking 3 cigarettes. A 2-can-a-week snuff dipper gets as much nicotine as a person who smokes 1½ packs a day.

Whether a person smokes tobacco products or uses smokeless tobacco, the amount of nicotine absorbed in the body is enough to make someone addicted. When this happens, the person continues to seek out the tobacco even though he or she understands the harm it causes. Nicotine addiction can cause:

- *Tolerance:* Over the course of a day, someone who uses tobacco products develops tolerance—more nicotine is required to produce the same initial effects. In fact, people who smoke often report that the first cigarette of the day is the strongest or the "best."
- *Withdrawal:* When people quit using tobacco products, they usually experience uncomfortable **withdrawal** symptoms, which often drive them back to tobacco use. Nicotine withdrawal symptoms include irritability, problems with thinking and paying attention, sleep problems, increased appetite, and craving, which may last 6 months or longer, and can be a major stumbling block to quitting.

6.4.4.2 Methamphetamine

Methamphetamine is a powerful, highly addictive stimulant that affects the central nervous system. Crystal methamphetamine is a form of the drug that looks like glass fragments or shiny, bluish-white rocks. It is chemically similar to amphetamine, a drug used to treat attention-deficit hyperactivity disorder (ADHD) and narcolepsy, a sleep disorder.

People can take methamphetamine by, smoking, swallowing (pill), snorting, or injecting the powder that has been dissolved.

Because the "high" from the drug both starts and fades quickly, people often take repeated doses in a "binge and crash" pattern. In some cases, people take methamphetamine in a form of binging known as a "run," giving up food and sleep while continuing to take the drug every few hours for up to several days.

6.4.4.2.1 *Methamphetamine Effects on the Brain and Behavior*

Methamphetamine increases the amount of the natural chemical **dopamine** in the brain. Dopamine is involved in body movement, motivation, and reinforcement of rewarding behaviors. The drug's ability to rapidly release high levels of dopamine in reward areas of the brain strongly reinforces drug-taking behavior, making the user want to repeat the experience.

6.4.4.2.2 *Short-Term Effects*

Taking even small amounts of methamphetamine can result in many of the same health effects as those of other stimulants, such as cocaine or amphetamines. These include, increased wakefulness and physical activity, decreased appetite, faster breathing, rapid and/or irregular heartbeat, and increased blood pressure and body temperature.

6.4.4.2.3 *Long-Term Effects*

Long-term methamphetamine use has many other negative consequences, including addiction, extreme weight loss, severe dental problems, anxiety, altered judgement and decision making, confusion, memory loss, sleeping problems, violent behavior, ***paranoia, and hallucinations***.

In addition, continued methamphetamine use can cause changes in brain structure and function. In particular, the brain's dopamine system is often affected, resulting in reduced coordination and impaired verbal learning. In studies of people who used methamphetamine over the long term, severe changes also affected areas of the brain involved with emotion and memory.

Although some of these brain changes may reverse after being off the drug for a year or more, other changes may not recover even after a long period of time. A recent study even suggests that people who once used methamphetamine have an increased the risk of developing Parkinson's disease, a disorder of the nerves that affects movement.

People who inject methamphetamine are at increased risk of contracting infectious diseases such as HIV and hepatitis B and C. In addition, methamphetamine use may worsen the progression of HIV/AIDS and its consequences. Studies indicate that HIV causes more injury to nerve cells and more cognitive problems in people who use methamphetamine than it does in people who have HIV and don't use the drug.

6.4.4.3 **MDMA (i.e. Ecstasy or Molly)**

3,4-methylenedioxy-methamphetamine (**MDMA**) is a synthetic, emphatogenic drug that alters mood and perception (awareness of surrounding objects and conditions). It is chemically similar to both stimulants and hallucinogens, producing feelings of increased energy, pleasure, emotional warmth, and distorted sensory and time perception.

MDMA was initially popular in the nightclub scene and at all-night dance parties ("raves"), but the drug now affects a broader range of people who more commonly call the drug Ecstasy or Molly.

6.4.4.3.1 *How do people use MDMA?*

People who use MDMA usually take it as a capsule or tablet, though some swallow it in liquid form or snort the powder. The popular nickname Molly (slang for "molecular") often refers to the supposedly "pure" crystalline powder form of MDMA, usually sold in capsules. However, people who purchase powder or capsules sold as Molly often actually get other drugs such as synthetic cathinones ("bath salts") instead. In addition, some people take MDMA in combination with other drugs such as alcohol or marijuana.

6.4.4.3.2 *MDMA Effects on the Brain and Behavior*

MDMA increases the activity of three brain chemicals:

- Dopamine—produces increased energy/activity and acts in the reward system to reinforce behaviors
- Norepinephrine—increases heart rate and blood pressure, which are particularly risky for people with heart and blood vessel problems
- Serotonin—affects mood, appetite, sleep, and other functions. It also triggers hormones that affect sexual arousal and trust. The release of large amounts of serotonin likely causes the emotional closeness, elevated mood, and empathy felt by those who use MDMA.

6.4.4.3.3 *Adverse Effects*

MDMA's effects last about 3 to 6 hours, although many users take a second dose as the effects of the first dose begin to fade. Negative health effects include nausea, muscle cramping, blurred vision, chills, sweating, and others. Over the course of the week following moderate use of the drug, a person may experience:

- irritability, impulsiveness and/or aggression
- depression
- sleep problems
- anxiety
- memory and attention problems
- decreased appetite

- decreased interest in and pleasure from sex

It's possible that some of these effects may be due to the combined use of MDMA with other drugs, especially marijuana.

6.4.4.3.4 *Addictive Potential*

Research results vary on whether MDMA is addictive. Experiments have shown that animals will self-administer MDMA—an important indicator of a drug's abuse potential—although to a lesser degree than some other drugs such as cocaine.

Some people report signs of addiction, including withdrawal symptoms, such as fatigue, loss of appetite, depression, and trouble concentrating.

6.4.4.3.5 *Therapeutic* Potential

MDMA was first used in the 1970s as an aid in psychotherapy (mental disorder treatment using "talk therapy"). The drug did not have the support of clinical trials (studies using humans) or approval from the U.S. Food and Drug Administration (FDA). In 1985, The U.S. Drug Enforcement Administration (DEA) labeled MDMA as an illegal drug with no recognized medicinal use. However, some researchers remain interested in its value in psychotherapy when given to patients under carefully controlled conditions. MDMA is currently in clinical trials as a possible treatment aid for post-traumatic stress disorder (PTSD); for anxiety in terminally ill patients; and for social anxiety in autistic adults. Recently, the FDA gave MDMA-assisted psychotherapy for PTSD a Breakthrough Therapy designation.

6.4.5 **Heroin**

Heroin is a very addictive drug made from morphine, a psychoactive (mind-altering) substance taken from the resin of the seed pod of the opium poppy plant. Heroin is part of a class of drugs called *opioids*. Other opioids include some prescription pain relievers, such as codeine, oxycodone, and hydrocodone.

Heroin use and overdose deaths have dramatically increased recently. This increase is related to the growing number of people misusing prescription opioid pain relievers like OxyContin and Vicodin. Some people who become addicted to those drugs switch to heroin because it produces similar effects but is cheaper and may be easier to get.

In fact, most people who use heroin report they first misused prescription opioids, but it is a small percentage of people who switch to heroin. The numbers of people misusing prescription drugs is so high, that even a small percentage translates to hundreds of thousands of heroin users.

Heroin is mixed with water and injected with a needle. It can also be sniffed, smoked, or snorted. People who use heroin sometimes combine it with other drugs, such as alcohol or cocaine (a "speedball"), which can be particularly dangerous and raise the risk of overdose.

6.4.5.0.1 *Heroin Effects on the Brain and Behavior*

When heroin enters the brain, it attaches to molecules on cells known as opioid receptors. These receptors are located in many areas of the brain and body, especially areas involved in the perception of pain and pleasure, as well as a part of the brain that regulates breathing.

Short-term effects of heroin include a rush of good feelings and clouded thinking. These effects can last for a few hours, and during this time people feel drowsy, and their heart rate and breathing slow down. When the drug wears off, people experience a depressed mood and often crave the drug to regain the good feelings.

Regular heroin use changes the functioning of the brain and can result in the following:

- *tolerance*: more of the drug is needed to achieve the same "high"
- *dependence*: the need to continue use of the drug to avoid withdrawal symptoms
- *addiction (i.e. substance use disorder):* a devastating brain disease where, without proper treatment, people have trouble stopping using drugs even when they really want to and even after it causes terrible consequences to their health and other parts of their lives.

6.4.5.0.1 *Adverse Effects of Heroin*

Opioid receptors are located in the brain, the brain stem, down the spinal cord, and in the lungs and intestines. Thus, using heroin can result in a wide variety of physical problems related to breathing and other basic life functions, some of which may be very serious. A sample of these effects are given below.

Short-Term Effects

- dry mouth
- warm flushing skin
- feeling sick to the stomach and throwing up
- severe itching
- clouded thinking
- going "on the nod," switching back and forth between being conscious and semi-conscious
- increased risk of HIV and hepatitis (a liver disease) through shared needles and poor judgment while “high” leading to other risky behaviors.

When mixed with alcohol, short-term effects can include:

- coma—a deep state of unconsciousness
- dangerously slowed (or even stopped) breathing that can lead to overdose death

Long-Term Effects

- problems sleeping
- damage to the tissues inside the nose for people who sniff or snort it
- infection of the heart
- constipation and stomach cramping
- liver and kidney disease
- lung problems
- mental health problems, such as depression
- sexual problems for men & changes in menstrual cycles for women

In addition to the effects of the drug itself, heroin bought on the street often contains a mix of substances, including the dangerous opioid called **fentanyl** (see section below). Drug dealers add fentanyl because it is cheap, and they can save money. Some of these substances can be toxic and can clog the blood vessels leading to the lungs, liver, kidney, or brain. This can cause permanent damage to those organs.

6.4.5.1 Fentanyl

Fentanyl is a powerful synthetic opioid that is similar to morphine but is 50 to 100 times more potent. It is a prescription drug that is also made and used illegally. Like morphine, it is a medicine that is typically used to treat patients with severe pain, especially after surgery. It is also sometimes used to treat patients with chronic pain who are physically tolerant to other opioids. Tolerance occurs when you need a higher and/or more frequent amount of a drug to get the desired effects.

Synthetic opioids, including fentanyl, are now the most common drugs involved in drug overdose deaths in the United States. In 2017, 59 percent of opioid-related deaths involved fentanyl compared to 14.3 percent in 2010.

When prescribed by a doctor, fentanyl can be given as a shot, a patch that is put on a person’s skin, or as lozenges that are sucked like cough drops. The illegally used fentanyl most often associated with recent overdoses is made in labs. This synthetic fentanyl is sold illegally as a powder, dropped onto blotter paper, put in eye droppers and nasal sprays, or made into pills that look like other prescription opioids.

Some drug dealers are mixing fentanyl with other drugs, such as heroin, cocaine, methamphetamine, and MDMA (Ectasy). This is because it takes very little to produce a high with fentanyl, making it a cheaper option. This is especially risky when people taking drugs don’t realize they might contain fentanyl as a cheap but dangerous additive. They might be taking stronger opioids than their bodies are used to and can be more likely to overdose.

6.4.5.1.1 *Fentanyl Effects on the Brain and Behavior*

Like heroin, morphine, and other opioid drugs, fentanyl works by binding to the body's opioid receptors, which are found in areas of the brain that control pain and emotions. After taking opioids many times, the brain adapts to the drug, diminishing its sensitivity, making it hard to feel pleasure from anything besides the drug.

Fentanyl's effects include

- extreme happiness
- drowsiness
- nausea
- confusion
- constipation
- sedation
- problems breathing
- unconsciousness

Fentanyl is addictive because of its potency. A person taking prescription fentanyl as instructed by a doctor can experience dependence, which is characterized by withdrawal symptoms when the drug is stopped. A person can be dependent on a substance without being addicted, but dependence can sometimes lead to addiction.

People addicted to fentanyl who stop using it can have severe withdrawal symptoms that begin as early as a few hours after the drug was last taken. These symptoms include:

- muscle and bone pain
- sleep problems
- diarrhea and vomiting
- cold flashes with goose bumps
- uncontrollable leg movements
- severe cravings

These symptoms can be extremely uncomfortable and are the reason many people find it so difficult to stop taking fentanyl. There are medicines being developed to help with the withdrawal process for fentanyl and other opioids. The FDA has approved lofexidine, a non-opioid medicine designed to reduce opioid withdrawal symptoms. Also, the NSS-2 Bridge device is a small electrical nerve stimulator placed behind the person’s ear, that can be used to try to ease symptoms for up to five days during the acute withdrawal phase. In December 2018, the FDA cleared a mobile medical application, reSET®, to help treat opioid use disorders. This application is a prescription cognitive behavioral therapy and should be used in conjunction with treatment that includes buprenorphine and contingency management.

6.4.5.1.2 *Overdose Risk*

The rise of fentanyl misuse has coincided with an unprecedented rise in opioid-related overdose deaths. An overdose occurs when a drug produces serious adverse effects and life-threatening symptoms. When people overdose on fentanyl or other drugs mixed with fentanyl, their breathing can slow or stop. This can decrease the amount of oxygen that reaches the brain, a condition called *hypoxia*. Hypoxia can lead to a coma and permanent brain damage, and even death.

Attributions

Adapted from National Institute of Drug Abuse (NIDA) - Drug Facts and National Institute of Drug Abuse (NIDA) - Alcohol. 2021, June 10. Drug Facts. **License**: *Public Domain: No Known Copyright*

6.4.5.2 References

Chaffee BW, Watkins SL, Glantz SA. Electronic Cigarette Use and Progression From Experimentation to Established Smoking. Pediatrics. March 2018:e20173594. doi:10.1542/peds.2017-3594.

Hasin DS, Saha TD, Kerridge BT, et al. Prevalence of Marijuana Use Disorders in the United States Between 2001-2002 and 2012-2013. *JAMA Psychiatry*. 2015;72(12):1235-1242. doi:10.1001/jamapsychiatry.2015.1858.

Hess CA, Olmedo P, Goessler W, Cohen E, Rule AM. E-cigarettes as a source of toxic and potentially carcinogenic metals. *Environmental Research.* 2017;152:221-221.

National Institute of Drug Abuse (NIDA). 2021, June 10. Drug Facts. Retrieved from https://teens.drugabuse.gov/teens/drug-facts on 2022, March 7Winters KC, Lee C-YS. Likelihood of developing an alcohol and cannabis use disorder during youth: association with recent use and age. *Drug Alcohol Depend*. 2008;92(1-3):239-247. doi:10.1016/j.drugalcdep.2007.08.005.

Rigotti NA. e-Cigarette use and subsequent tobacco use by adolescents: new evidence about a potential risk of e-cigarettes. *JAMA.* 2015;314(7):673-674.

U.S. Department of Health, and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease, Prevention and Health Promotion, Office on Smoking and Health. E-Cigarette Use Among Youth And Young Adults: A Report of the Surgeon General — Executive Summary.; 2016. **https://e-cigarettes.surgeongeneral....R_Exec_Summ_50….** Accessed February 21, 2017.

This page titled 6.4: Specific Drug Effects, Misuse, and Risks is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

6.5: Psychotherapeutics

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Differentiate the various classifications of Psychotherapeutic drugs based on their therapeutic action, effects on steps of neurotransmission, and neurotransmitter specificity
- Compare and contrast the effectiveness and side effects of Psychotherapeutics

6.5.1 Overview

Psychotherapeutic medications can play a role in treating several mental disorders and conditions. Treatment may also include psychotherapy (also called “talk therapy”) and a variety of other medical and non-medical interventions. In some cases, psychotherapy alone may be the best treatment option. Choosing the right treatment plan should be based on a person's individual needs and medical situation, and under a mental health professional’s care.

The National Institute of Mental Health (NIMH) also does not endorse or recommend any particular drug, herb, or supplement. Results from NIMH-supported clinical trials that examine the effectiveness of treatments, including medications, are reported in the medical literature. This section is intended to provide basic information about mental health medications. It is not a complete source for all medications available and should not be used as a guide for making medical decisions.

Information about medications changes frequently. Check the U.S. Food and Drug Administration (FDA) website for the latest warnings, patient medication guides, or newly approved medications. Brand names are not referenced on this page, but you can search by brand name on MedlinePlus Drugs, Herbs and Supplements Drugs website. The MedlinePlus website also provides additional information about each medication, including side effects and FDA warnings.

6.5.2 Psychotherapeutic Classification

Much like the general classification of psychoactive drugs, there are multiple schemes of categorization that can be used. The most clear and useful system is based on the therapeutic purpose of the drugs rather than their synaptic mechanism of action or physiological effects (i.e. stimulant, depressant, etc.). Comprehensive information on various psychotherapeutic medications is available from the National Institute of Mental Health.

6.5.3 Antidepressants

Antidepressants are medications commonly used to treat depression. Antidepressants are also used for other health conditions, such as anxiety, pain and insomnia. Although antidepressants are not FDA-approved specifically to treat ADHD, antidepressants are sometimes used to treat ADHD in adults.

The most popular types of antidepressants are called selective serotonin reuptake inhibitors (**SSRIs**). Examples of SSRIs include, Fluoxetine (Prozac), Citalopram (Celexa), and Escitalopram (Lexapro). Other types of antidepressants are serotonin and norepinephrine reuptake inhibitors (SNRIs). SNRIs are similar to SSRIs and include venlafaxine (Effexor) and duloxetine (Cymbalta). By inhibiting the reuptake of either serotonin or norepinephrine, these drugs enhance the activity of these systems by allowing these transmitters to stay in the synapse longer.

Another antidepressant that is commonly used is bupropion. **Bupropion** is a third type of antidepressant which **inhibits the reuptake** of **both norepinephrine and dopamine**. Bupropion is also used to treat seasonal affective disorder and to help people stop smoking.

SSRIs, SNRIs, and bupropion are popular because they do not cause as many side effects as older classes of antidepressants, and seem to help a broader group of depressive and anxiety disorders. Older antidepressant medications include tricyclics, tetracyclics, and monoamine oxidase inhibitors (MAOIs). For some people, these may be the best medications.

6.5.3.1 *How do people respond to antidepressants?*

According to a research review by the Agency for Healthcare Research and Quality, all antidepressant medications work about as well as each other to improve symptoms of depression and to keep depression symptoms from coming back. For reasons not yet well understood, some people respond better to some antidepressant medications than to others.

Therefore, it is important to know that some people may not feel better with the first medicine they try and may need to try several medicines to find the one that works for them. Others may find that a medicine helped for a while, but their symptoms came back. It is important to carefully follow your doctor's directions for taking your medicine at an adequate dose and over an extended period of time (often 4 to 6 weeks) for it to work.

Once a person begins taking antidepressants, it is important to not stop taking them without the help of a doctor. Sometimes people taking antidepressants feel better and stop taking the medication too soon, and the depression may return. When it is time to stop the medication, the doctor will help the person slowly and safely decrease the dose. It's important to give the body time to adjust to the change. People don't get addicted (or "hooked") on these medications, but stopping them abruptly may also cause withdrawal symptoms

6.5.3.2 *Side Effects of Antidepressants*

Some antidepressants may cause more side effects than others. You may need to try several different antidepressant medications before finding the one that improves your symptoms and that causes side effects that you can manage.

The most common side effects listed by the FDA include:

- Nausea and vomiting
- Weight gain
- Diarrhea
- Sleepiness
- Sexual problems

Combining the newer SSRI or SNRI antidepressants with one of the commonly-used "triptan" medications used to treat migraine headaches could cause a life-threatening illness called "serotonin syndrome." A person with serotonin syndrome may be agitated, have hallucinations (see or hear things that are not real), have a high temperature, or have unusual blood pressure changes. Serotonin syndrome is usually associated with the older antidepressants called MAOIs, but it can happen with the newer antidepressants as well, if they are mixed with the wrong medications.

6.5.4 Anti-Anxiety Medications

Anti-anxiety medications help reduce the symptoms of anxiety, such as panic attacks, or extreme fear and worry. The most common anti-anxiety medications are called **benzodiazepines** which act to increase GABA function. Benzodiazepines can be effective in treating generalized anxiety disorder, but in the case of panic disorder or social phobia (social anxiety disorder), benzodiazepines are usually second-line treatments, behind SSRIs or other antidepressants.

Benzodiazepines used to treat anxiety disorders include, Clonazepam (Klonopin), Alprazolam (Xanax), and Lorazepam (Ativan).

Short half-life (or short-acting) benzodiazepines (such as Lorazepam) and beta-blockers (adrenaline inhibitors) are used to treat the short-term symptoms of anxiety. Beta-blockers help manage physical symptoms of anxiety, such as trembling, rapid heartbeat, and sweating that people with phobias (an overwhelming and unreasonable fear of an object or situation, such as public speaking) experience in difficult situations. Taking these medications for a short period of time can help the person keep physical symptoms under control and can be used "as needed" to reduce acute anxiety.

Buspirone (which is unrelated to the benzodiazepines) is sometimes used for the long-term treatment of chronic anxiety. In contrast to the benzodiazepines, buspirone has both agonist and antagonist properties and effects serotonin and dopamine systems in complex ways. This drug must be taken every day for a few weeks to reach its full effect, thus, it is not useful for immediate treatment of anxiety symptoms.

MDMA is an empathogen with both stimulant and hallucinogenic properties. As mentioned in Section 6.4, it was first used in the 1970s as an aid in psychotherapy but faced many legal and regulatory challenges in subsequent years that stalled both the research and clinical tests of its effectiveness. However, some researchers remained interested in its value in psychotherapy when given to

patients under carefully controlled conditions. MDMA is currently in clinical trials as a possible treatment aid for post-traumatic stress disorder (PTSD); for anxiety in terminally ill patients; and for social anxiety in autistic adults. Recently, the FDA gave MDMA-assisted psychotherapy for PTSD a "Breakthrough Therapy" designation.

6.5.4.1 How do people respond to anti-anxiety medications?

Anti-anxiety medications such as benzodiazepines are effective in relieving anxiety and take effect more quickly than the antidepressant medications (or buspirone) often prescribed for anxiety. However, people can build up a tolerance to benzodiazepines if they are taken over a long period of time and may need higher and higher doses to get the same effect. Some people may even become dependent on them. To avoid these problems, doctors usually prescribe benzodiazepines for short periods, a practice that is especially helpful for older adults, people who have substance abuse problems and people who become dependent on medication easily. If people suddenly stop taking benzodiazepines, they may have withdrawal symptoms or their anxiety may return. Therefore, benzodiazepines should be tapered off slowly.

6.5.4.2 *Side Effects of A*nti-Anxiety Medications

Like other medications, anti-anxiety medications may cause side effects. Some of these side effects and risks are serious. The most common side effects for benzodiazepines are drowsiness and dizziness. For beta-blockers the common side effects include, fatigue, cold hands, dizziness or light-headedness, and weakness. Beta-blockers generally are not recommended for people with asthma or diabetes because they may worsen symptoms related to both. Possible side effects from buspirone include, dizziness, headaches, nausea, nervousness, lightheadedness, excitement, and trouble sleeping

6.5.5 Stimulants

As the name suggests, stimulants increase alertness, attention, and energy, as well as elevate blood pressure, heart rate, and respiration (National Institute on Drug Abuse, 2014). Stimulant medications are often prescribed to treat children, adolescents, or adults diagnosed with Attention Deficit, Hyperactivity Disorder (ADHD).

Two of the most common stimulants used to treat ADHD are Methylphenidate (Ritalin), and Amphetamine/Dextroamphetamine (Adderall)

Stimulants are also prescribed to treat other health conditions, including narcolepsy, and occasionally depression (especially in older or chronically medically ill people and in those who have not responded to other treatments).

6.5.5.1 *How do people respond to stimulants?*

Stimulants have the potential to cause excessive activity and physiological activation, however, for individuals with ADHD the opposite is true in that it can have a calming and “focusing” effect. Stimulant medications are safe when given under a doctor's supervision. Some children taking them may feel slightly different or "funny."

Some parents worry that stimulant medications may lead to drug abuse or dependence, but there is little evidence of this when they are used properly as prescribed. Additionally, research shows that teens with ADHD who took stimulant medications were less likely to abuse drugs than those who did not take stimulant medications. Despite the findings noted above, there is also research indicating the risk of "diversion" of prescribed stimulants for nonmedical uses between students at educational institutions. This diversion could put individuals at a higher risk for stimulant abuse.

6.5.5.2 *Side Effects of S*timulants

Most side effects are minor and disappear when dosage levels are lowered. The most common side effects include:

- Difficulty falling asleep or staying asleep
- Loss of appetite
- Stomach pain
- Headache

Less common side effects include:

- Motor tics or verbal tics (sudden, repetitive movements or sounds)
- Personality changes, such as appearing “flat” or without emotion

6.5.6 **Antipsychotics**

Antipsychotic medicines are primarily used to manage psychosis. The word “psychosis” is used to describe conditions that affect the mind, and in which there has been some loss of contact with reality, often including delusions (false, fixed beliefs) or hallucinations (hearing or seeing things that are not really there). It can be a symptom of a physical condition such as drug abuse or a mental disorder such as schizophrenia, bipolar disorder, or very severe depression (also known as “psychotic depression”).

Antipsychotic medicines do not cure these conditions. They are used to help relieve symptoms and improve quality of life.

Older or first-generation antipsychotic medications are also called conventional "typical" antipsychotics or “neuroleptics”. Most of the drugs in this class act as potent dopamine antagonists primarily effecting the dopamine D2 receptor. This antagonistic effect reduces the dopamine overactivity found in psychotic disorders, such as Schizophrenia. Some of the common first-generation antipsychotics include:

- Chlorpromazine
- Haloperidol
- Fluphenazine

Newer or second generation medications are also called "atypical" antipsychotics. These drugs act on various neurotransmitters systems in addition to dopamine and, as such are not as potent in their dopamine antagonism. Some of the common atypical antipsychotics include:

- Risperidone
- Olanzapine
- Quetiapine

Both typical and atypical antipsychotics have been shown to work to treat symptoms of schizophrenia and the manic phase of bipolar disorder. In addtion, several atypical antipsychotics have a “broader spectrum” of action than the older medications, and are used for treating bipolar depression or depression that has not responded to an antidepressant medication alone.

6.5.6.3 *How do people respond to antipsychotics?*

Certain symptoms, such as feeling agitated and having hallucinations, usually go away within days of starting an antipsychotic medication. Symptoms like delusions usually go away within a few weeks, but the full effects of the medication may not be seen for up to six weeks. Every patient responds differently, so it may take several trials of different antipsychotic medications to find the one that works best.

Some people may have a relapse—meaning their symptoms come back or get worse. Usually relapses happen when people stop taking their medication, or when they only take it sometimes. Some people stop taking the medication because they feel better or they may feel that they don't need it anymore, but no one should stop taking an antipsychotic medication without talking to his or her doctor.When a doctor says it is okay to stop taking a medication, it should be gradually tapered off— never stopped suddenly. Many people must stay on an antipsychotic continuously for months or years in order to stay well; treatment should be personalized for each individual.

6.5.6.4 *Side Effects of A**ntipsychotics?***

Antipsychotics have many side effects (or adverse events) and risks. The FDA lists the following side effects of antipsychotic medicines:

- Drowsiness
- Dizziness
- Restlessness
- Weight gain (the risk is higher with some atypical antipsychotic medicines)
- Dry mouth
- Constipation
- Nausea

- Vomiting
- Blurred vision
- Low blood pressure
- Uncontrollable movements, such as tics and tremors (the risk is higher with typical antipsychotic medicines)
- Seizures
- A low number of white blood cells, which fight infections

The older, typical antipsychotic medications can also cause additional side effects related to physical movement, such as:

- Rigidity
- Persistent muscle spasms
- Tremors
- Restlessness

Long-term use of typical antipsychotic medications may lead to a condition called tardive dyskinesia (TD). TD causes muscle movements, commonly around the mouth, that a person can't control. TD can range from mild to severe, and in some people, the problem cannot be cured. Sometimes people with TD recover partially or fully after they stop taking typical antipsychotic medication. People who think that they might have TD should check with their doctor before stopping their medication. TD rarely occurs while taking atypical antipsychotics.

6.5.7 Mood Stabilizers

Mood stabilizers are used primarily to treat bipolar disorder, mood swings associated with other mental disorders, and in some cases, to augment the effect of other medications used to treat depression. **Lithium**, which is an effective mood stabilizer, has a broad spectrum of action including the inhibition of dopamine and glutamate systems and increased GABA action (Malhi, et al., 2013). Lithium is approved for the treatment of mania and the maintenance treatment of bipolar disorder. Additionally, a number of cohort studies describe anti-suicide benefits of lithium for individuals on long-term maintenance. Because mood stabilizers work primarily by decreasing abnormal activity in the brain they are also sometimes used to treat:

- Depression (usually along with an antidepressant)
- Schizoaffective Disorder
- Disorders of impulse control
- Certain mental illnesses in children

Anticonvulsant medications that primarily act as agonists to GABA neurotransmitter systems are also used as mood stabilizers. They were originally developed to treat seizures, but they were found to help control unstable moods as well. One anticonvulsant commonly used as a mood stabilizer is valproic acid (also called divalproex sodium). For some people, especially those with “mixed” symptoms of mania and depression or those with rapid-cycling bipolar disorder, valproic acid may work better than lithium. Other anticonvulsants used as mood stabilizers include:

- Carbamazepine
- Lamotrigine
- Oxcarbazepine

6.5.7.1 *Side Effects of M*ood Stabilizers

Mood stabilizers can cause several side effects, and some of them may become serious, especially at excessively high blood levels. These side effects include:

- Itching, rash
- Excessive thirst
- Frequent urination
- Tremor (shakiness) of the hands
- Nausea and vomiting
- Slurred speech
- Fast, slow, irregular, or pounding heartbeat

- Blackouts
- Changes in vision
- Seizures
- Hallucinations (seeing things or hearing voices that do not exist)
- Loss of coordination
- Swelling of the eyes, face, lips, tongue, throat, hands, feet, ankles, or lower legs.

Some possible side effects linked to anticonvulsants (such as valproic acid) include:

- Drowsiness
- Dizziness
- Headache
- Diarrhea
- Constipation
- Changes in appetite and/or weight
- Agitation
- Mood swings
- Abnormal thinking
- Uncontrollable shaking or movement of a part of the body
- Loss of coordination
- Blurred or double vision
- Ringing in the ears
- Hair loss

6.5.8 Summary

It should be clear that psychotherapeutic medications have been proven effective at treating a wide variety of psychological disorders, with only a few major types reviewed here. The drugs that treat depression, anxiety, psychosis, and various other disorders of mood and cognition do so by a variety of means. Each category of psychotherapeutic drug can include various different drugs with different actions at the synaptic and neurotransmitter levels. Many other drugs not referenced here, such as hallucinogens and marijuana have also been studied for their potential effectiveness in treating disorders, such as Post Traumatic Stress Disorder, Anxiety, and depression. Some positive outcomes have been found, but more definitive research is needed in order to validate these findings.

6.5.9 Attributions

Adapted from "Mental Health Medications." **Authored and Provided by** the National Institute of Mental Health. . **License**: *Public Domain: No Known Copyright*

6.5.10 References

Malhi GS, Tanious M, Das P, Coulston CM, Berk M. (2013) Potential mechanisms of action of lithium in bipolar disorder. Current understanding. *CNS Drugs*. Feb;27(2):135-53. doi: 10.1007/s40263-013-0039-0. PMID: 23371914.

Vrecko S. (2015). Everyday drug diversions: a qualitative study of the illicit exchange and non-medical use of prescription stimulants on a university campus. *Social science & medicine*, *131*, 297–304. https://doi.org/10.1016/j.socscimed.2014.10.01

This page titled 6.5: Psychotherapeutics is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

CHAPTER OVERVIEW

7: Senses

> This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

7.1: Senses Overview
7.2: Vision
7.3: Hearing
7.4: Touch
7.5: Taste and Smell

This page titled 7: Senses is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

7.1: Senses Overview

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

7.1.1 Senses Overview

Organisms such as humans have been identified as having five basic senses. The five basic senses that are most commonly identified in humans are sight, hearing, touch, smell, and taste. These senses help individuals understand and make sense of the environment around them. How do the senses work? Organisms and humans have sensory receptors which are specialized neurons that react to specific types of stimuli in the environment. When a stimulus is detected by a sensory receptor, a sensation response has occurred**.** There is typically a specified area or space in which a given sensory receptor can identify or respond to a stimulus, be it far away or in close proximity to the body; that is the receptor's **receptive field**. Think of everyday experiences organisms may have with the environment around them and how different receptive fields are different for the different senses. For example: for the sense of touch, a stimulus must come into contact with the body. For the sense of hearing, a stimulus can be a moderate distance away (some baleen whale sounds can propagate for many kilometers). For vision, a stimulus can be very far away; for example, the visual system perceives light from stars at enormous distances. A sensory receptor identifies stimuli in the environment and the conversion from sensory stimulus energy to action potential is known as **transduction**. Transduction in the nervous system typically refers to stimulus-alerting events wherein a physical stimulus is converted into an action potential (see image below), which is transmitted along axons towards the central nervous system for integration. It is a step in the larger process of sensory processing (see Figure **7.1.1**).

Information about reusing

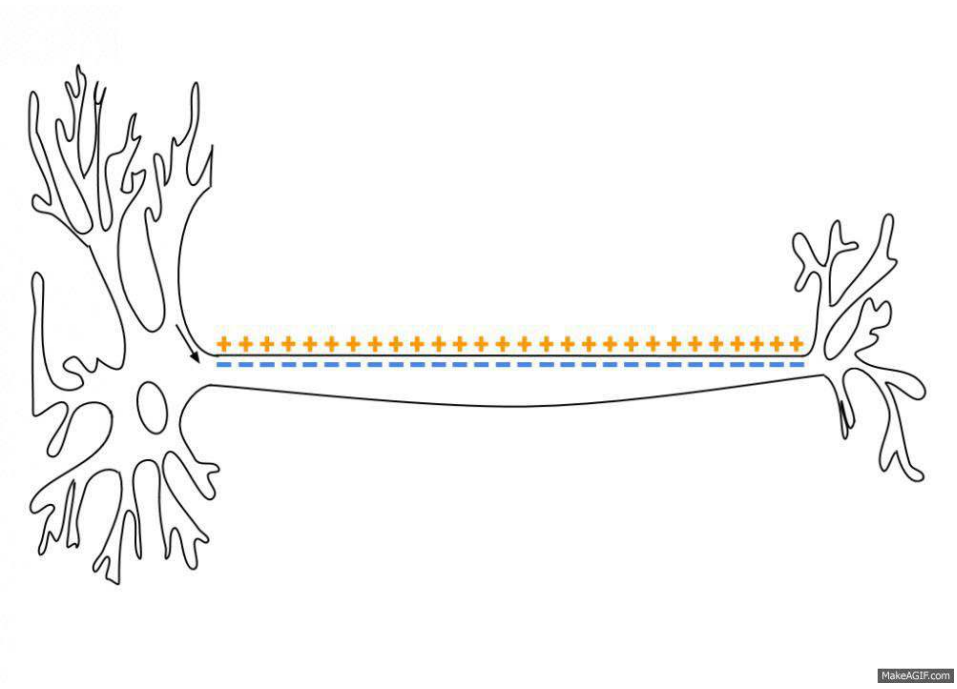


Figure **7.1.1**: As an action potential (nerve impulse) travels down an axon there is a change in polarity across the membrane of the axon. In response to a signal from another neuron, sodium- (Na^+) and potassium- (K^+) gated ion channels open and close as the membrane reaches its threshold potential. Na^+ channels open at the beginning of the action potential, and Na^+ moves into the axon, causing depolarization. Repolarization occurs when the K^+ channels open and K^+ moves out of the axon, creating a change in polarity between the outside of the cell and the inside. The impulse travels down the axon in one direction only, to the axon terminal where it signals other neurons. By Laurentaylorj - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/inde...curid=26311114

This chapter will later go into more depth about the difference between sensation and perception but the initial part of the chapter will focus on sensation. In short, senses help organisms identify a familiar sound, or recognize smoke when there is a fire. The receptors are different for each sense and they are specialized according to the type of stimulus they sense. For example, touch receptors, light receptors, and sound receptors are each activated by different stimuli. Touch receptors are not sensitive to light or sound; they are sensitive only to touch or pressure. However, stimuli may be combined at higher levels in the brain, as happens with olfaction, contributing to our sense of taste. These senses will be discussed in more detail throughout this chapter along with a brief introduction into how these senses at times work together to help individuals have a deeper understanding of the environment around them and how this deeper understanding can help individuals respond to that environment.

7.1.2 Attributions

Sensation vs Perception adapted by Isaias Hernandez from "PressBooks licensed CC BY-NC-SA 4.0"

Adapted from 5.1 Sensation versus Perception by Kathryn Dumper, William Jenkins, Arlene Lacombe, Marilyn Lovett, and Marion Perimutter is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, except where otherwise noted.

Auto Attribution

5.1 Sensation vs. Perception by PressBooks is licensed CC BY-NC-SA 4.0

This page titled 7.1: Senses Overview is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

7.2: Vision

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

7.2.1 Vision

Learning Objectives

- Identify the key structures of the eye and the role they play in vision.
- Summarize how the eye and the visual cortex work together to sense and perceive the visual stimuli in the environment, including processing colors, shape, depth, and motion.
- Explain the benefit of having two eyes.
- Understand differences in color vision.
- Understand depth perception and vision.
- Understand motion detection.

Humans rely largely on vision and seeing the world around them through sight. Most non human organisms rely more on the other senses to interpret the world around them. Research indicates that a large part of the cerebral cortex in humans is dedicated to vision. As the information reaches the visual cortex in the brain, multiple neurons identify shapes, colors, and motions. When light falls on the eyes, sensory receptors begin the process of transduction. As this process occurs, individuals begin to perceive the environment around them.

Think about entering a dark theater or room if you have been outside on a sunny day; often times it takes time to adjust to the change of light intensity in the environment. The muscles in the iris of the eye adjust to allow more light to enter the pupil (**the round opening in the center of the iris**). The iris of the eye is the colored part of the eye that constricts and dilates the pupil to adjust for different light intensities that we may encounter throughout the day. Photoreceptors help distinguish the different characteristics of different objects. The types of photoreceptors in the human eye include rods and cones; rods are typically more prevalent than cones. Rods give us sensitivity under dim lighting conditions and allow us to see at night. Cones allow us to see fine details in bright light and give us the ability to distinguish color. Cones are tightly packed around the fovea (the small central region located at the back of the retina) and more sparsely elsewhere. Rods populate the periphery (the region surrounding the fovea) and are almost absent from the fovea. As organisms move about their environment throughout the day the information by light from objects both near and far is encoded by the brain and the identification of colors, shapes, and motion occurs (see Figure **7.2.1**).

Anatomy of the Human Eye

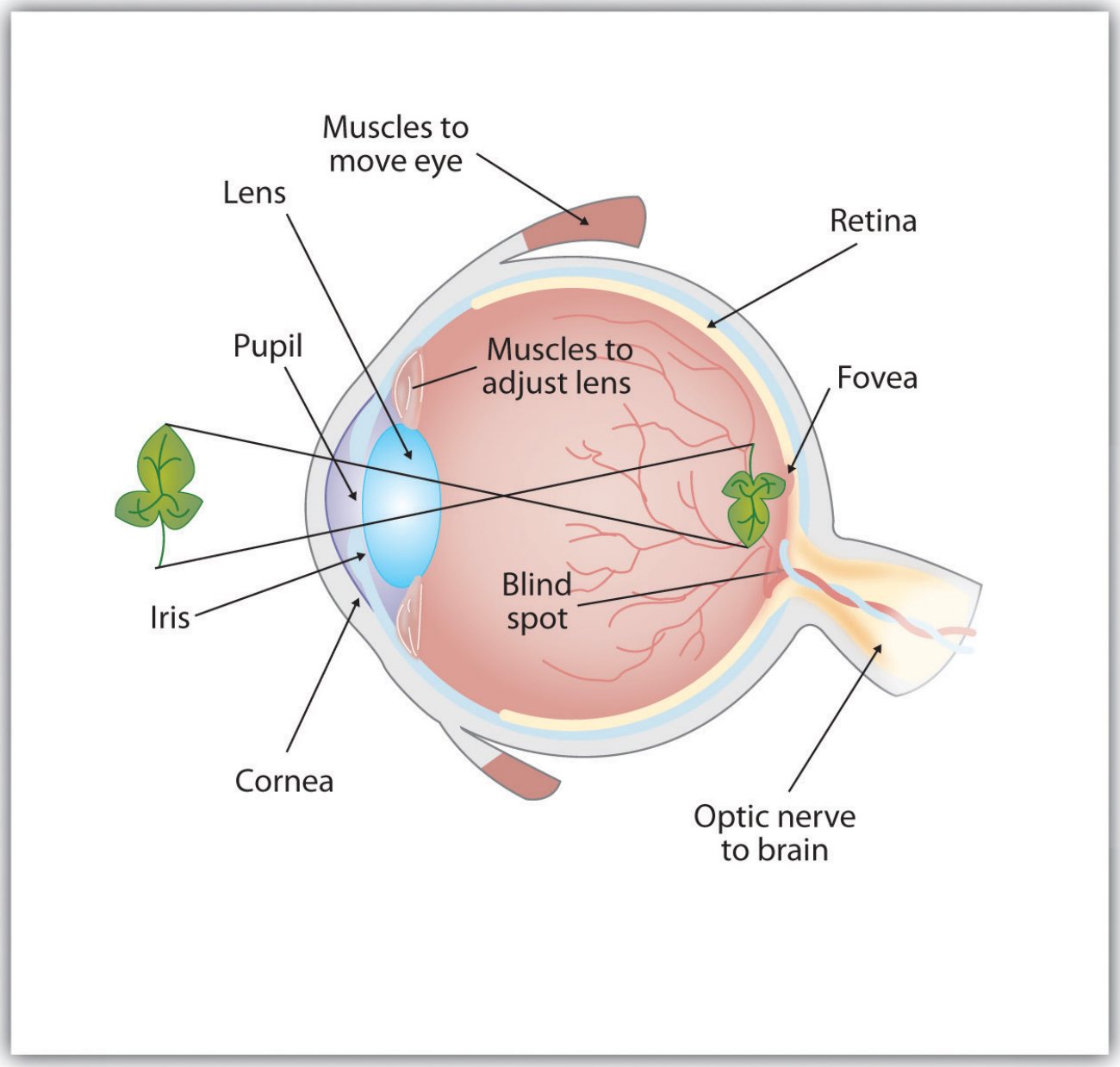


Figure 7.2.1: Light enters the eye through the transparent cornea, passing through the pupil at the center of the iris. The lens adjusts to focus the light on the retina, where it appears upside down and backward. Receptor cells on the retina send information via the optic nerve to the visual cortex. https://open.lib.umn.edu/intropsyc/c...04_s02_s01_f02

Normal, Nearsighted, and Farsighted Eyes Figure 7.2.2

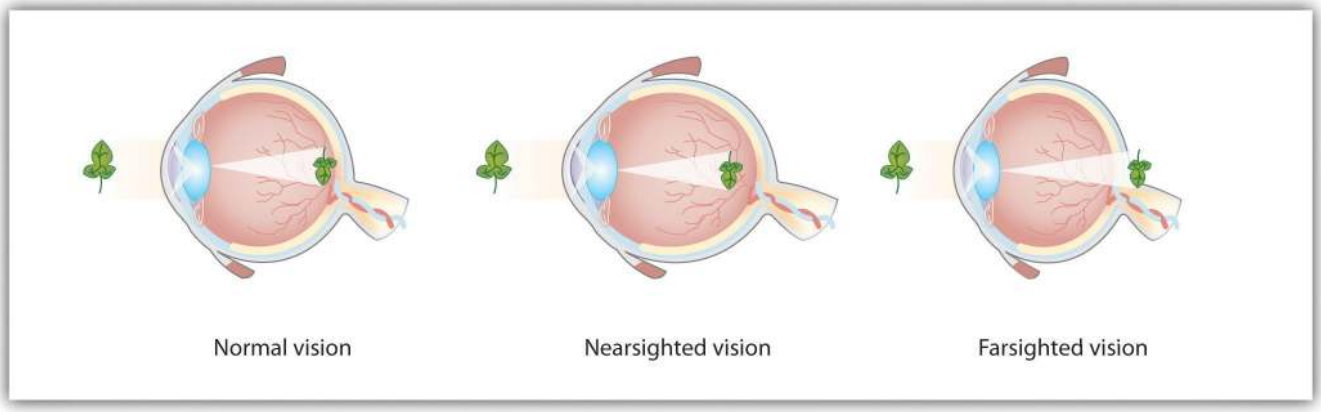


Figure **7.2.2** For people with normal vision (left), the lens properly focuses incoming light on the retina. For people who are nearsighted (center), images from far objects focus too far in front of the retina, whereas for people who are farsighted (right), images from near objects focus too far behind the retina. Eyeglasses solve the problem by adding a secondary, corrective, lens. https://open.lib.umn.edu/intropsyc/chapter/4-2-seeing/#stangor-ch04_s02_s01_f02

Research indicates that humans can distinguish between millions of different color variations (Gerald , 1972). Although humans can distinguish between a wide variety of variations there are three primary colors that form the base of these variations; red, green, and blue. The ability for these colors to form these variations is based on the hue or shade of the colors. The electromagnetic spectrum illustrated below shows that visible light for humans is just a small slice of the entire spectrum, which includes radiation that we cannot see as light because it is below the frequency of visible red light and above the frequency of visible violet light.

The Electromagnetic Spectrum Figure **7.2.3**

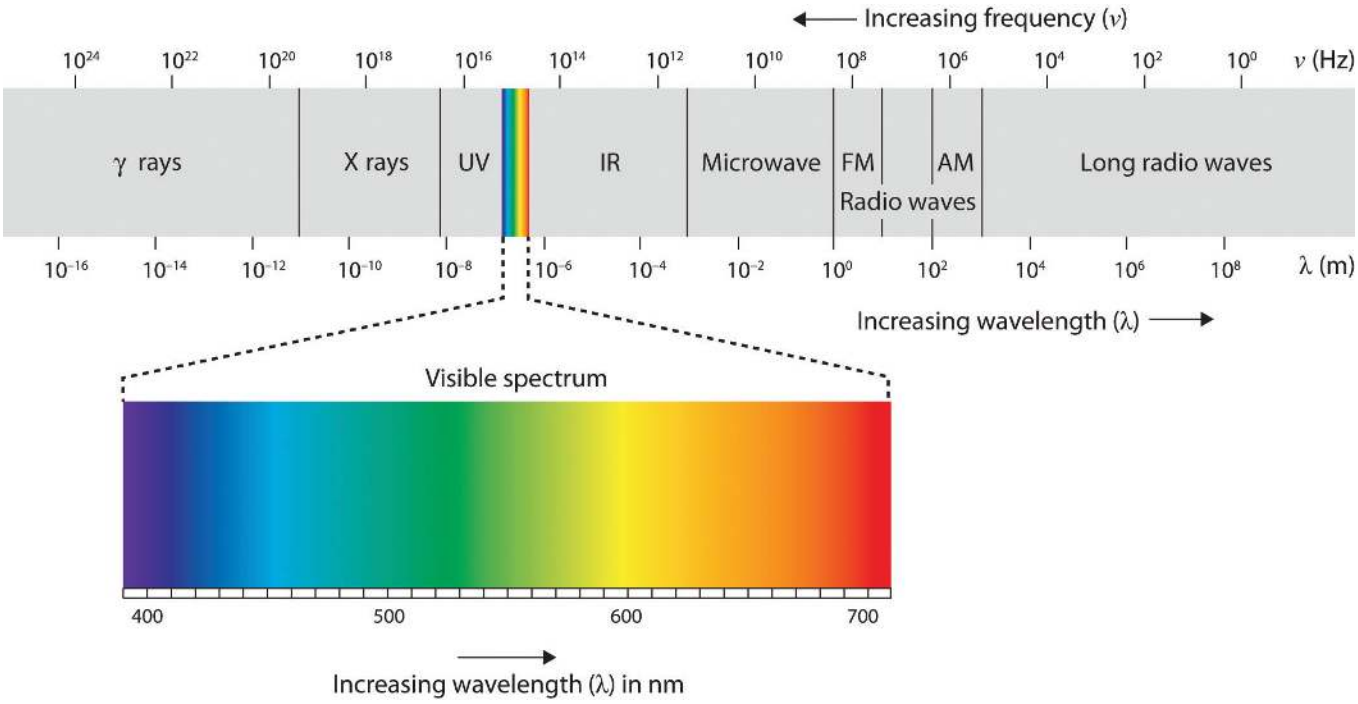


Figure **7.2.3**: Only a small fraction of the electromagnetic energy that surrounds us (the visible spectrum) is detectable by the human eye.

In his important research on color vision, Hermann von Helmholtz theorized that color is perceived because the cones in the retina come in three types. One type of cone reacts primarily to blue light (short wavelengths), another reacts primarily to green light (medium wavelengths), and a third reacts primarily to red light (long wavelengths). The visual cortex then detects and compares

the strength of the signals from each of the three types of cones, creating the experience of color. According to this Young-Helmholtz trichromatic color theory, *what color we see depends on the mix of the signals from the three types of cones*. If the brain is receiving primarily red and blue signals, for instance, it will perceive purple; if it is receiving primarily red and green signals it will perceive yellow; and if it is receiving messages from all three types of cones it will perceive white.

The different functions of the three types of cones are apparent in people who experience color blindness. Color blindness can be defined as having decreased capacity to see color or distinguish differences between colors. Red–green color blindness is the most common form, followed by blue–yellow color blindness and total color blindness. Red–green color blindness affects up to 8% of males and 0.5% of females of Northern European descent. Red-green color blindness is a result of an absence of long or medium wavelength cones or to the production of abnormal opsin pigments in these cones that affect red-green color vision (see Figure **7.2.4**).

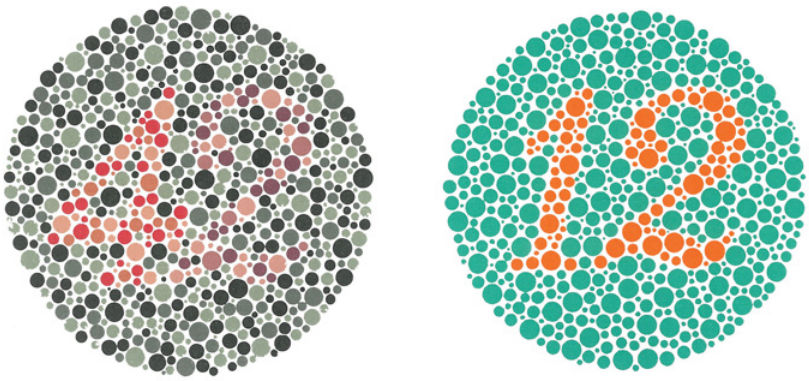


Figure **7.2.4**: People with normal color vision can see the number 42 in the first image and the number 12 in the second (they are vague but apparent). However, people who are color blind cannot see the numbers at all. Wikimedia Commons.

An alternative approach to the Young-Helmholtz theory, known as the opponent-process color theory, *proposes that we analyze sensory information not in terms of three colors but rather in three sets of "opponent colors": red-green, yellow-blue, and white-black.* Evidence for the opponent-process theory comes from the fact that some neurons in the retina and in the visual cortex are excited by one color (e.g., red) but inhibited by another color (e.g., green).

One example of opponent processing occurs in the experience of an afterimage. If you stare at the flag on the left for about 30 seconds (the longer you look, the better the effect), and then move your eyes to the blank area to the right of it, you will see the afterimage. When we stare at the green stripes, our green receptors habituate and begin to process less strongly, whereas the red receptors remain at full strength. When we switch our gaze, we see primarily the red part of the opponent process. Similar processes create blue after yellow and white after black (see Figure **7.2.5**).



Figure **7.2.5**: U.S. Flag. The presence of an afterimage is best explained by the opponent-process theory of color perception. Stare at the flag for a few seconds, and then move your gaze to the blank space next to it. Do you see the afterimage? Mike Swanson – U.S.

Flag (inverted) – public domain.

Depth Perception

Depth perception is *the ability to perceive three-dimensional space and to accurately judge distance*. Without depth perception, we would be unable to drive a car, thread a needle, or simply navigate our way around the supermarket (Howard & Rogers, 2001). Research has found that depth perception is in part based on innate capacities and in part learned through experience (Witherington, 2005).

Psychologists Eleanor Gibson and Richard Walk (1960) tested the ability to perceive depth in 6- to 14-month-old infants by placing them on a **visual cliff**, *a mechanism that gives the perception of a dangerous drop-off, in which infants can be safely tested for their perception of depth* (see Figure 7.2.6). The infants were placed on one side of the "cliff," while their mothers called to them from the other side. Gibson and Walk found that most infants either crawled away from the cliff or remained on the board and cried because they wanted to go to their mothers, but the infants perceived a chasm that they instinctively could not cross. Further research has found that even very young children who cannot yet crawl are fearful of heights (Campos, Langer, & Krowitz, 1970). On the other hand, studies have also found that infants improve their hand-eye coordination as they learn to better grasp objects and as they gain more experience in crawling, indicating that depth perception is also learned (Adolph, 2000).

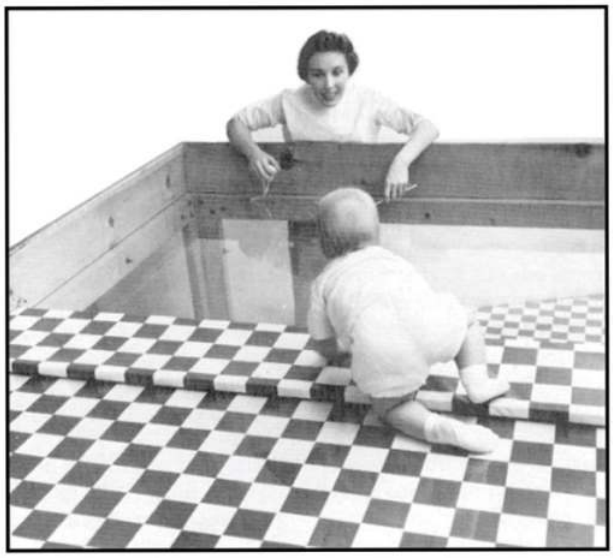


Figure 7.2.6: Visual Cliff- A mother urging her child from across the deep side of the visual cliff. Despite a transparent surface covering the cliff, the child hesitates to move forward. (From Gibson and Walk (1960). Copyright 1960 Nature Publishing Group., CC BY 4.0 <https://creativecommons.org/licenses/by/4.0>, via Wikimedia Commons)

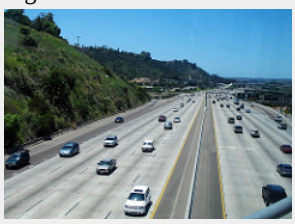
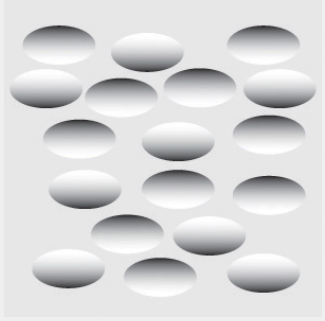
Depth perception is the result of our use of **depth cues**, *messages from our bodies and the external environment that supply us with information about space and distance*. **Binocular depth cues** are *depth cues that are created by retinal image disparity—that is, the space between our eyes, and thus which require the coordination of both eyes.* One outcome of retinal disparity is that the images projected on each eye are slightly different from each other. The visual cortex automatically merges the two images into one, enabling us to perceive depth. Three-dimensional movies make use of retinal disparity by using 3-D glasses that the viewer wears to create a different image on each eye. The perceptual system quickly, easily, and unconsciously turns the disparity into 3-D.

An important binocular depth cue is **convergence**, *the inward turning of our eyes that is required to focus on objects that are less than about 50 feet away from us*. The visual cortex uses the size of the convergence angle between the eyes to judge the object's distance. You will be able to feel your eyes converging if you slowly bring a finger closer to your nose while continuing to focus on it. When you close one eye, you no longer feel the tension—convergence is a binocular depth cue that requires both eyes to work.

The visual system also uses *accommodation* to help determine depth. As the lens changes its curvature to focus on distant or close objects, information relayed from the muscles attached to the lens helps us determine an object's distance. Accommodation is only effective at short viewing distances, however, so while it comes in handy when threading a needle or tying shoelaces, it is far less effective when driving or playing sports.

Although the best cues to depth occur when both eyes work together, we are able to see depth even with one eye closed. **Monocular depth cues** are *depth cues that help us perceive depth using only one eye* (Sekuler & Blake, 2006). Some of the most important are summarized in Table **7.2.1** "Monocular Depth Cues That Help Us Judge Depth at a Distance".

Table **7.2.1** Monocular Depth Cues That Help Us Judge Depth at a Distance

Name	Description	Example	Image
Position	We tend to see objects higher up in our field of vision as farther away.	The fence posts at right appear farther away not only because they become smaller but also because they appear higher up in the picture.	Figure  Andrew Huff – Rotted Fence – CC BY 2.0.
Relative size	Assuming that the objects in a scene are the same size, smaller objects are perceived as farther away.	At right, the cars in the distance appear smaller than those nearer to us.	Figure  Allan Ferguson – Trolley Crosses Freeway – CC BY 2.0.
Linear perspective	Parallel lines appear to converge at a distance.	We know that the tracks at right are parallel. When they appear closer together, we determine they are farther away.	Figure  Bo Insogna, TheLightningMan.com – Lightning Striking By The Train Tracks – CC BY-NC-ND 2.0.
Light and shadow	The eye receives more reflected light from objects that are closer to us. Normally, light comes from above, so darker images are in shadow.	We see the images at right as extending and indented according to their shadowing. If we invert the picture, the images will reverse.	Figure 

Name	Description	Example	Image
Interposition	When one object overlaps another object, we view it as closer.	At right, because the blue star covers the pink bar, it is seen as closer than the yellow moon.	Figure 
Aerial perspective	Objects that appear hazy, or that are covered with smog or dust, appear farther away.	The artist who painted the picture on the right used aerial perspective to make the distant hills more hazy and thus appear farther away.	Figure  Frans Koppelaar – Landscape near Bologna – CC BY-SA 2.5.

Motion

Many animals, including human beings, have very sophisticated perceptual skills that allow them to coordinate their own motion with the motion of moving objects in order to create a collision with that object. Bats and birds use this mechanism to catch up with prey, dogs use it to catch a Frisbee, and humans use it to catch a moving football. The brain detects motion partly from the changing size of an image on the retina (objects that look bigger are usually closer to us) and in part from the relative brightness of objects.

We also experience motion when objects near each other change their appearance. The **beta effect** refers to *the perception of motion that occurs when different images are presented next to each other in succession* (see Note "Beta Effect and Phi Phenomenon"). The visual cortex fills in the missing part of the motion and we see the object moving. The beta effect is used in movies to create the experience of motion. A related effect is the **phi phenomenon**, in which *we perceive a sensation of motion caused by the appearance and disappearance of objects that are near each other*. The phi phenomenon looks like a moving zone or cloud of background color surrounding the flashing objects. The beta effect and the phi phenomenon are other examples of the importance of the gestalt—our tendency to "see more than the sum of the parts."

7.2.1.0.1 Beta Effect and Phi Phenomenon

In the beta effect, our eyes detect motion from a series of still images, each with the object in a different place. This is the fundamental mechanism of motion pictures (movies). In the phi phenomenon, the perception of motion is based on the momentary hiding of an image.

Phi phenomenon: http://upload.wikimedia.org/Wikipedia/commons/6/6e/Lilac-Chaser.gif

Beta effect: http://upload.wikimedia.org/Wikipedia/commons/0/09/Phi_phenomenom_no_watermark.gif

7.2.2 Attribution

- Vision sections adapted by Isaias Hernandez from "NOBA-Vision"; https://nobaproject.com/modules/vision
- Vision sections adapted by Isaias Hernandez from "Psychology-Penn State"; https://psu.pb.unizin.org/intropsych...eption-vision/

- Vision sections adapted by Isaias Hernandez from "Introduction to Psychology-University of Minnesota"; https://open.lib.umn.edu/intropsyc/c...04_s02_s01_f02

7.2.3 References

https://www.nei.nih.gov/learn-about-...olor-blindness

This page titled 7.2: Vision is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

7.3: Hearing

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

7.3.1 Hearing

Learning Objectives

1. Label key structures of the ear, identify their functions, and describe the role they play in hearing.
2. Explain how we encode and perceive pitch.
3. Explain how we localize sound.
4. Describe how hearing and the vestibular system are associated.

This section will provide an overview of the basic anatomy and function of the auditory system, how we perceive pitch, and how we know where sound is coming from. It has been argued that the human ear can hear a variety of sounds and can distinguish between a large variation of these sounds. The complex anatomy of the ear in humans can be divided into three sections (the three sections are shown in the image below). The neural impulses of these sound waves are sent to the brain and our past experiences integrate with those sound waves to help us make sense of the sounds we are hearing. The auditory system converts the sound waves into electrical signals that the brain interprets. The illustration below shows the complex system of the ear and how the anatomical structures lead to the ability to hear the sounds of nature, appreciate beauty of music and utilize language to communicate with others who speak the same language (see Figure **7.3.1**).

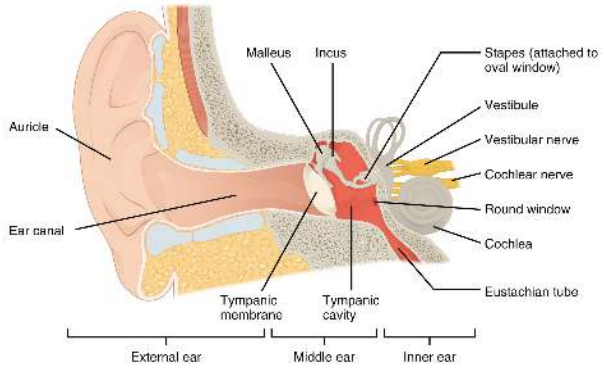


Figure **7.3.1** – ***Structures of the Ear:*** *The external ear contains the auricle, ear canal, and tympanic membrane. The middle ear contains the ossicles and is connected to the pharynx by the Eustachian tube. The inner ear contains the cochlea and vestibule, which are responsible for audition and equilibrium, respectively* https://open.oregonstate.education/a.../15-3-hearing/.

In particular, the human ear is most responsive to sounds that are in the same frequency as the human voice. This is why parents and mothers in particular are able to pick out the sound of their children's voice amongst other children's voices and we are often able to identify another person from the sound of their voice without having to see them physically. The complex system of the ear allows us to process sounds almost instantly.

Unlike light waves which travel in a vacuum, sound waves are transferred when molecules bump into each other in the air and produce sound waves that allow us to identify the source of the sounds that we encounter. These sound waves create different frequencies, with low frequency sounds being lower pitched and higher frequency sounds being higher pitched. There are a few theories that have been proposed to help account as to why individuals can distinguish between pitch perception and frequencies.

The **temporal theory** of pitch perception asserts that frequency is coded by the activity level of a sensory neuron. This would mean that a given hair cell would fire action potentials related to the frequency of the sound wave. While this is a very intuitive explanation, we detect such a broad range of frequencies (20–20,000 Hz) that the frequency of action potentials fired by hair cells cannot account for the entire range. Because of properties related to sodium channels on the neuronal membrane that are involved in action potentials, there is a point at which a cell cannot fire any faster (Shamma, 2001).

The **place theory** of pitch perception suggests that different portions of the basilar membrane **(a stiff structural element within the cochlea of the inner ear)** moves up and down in response to incoming sound waves. More specifically, the base of the basilar membrane responds best to high frequencies and the tip of the basilar membrane responds best to low frequencies. Therefore, hair cells that are in the base portion would be labeled as high-pitch receptors, while those in the tip of basilar membrane would be labeled as low-pitch receptors (Shamma, 2001). In reality, both theories explain different aspects of pitch perception. At frequencies up to about 4000 Hz, it is clear that both the rate of action potentials and place contribute to our perception of pitch. However, much higher frequency sounds can only be encoded using place cues (Shamma, 2001).

Similar to the need for recognizing different pitches and frequencies, knowing where particular sounds are coming from (sound localization) is an important part of navigating the environment around us. The auditory system has the ability to use monaural (one ear) and binaural (two ears) cues to locate where a particular sound might be coming from. Each pinna interacts with incoming sound waves differently, depending on the sound's source relative to our bodies. This interaction provides a monaural cue that is helpful in locating sounds that occur above or below and in front or behind us. The sound waves received by your two ears from sounds that come from directly above, below, in front, or behind you would be identical; therefore, monaural cues are essential (Grothe, Pecka, & McAlpine, 2010).

Binaural cues, on the other hand, provide information on the location of a sound along a horizontal axis by relying on differences in patterns of vibration of the eardrum between our two ears. If a sound comes from an off-center location, it creates two types of binaural cues: interaural level differences and interaural timing differences. **Interaural level difference** refers to the fact that a sound coming from the right side of your body is more intense at your right ear than at your left ear because of the attenuation of the sound wave as it passes through your head. **Interaural timing difference** refers to the small difference in the time at which a given sound wave arrives at each ear, illustrated in Figure **7.3.2**. Certain brain areas monitor these differences to construct where along a horizontal axis a sound originates (Grothe et al., 2010).

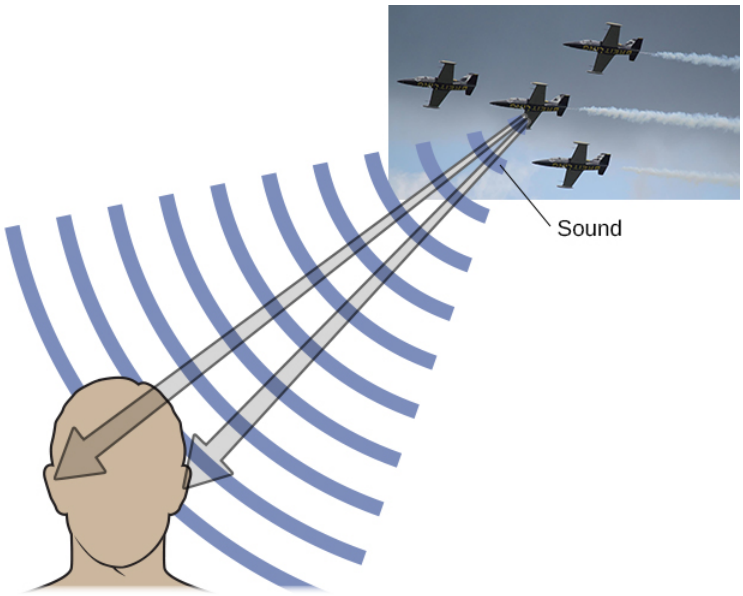


Figure **7.3.2**: Localizing sound involves the use of both monaural and binaural cues. (credit "plane": modification of work by Max Pfandl)

7.3.1.1 Hearing Loss

Hearing Loss

- About 2 to 3 out of every 1,000 children in the United States are born with a detectable level of hearing loss in one or both ears (https://www.nidcd.nih.gov/health/sta...istics-hearing).
- More than 90 percent of deaf children are born to hearing parents (https://www.nidcd.nih.gov/health/sta...istics-hearing).
- Approximately 15% of American adults (37.5 million) aged 18 and over report some trouble hearing (https://www.nidcd.nih.gov/health/statistics/quick-statistics-hearing)

Conductive hearing loss can be caused by physical damage to the ear (such as to the eardrums); this condition reduces the ability of the ear to transfer vibrations from the outer ear to the inner ear. Conductive hearing loss can also be a result of fusion of the ossicles **(three bones in the middle ear)**. Sensorineural hearing loss, which is caused by damage to the cilia (of the hair cells) or to the auditory nerve, is not as common as conductive hearing loss but the likelihood of this condition increases with age (Tennesen, 2007). As we continue to get older damage to the cilia increases; by the age of 65 years old 40% of individuals will have had damage to the cilia (Chisolm, Willott, & Lister, 2003).

Individuals who have experienced sensorineural hearing loss may benefit from a cochlear implant. Data From the National Institutes of Health shows that as of December 2019, approximately 736,900 cochlear implants have been implanted worldwide. In the United States, roughly 118,100 devices have been implanted in adults and 65,000 in children The following video explains the process of a cochlear implant.

Cochlear implant surgeries and how they work:

https://www.youtube.com/watch?v=AqXBrKwB96E

Deafness and Deaf Culture

In most modern nations people who are born or become deaf at an early age have developed their own system of communication and culture amongst themselves and people close to them. It has been argued that encouraging deaf people to sign is a more appropriate adjustment as opposed to encouraging them to speak, read lips or have cochlear implant surgeries. However, more recent studies suggest that due to advancements in technology, cochlear implants increase the likelihood of a person being able to have and engage in some auditory and speaking activities if implanted early enough (Dettman, Pinder, Briggs, Dowell, & Leigh, 2007; Dorman & Wilson, 2004). As a result parents often face the difficult decision of whether to take advantage of new technologies and approaches of providing support to deaf students in mainstream classroom settings or utilizing American Sign Language (ASL) schools and encouraging more immersion in those settings.

Hearing and the Vestibular System

The vestibular system has some similarities with the auditory system. It utilizes hair cells just like the auditory system, but it excites them in different ways. There are five vestibular receptor organs in the inner ear: the utricle, the saccule, and three semicircular canals. Together, they make up what's known as the vestibular labyrinth that is shown in Figure **7.3.3**. The utricle and saccule respond to acceleration in a straight line, such as gravity. The roughly 30,000 hair cells in the utricle and 16,000 hair cells in the saccule lie below a gelatinous layer, with their stereocilia projecting into the gelatin. Embedded in this gelatin are calcium carbonate crystals—like tiny rocks. When the head is tilted, the crystals continue to be pulled straight down by gravity, but the new angle of the head causes the gelatin to shift, thereby bending the stereocilia. The bending of the stereocilia stimulates the neurons, and they signal to the brain that the head is tilted, allowing the maintenance of balance. It is the vestibular branch of the vestibulocochlear cranial nerve that deals with balance.

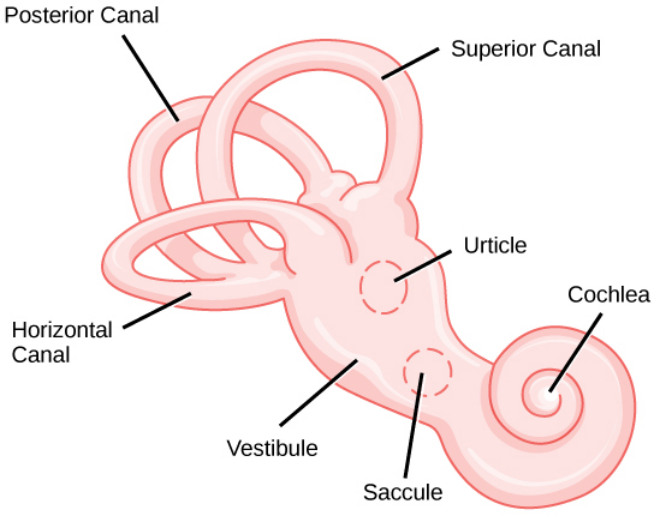


Figure **7.3.3** The structure of the vestibular labyrinth is shown. (credit: modification of work by NIH)

7.3.2 Attributions

- Hearing sections adapted by Isaias Hernandez from "Anatomy and Physiology from Oregon State University"; https://open.oregonstate.education/aandp/chapter/15-3-hearing/
- Hearing sections adapted by Isaias Hernandez from "Psychology 2e from OpenStax"; https://openstax.org/books/psychology-2e/pages/5-4-hearing

7.3.3 References

https://www.nidcd.nih.gov/health/sta...istics-hearing

This page titled 7.3: Hearing is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

7.4: Touch

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

7.4.1 Touch

Learning Objectives

1. Describe the role of social and affective touch in development and bonding.

2. Describe how the somatosensory system allows for the sensing of touch, temperature, and position in space.

3. Describe the process of transduction in the senses of touch and proprioception.

4. Explain how expectations and context affect pain and touch experiences.

The sense of touch is one of the most important senses in human development, particularly with regard to attachment to others. It has long been viewed as an essential part of childhood development (Baysinger, Plubell, & Harlow, 1973; Feldman, 2007; Haradon, Bascom, Dragomir, & Scripcaru, 1994) and also helps us feel socially attached to others (Field et al., 1997; Kelter, 2009). Touch is an essential part of everyday activities ranging from touching our food to eat to sharing human affection with another individual. Recent studies indicate that slow affectionate touch, rather than faster neutral touch, can serve as a buffer against physical pain. In addition, the same slow affectionate touch can reduce feelings of social exclusion and increase social bonding.

Skin is the largest organ in our body and has a variety of nerve endings that allow us to feel sensations such as pressure, pain, cold, and heat. The skin on our bodies also serves as a protector to important vital organs.

Touch and pain are aspects of the somatosensory system, which provides our brain with information about our own body (**interoception**) and properties of the immediate external world (**exteroception**) (Craig, 2002). We have somatosensory receptors located all over the body, from the surface of our skin to the depth of our joints. The information they send to the central nervous system is generally divided into four modalities: **cutaneous senses** (senses of the skin), **proprioception** (body position), **kinesthesis** (body movement), and **nociception** (pain, discomfort).

Cutaneous Senses (senses of the skin)

Cutaneous senses respond to tactile, thermal, and pruritic (itchy) stimuli, and events that cause tissue damage (hence pain). Cutaneous mechanoreceptors are located in the skin. They provide the senses of touch, pressure, vibration, proprioception, and others. They are all innervated by Aβ fibers (**Aβ fibers from the skin are mostly dedicated to touch)**, except the mechanoreceiving free nerve endings, which are innervated by Aδ fibers **(Aδ fibers serve to receive and transmit information primarily relating to acute pain; sharp, immediate, and relatively short-lasting)**.

The can be categorized by morphology, by the type of sensation they perceive, and by the rate of adaptation. Furthermore, each has a different receptive field:

- Ruffini's end organs detect tension deep in the skin.
- Meissner's corpuscles detect changes in texture (vibrations around 50 Hz) and adapt rapidly.
- Pacinian corpuscles detect rapid vibrations (about 200–300 Hz).
- Merkel's discs detect sustained touch and pressure.
- Mechanoreceiving free nerve endings detect touch, pressure, and stretching.
- Hair follicle receptors are located in hair follicles and sense the position changes of hair strands.

Ruffini Ending

The Ruffini ending (Ruffini corpuscle or bulbous corpuscle) is a class of slowly adapting mechanoreceptors thought to exist only in the glabrous dermis and subcutaneous tissue of humans. It is named after Angelo Ruffini (an Italian histologist and embryologist).

This spindle-shaped receptor is sensitive to skin stretch, and contributes to the kinesthetic sense of and control of finger position and movement. It is believed to be useful for monitoring the slippage of objects along the surface of the skin, allowing the

modulation of grip on an object.

Ruffini endings are located in the deep layers of the skin. They register mechanical information within joints, more specifically angle change, with a specificity of up to two degrees, as well as continuous pressure states. They also act as thermoreceptors that respond for a long time, such as holding hands with someone during a walk. In a case of a deep burn to the body, there will be no pain as these receptors will be burned off.

Meissner's Corpuscles

Meissner's corpuscles (or tactile corpuscles) are responsible for sensitivity to light touch. In particular, they have the highest sensitivity (lowest threshold) when sensing vibrations lower than 50 hertz. They are rapidly adaptive receptors.

Pacinian Corpuscles

Pacinian corpuscles (or lamellar corpuscles) are responsible for sensitivity to vibration and pressure. The vibrational role may be used to detect surface texture, e.g., rough versus smooth.

Merkel Nerve

Merkel nerve endings are mechanoreceptors found in the skin and mucosa of vertebrates that provide touch information to the brain. The information they provide are those regarding pressure and texture. Each ending consists of a Merkel cell in close apposition with an enlarged nerve terminal.

This is sometimes referred to as a Merkel cell–neurite complex, or a Merkel disk receptor. A single afferent nerve fiber branches to innervate up to 90 such endings. They are classified as slowly adapting type I mechanoreceptors.

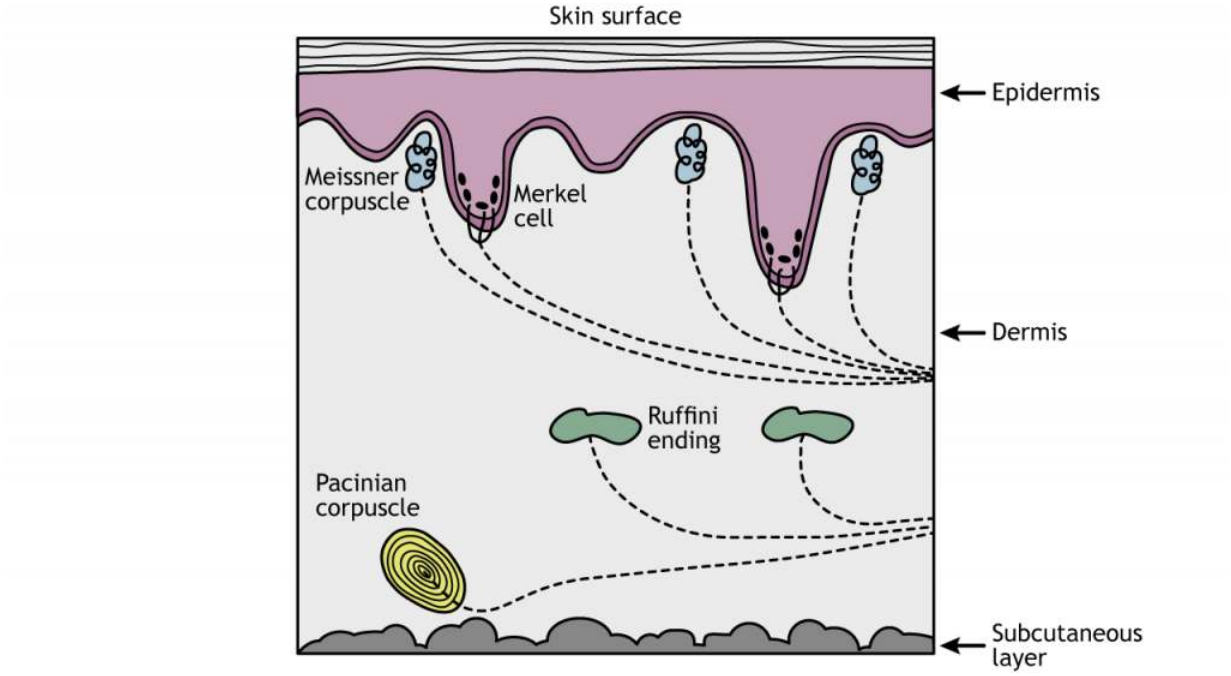


Figure **7.4.1**: *The different mechanoreceptor types are located in different regions of the skin and are responsible for perception of different characteristics of a touch stimulus. Pacinian corpuscles and Ruffini endings are located deep in the dermis. Meissner corpuscles are located in the dermis near the epidermis, and Merkel cells are located in the epidermis, near the surface of the skin. 'Mechanoreceptors' by Casey Henley is licensed under a Creative Commons Attribution Non-Commercial Share-Alike (CC BY-NC-SA) 4.0 International License.*

Proprioception (body position)

Proprioception refers to the sense of knowing how one's body is positioned in three-dimensional space. Proprioception is the sense of the relative position of neighboring parts of the body and the strength of effort being employed in movement. It is distinguished from exteroception (perception of the outside world) and interoception (perception of pain, hunger, and the movement of internal organs). The initiation of proprioception is the activation of a proprioreceptor in the periphery. The proprioceptive sense is believed to be composed of information from sensory neurons located in the inner ear (motion and orientation) and in the stretch receptors

located in the muscles and the joint-supporting ligaments (stance). Conscious proprioception is communicated by the posterior (dorsal) column–medial lemniscus pathway to the cerebrum. Unconscious proprioception is communicated primarily via the dorsal and ventral spinocerebellar tracts to the cerebellum.

An unconscious reaction is seen in the human proprioceptive reflex, or Law of Righting, a reflex that corrects the orientation of the body when it is taken out of its normal upright position. In the event that the body tilts in any direction, the person will tilt their head back to level the eyes against the horizon. Infants are often seen doing this as soon as they gain control of their neck muscles. This control comes from the cerebellum, the part of the brain that affects balance.

Muscle spindles are sensory receptors within the belly of a muscle that primarily detect changes in the length of a muscle. They convey length information to the central nervous system via sensory neurons. This information can be processed by the brain to determine the position of body parts. The responses of muscle spindles to changes in length also play an important role in regulating the contraction of muscles.

The Golgi organ (also called Golgi tendon organ, tendon organ, neurotendinous organ or neurotendinous spindle) is a proprioceptive sensory receptor that provides the sensory component of the Golgi tendon reflex. The Golgi organ should not be confused with the Golgi apparatus—an organelle in the eukaryotic cell —or the Golgi stain, which is a histologic stain for neuron cell bodies. The Golgi tendon reflex is a normal component of the reflex arc of the peripheral nervous system. In a Golgi tendon reflex, skeletal muscle contraction causes the agonist muscle to simultaneously lengthen and relax. This reflex is also called the inverse myotatic reflex, because it is the inverse of the stretch reflex. Although muscle tension is increasing during the contraction, alpha motor neurons in the spinal cord that supply the muscle are inhibited and antagonistic muscles are activated.

Kinesthesis (body movement)

Kinesthesis or Kinesthesia is a term that is often used interchangeably with proprioception. Some users differentiate the kinesthetic sense from proprioception by excluding the sense of equilibrium or balance from kinesthesia. An inner ear infection, for example, might degrade the sense of balance. This would degrade the proprioceptive sense, but not the kinesthetic sense. The infected person would be able to walk, but only by using their sense of sight to maintain balance; the person may have difficulty walking when their eyes are closed.

Proprioception and kinesthesia are seen as interrelated and there is considerable disagreement regarding the definition of these terms. Some of this difficulty stems from Sherrington's original description of joint position sense (or the ability to determine where a particular body part exactly is in space) and kinesthesia (or the sensation that the body part has moved) under a more general heading of proprioception. Clinical aspects of proprioception are measured in tests that measure a subject's ability to detect an externally imposed passive movement, or the ability to reposition a joint to a predetermined position. Often it is assumed that the ability of one of these aspects will be related to another, unfortunately experimental evidence suggests there is no strong relation between these two aspects. This suggests that while these components may well be related in a cognitive manner, they seem to be separate physiologically.

Much of the forgoing work is dependent on the notion that proprioception is essentially a feedback mechanism: that is the body moves (or is moved) and then the information about this is returned to the brain whereby subsequent adjustments could be made. More recent work into the mechanism of ankle sprains suggest that the role of reflexes may be more limited due to their long latencies (even at the spinal cord level) as ankle sprain events occur in perhaps 100 milliseconds or less. As a result, a model has been proposed to include a 'feedforward' component of proprioception where the subject will also have central information about the body's position prior to attaining it.

Kinesthesia is a key component in **muscle memory** and **hand-eye coordination** and training can improve this sense. The ability to swing a golf club, or to catch a ball, requires a finely-tuned sense of the position of the joints. This sense needs to become automatic through training to enable a person to concentrate on other aspects of performance, such as maintaining motivation or seeing where other people are.

Nociception (pain, discomfort)

Nociception deals with a series of events and processes required for an organism to receive a painful stimulus, convert it to a molecular signal, and recognize and characterize the signal in order to trigger an appropriate defense response. This serves as an important function because our brains can often turn off or reduce the feelings of pain depending on the situation we may be

experiencing or involved in. An example involves athletes who often do not feel pain until after they have completed the game or event they are participating in (Bantick, Wise, Ploghaus, Clare, Smith, & Tracey, 2002). In essence their brain is involved in a complicated process that requires them to utilize multiple systems to stay focused on the activity. The endorphins that are released when we are engaged or excited by a situation or event around us act as natural pain killers (Sternberg, Bailin, Grant, & Gracely, 1998).

The Primary Somatosensory Cortex

The primary somatosensory cortex is divided into four regions, each with its own input and function: areas 1, 2, 3a, and 3b. Most touch information from mechanoreceptors inputs to region 3b, whereas most proprioceptive information from the muscles inputs to region 3a. These regions then send and receive information from areas 1 and 2. As processing of somatosensory information continues, the stimuli required to activate neurons becomes more complex. For example, area 1 is involved in sensing texture, and area 2 is involved in sensing size and shape of an object. The posterior parietal cortex, an important output region of the somatosensory cortex, lies caudal to the postcentral gyrus; areas 5 and 7 are downstream structures that continue to process touch (see Figure **7.4.2**).

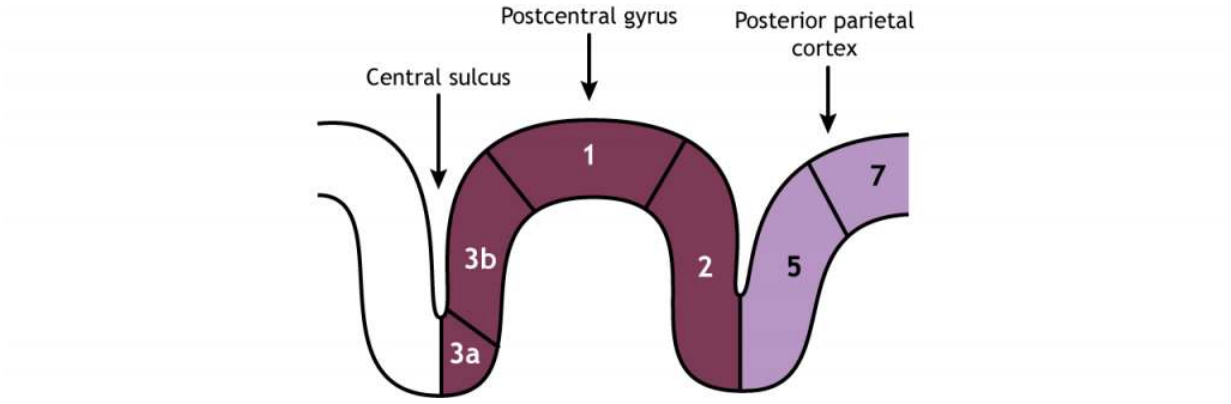


Figure **7.4.2**: *The somatosensory cortex, located in the postcentral gyrus, just posterior to the central sulcus, is divided into 4 areas: 3a, 3b, 1, and 2. The posterior parietal cortex, an output region of the somatosensory cortex, lies just posterior to the postcentral gyrus and is divided into areas 5 and 7. 'Postcentral Gyrus' by Casey Henley is licensed under a Creative Commons Attribution Non-Commercial Share-Alike (CC BY-NC-SA) 4.0 International License.*

Testing of the senses begins with examining the regions known as dermatomes that connect to the cortical region where somatosensation is perceived in the postcentral gyrus. To test the sensory fields, a simple stimulus of the light touch of the soft end of a cotton-tipped applicator is applied at various locations on the skin. The spinal nerves, which contain sensory fibers with dendritic endings in the skin, connect with the skin in a topographically organized manner, illustrated as dermatomes (Figure **7.4.3** - it is not necessary to memorize the terms in this figure). For example, the fibers of the eighth cervical nerve innervate the medial surface of the forearm and extend out to the fingers. In addition to testing perception at different positions on the skin, it is necessary to test sensory perception within the dermatome from distal to proximal locations in the appendages, or lateral to medial locations in the trunk. In testing the eighth cervical nerve, the patient would be asked if the touch of the cotton to the fingers or the medial forearm was perceptible, and whether there were any differences in the sensations.

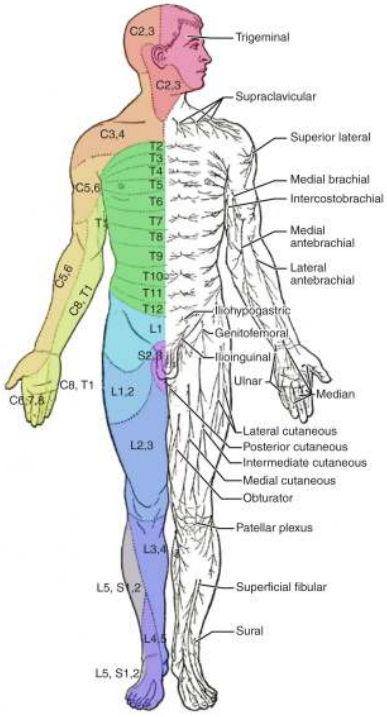


Figure **7.4.3** – ***Dermatomes:*** *The surface of the skin can be divided into topographic regions that relate to the location of sensory endings in the skin based on the spinal nerve that contains those fibers. (credit: modification of work by Mikael Häggström)* *https://open.oregonstate.education/a...otor-pathways/*

7.4.2 Attributions

- Touch sections adapted by Isaias Hernandez from "NOBA-Biological Psychology, Touch and Pain"; https://nobaproject.com/modules/touch-and-pain#:~:text=Touch%20and%20pain%20are%20aspects,)%20(Craig%2C%202002).
- Touch sections adapted by Isaias Hernandez from "Introduction to Psychology, University of Minnesota'"; https://open.lib.umn.edu/intropsyc/chapter/4-4-tasting-smelling-and-touching/
- Touch sections adapted by Isaias Hernandez from "Boundless Anatomy and Physiology from Lumen"; https://courses.lumenlearning.com/boundless-ap/chapter/pain/
- Touch sections adapted by Isaias Hernandez from "Kinesthia, Psychology Wiki"; https://en.Wikipedia.org/wiki/Kinaesthetics
- Touch sections adapted by Isaias Hernandez from "Nociception, Wikipedia"; https://en.Wikipedia.org/wiki/Nociception
- Touch sections adapted by Isaias Hernandez from "Foundations of Neuroscience, Michigan State University"; https://openbooks.lib.msu.edu/neuroscience/chapter/touch-the-skin/
- Touch sections adapted by Isaias Hernandez from "Anatomy and Physiology, Oregon State University"; https://open.oregonstate.education/aandp/chapter/14-5-sensory-and-motor-pathways/

7.4.3 References

- von Mohr, M., Kirsch, L.P. & Fotopoulou, A. The soothing function of touch: affective touch reduces feelings of social exclusion. *Sci Rep* **7,** 13516 (2017). https://doi.org/10.1038/s41598-017-13355-7

This page titled 7.4: Touch is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

7.5: Taste and Smell

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

7.5.1 Taste and Smell

Learning Objectives

1. Explain why taste and smell are the most interconnected senses
2. Identify the five primary tastes in humans: sweet, sour, bitter, salty, and umami
3. Identify how the five primary tastes allow individuals to adapt to changes in the environment and increase chances of survival
4. Explain how olfactory receptors are responsive to different odorants

Being able to sense chemicals in the environment through taste (gustation) and smell (olfaction) can help an organism find food, avoid poisons, and attract mates. Humans can perceive five basic tastes: salty, sour, bitter, sweet, and umami. Bitter taste often indicates a dangerous substance like a poison, sweet taste signifies a high energy food, salty taste indicates a substance with high salt content, sour taste indicates an acidic food, and umami taste indicates a high protein food.

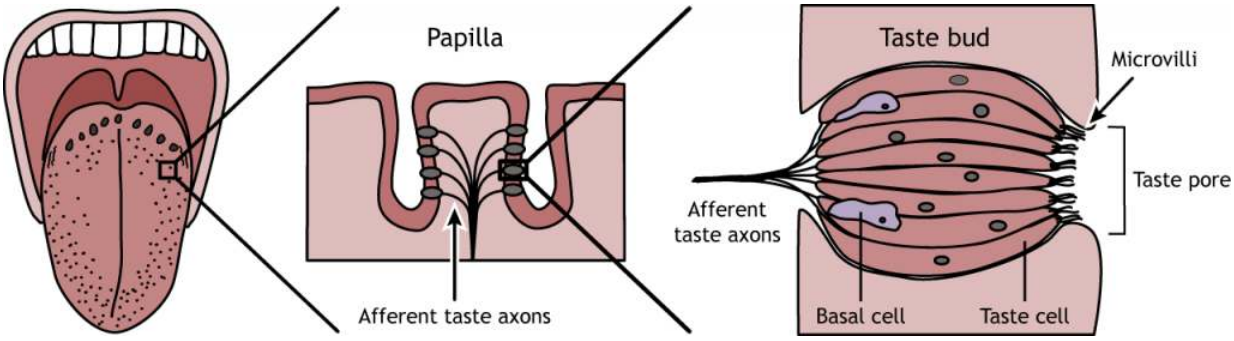


Figure **7.5.1**: *The visible bumps on the surface of the tongue are papillae that house taste buds. Taste buds are made up of taste cells and basal cells. The taste cells synapse on afferent axons that send information to the central nervous system. Tastants in food access the taste cells via the taste pore, where the food particles interact with the microvilli of the taste cells. 'Tongue Anatomy' by Casey Henley is licensed under a Creative Commons Attribution Non-Commercial Share-Alike (CC BY-NC-SA) 4.0 International License.*

Taste and smell, are the most interconnected senses in that both involve molecules of the stimulus entering the body and bonding to receptors. Smell lets an animal sense the presence of food and other chemicals in the environment that can impact their survival. Similarly, the sense of taste allows animals to discriminate between types of foods. Different tasting foods have different attributes, both helpful and harmful. For example, sweet-tasting substances tend to be highly caloric, which could be necessary for survival in lean times. Bitterness is often associated with toxicity, and sourness is often associated with spoiled food. Salty foods are valuable in maintaining homeostasis by helping the body retain water and by providing ions necessary for proper cell function.

Smell

Odorants (odor molecules) enter the nose and dissolve in the olfactory epithelium, the mucosa at the back of the nasal cavity (as illustrated in the figure below). The **olfactory epithelium** is a collection of specialized olfactory receptors in the back of the nasal cavity that spans an area about 5 cm^2 in humans. Recall that sensory cells are neurons. An **olfactory receptor**, which is a dendrite of a specialized neuron, responds when it binds certain molecules inhaled from the environment by sending impulses directly to the olfactory bulb of the brain. Humans have about 12 million olfactory receptors, distributed among hundreds of different receptor types that respond to different odors. Twelve million seems like a large number of receptors, but compare that to other animals: rabbits have about 100 million, most dogs have about 1 billion, and bloodhounds—dogs selectively bred for their sense of smell—have about 4 billion. The overall size of the olfactory epithelium also differs between species, with that of bloodhounds, for example, being many times larger than that of humans.

Olfactory neurons are **bipolar neurons** (neurons with two processes from the cell body). Each neuron has a single dendrite buried in the olfactory epithelium, and extending from this dendrite are 5 to 20 receptor-laden, hair-like cilia that trap odorant molecules. The sensory receptors on the cilia are proteins, and it is the variations in their amino acid chains that make the receptors sensitive to different odorants. Each olfactory sensory neuron has only one type of receptor on its cilia, and the receptors are specialized to detect specific odorants, so the bipolar neurons themselves are specialized. When an odorant binds with a receptor that recognizes it, the sensory neuron associated with the receptor is stimulated. Olfactory stimulation is the only sensory information that directly reaches the cerebral cortex, whereas other sensations are relayed through the thalamus.

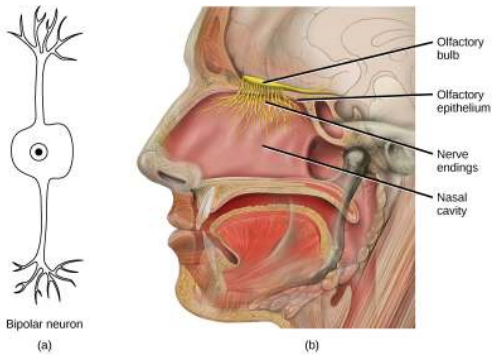


Figure **7.5.2**: In the human olfactory system, (a) bipolar olfactory neurons extend from (b) the olfactory epithelium, where olfactory receptors are located, to the olfactory bulb. (credit: modification of work by Patrick J. Lynch, medical illustrator; C. Carl Jaffe, MD, cardiologist)

Taste

Detecting a taste (gustation) is fairly similar to detecting an odor (olfaction), given that both taste and smell rely on chemical receptors being stimulated by certain molecules. The primary organ of taste is the taste bud. A **taste bud** is a cluster of gustatory receptors (taste cells) that are located within the bumps on the tongue called **papillae** (singular: papilla) (illustrated in the figure below). There are several structurally distinct papillae. Filiform papillae, which are located across the tongue, are tactile, providing friction that helps the tongue move substances, and contain no taste cells. In contrast, fungiform papillae, which are located mainly on the anterior two-thirds of the tongue, each contain one to eight taste buds and also have receptors for pressure and temperature. The large circumvallate papillae contain up to 100 taste buds and form a V near the posterior margin of the tongue.

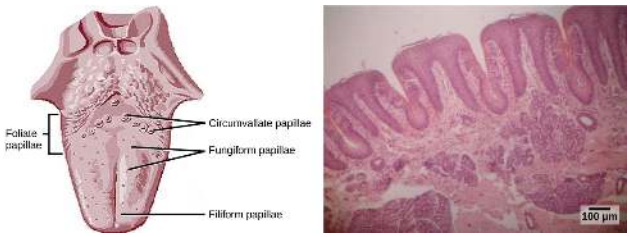


Figure **7.5.3**: The human tongue. (a) Foliate, circumvallate, and fungiform papillae are located on different regions of the tongue. (b) Foliate papillae are prominent protrusions on this light micrograph. (credit a: modification of work by NCI; scale-bar data from Matt Russell)

Each taste bud's taste cells are replaced every 10 to 14 days. These are elongated cells with hair-like processes called microvilli at the tips that extend into the taste bud pore. Food molecules (**tastants**) are dissolved in saliva, and they bind with and stimulate the receptors on the microvilli. The receptors for tastants are located across the outer portion and front of the tongue, outside of the middle area where the filiform papillae are most prominent.

In humans, there are five primary tastes, each receptor is specific to its stimulus (tastant). Transduction of the five tastes happens through different mechanisms that reflect the molecular composition of the tastant. These tastants bind to their respective receptors, thereby exciting the specialized neurons associated with them. Tasting abilities and sense of smell change with age. In humans, the senses decline dramatically by age 50 and continue to decline. A child may find a food to be too spicy, whereas an elderly person may find the same food to be bland and unappetizing.

7.5.2 Attributions

- Taste and Smell sections adapted by Isaias Hernandez from "Human Biology, OpenStax"; https://openstax.org/books/biology-2e/pages/36-3-taste-and-smell
- Taste and Smell sections adapted by Isaias Hernandez from "Foundations of Neuroscience, Michigan State University"; https://openbooks.lib.msu.edu/neuroscience/chapter/taste/

This page titled 7.5: Taste and Smell is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

CHAPTER OVERVIEW

8: Sensation and Perception

> This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

8.1: Sensation vs Perception
8.2: Psychophysical Methods
8.3: Perception and Action
8.4: Gestalt Principles of Perception

This page titled 8: Sensation and Perception is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

8.1: Sensation vs Perception

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Distinguish between sensation and perception
- Distinguish between top-down and bottom-up contributions to perception
- Describe key principles, such as transduction and sensory adaptation

Brief Overview:

The topics of sensation and perception are among the oldest and most important in all of psychology. People are equipped with senses such as sight, hearing and taste that help us to take in the world around us. Amazingly, our senses have the ability to convert real-world information into electrical information that can be processed by the brain. The way we interpret this information-- our perceptions-- is what leads to our experiences of the world. In this module, you will learn about the biological processes of sensation and perceptions as well as key differences between these two processes.

Sensation and perception are often intertwined, however, there are important distinctions between the two. The *physical* process during which our sensory organs—those involved with vision and hearing, for example—respond to external stimuli is called **sensation**. Sensation happens when you taste noodles or feel the wind on your face or hear a car horn honking in the distance. During sensation, our sense organs are engaging in **transduction**, the conversion of one form of energy into another. Physical energy such as light or sound is converted into a form of energy the brain can understand: electrical stimulation (i.e. action potentials). After our brain receives the electrical signals, we make sense of all this stimulation and begin to appreciate the complex world around us. This *psychological* process—making sense of the stimuli—is called **perception**. It is during this process that you are able to *identify* a gas leak in your home or a song that reminds you of a specific afternoon spent with friends.

While our sensory receptors are constantly collecting information from the environment, it is ultimately how we interpret that information that affects how we interact with the world. Perception involves both bottom-up and top-down processing. Bottom-up processing refers to the fact that perceptions are built from sensory input, your eyes are not the only way to use bottom-up processing. Imagine you are in the forest walking and you are admiring the trees and nature around. As you continue to walk you hear crackling noises and you begin to smell burning wood. As you continue walking you hear people talking and the crackling and smell of burning wood becomes stronger, at some point you may realize you are entering campgrounds, you can't see the campgrounds but bottom-up processing with the help of your other senses tells you what is going on.

On the other hand, how we interpret those sensations is influenced by our available knowledge, our experiences, and our thoughts. This is called top-down processing. One way to illustrate these two concepts is with our ability to read. Read the following quote pictured in Figure 8.1.1 out loud:



Figure 8.1.1 An example of stimuli processing.

Did you notice anything odd while you were reading the text in the triangle? Did you notice the second “the”? If not, it’s likely because you were reading this from a top-down approach. Having a second “the” doesn’t make sense. We know this. Our brain knows this and doesn’t *expect* there to be a second one, so we have a tendency to skip right over it. In other words, your past experience has changed the way you perceive the writing in the triangle! A beginning reader—one who is using a bottom-up approach by carefully attending to each piece—would be less likely to make this error. The above demonstration illustrates how our experiences can influence the way our brain processes sensory information.

Another way to distinguish between perception and sensation is that sensation is a physical process, whereas perception is psychological. For example, upon walking into a kitchen and smelling the scent of baking cinnamon rolls, the sensation is the scent receptors detecting the odor of cinnamon, but the perception may be “Mmm, this smells like the bread Grandma used to bake when the family gathered for holidays.” Although our perceptions are built from sensations, not all sensations result in perception. In fact, we often don’t perceive stimuli that remain relatively constant over prolonged periods of time, which is known as **sensory adaptation.** Imagine entering a classroom with an old analog clock. Upon first entering the room, you can hear the ticking of the clock; as you begin to engage in conversation with classmates or listen to your professor greet the class, you are no longer aware of the ticking. The clock is still ticking, and that information is still affecting sensory receptors of the auditory system. The fact that you no longer perceive the sound demonstrates sensory adaptation and shows that while closely associated, sensation and perception are different.

When we experience a sensory stimulus that doesn’t change, we stop paying attention to it. This is why we don’t feel the weight of our clothing, hear the hum of a projector in a lecture hall, or see all the tiny scratches on the lenses of our glasses. Thus, when a stimulus is constant and unchanging, we experience sensory adaptation. This occurs because if a stimulus does not change, our brain quits responding to it. A great example of this occurs when we leave the radio on in our car after we park it at home for the night. When we listen to the radio on our way home, the volume seems reasonable. However, the next morning when we start the car, we might be startled by how loud the radio is. We don’t remember it being that loud last night. What happened? We adapted to the constant stimulus (the radio volume) over the course of the previous day and increased the volume at various times.

8.1.1 Attribution

- Sensation and Perception sections adapted by Isaias Hernandez from "NOBA-Vision"; https://nobaproject.com/modules/sensation-and-perception
- Vision sections adapted by Isaias Hernandez from "Sensation vs Perception"; https://www.oercommons.org/courses/i...erception/view

---

This page titled 8.1: Sensation vs Perception is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

8.2: Psychophysical Methods

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

Differentiate between experiments measuring absolute thresholds, difference thresholds, and magnitude estimations

Describe at least two different methods of estimating a threshold

Understand what a subliminal message is

Explain Weber's law (also called Weber-Fechner law)

Brief Overview

Psychophysics *quantitatively investigates the relationship between physical stimuli and the sensations and perceptions they produce. Psychophysics has been described as the scientific study of the relation between stimulus and sensation or, more completely, as the analysis of perceptual processes by studying the effect on a subject's experience or behaviour of systematically varying the properties of a stimulus along one or more physical dimensions.*

The sensitivity of a given sensory system to the relevant stimuli can be expressed as an **absolute threshold**. **Absolute threshold** refers to the minimum amount of stimulus energy that must be present for the stimulus to be detected 50% of the time. Another way to think about this is by considering how dim a light can be or how soft a sound can be and still be detected half of the time. The sensitivity of our sensory receptors can be quite amazing. It has been estimated that on a clear night, the rods (sensitive sensory cells in the back of the eye) can detect a candle flame 30 miles away (Okawa & Sampath, 2007). Under quiet conditions, the hair cells (receptor cells of the inner ear) can detect the tick of a clock 20 feet away (Galanter, 1962) see figure 8.4.1 below.

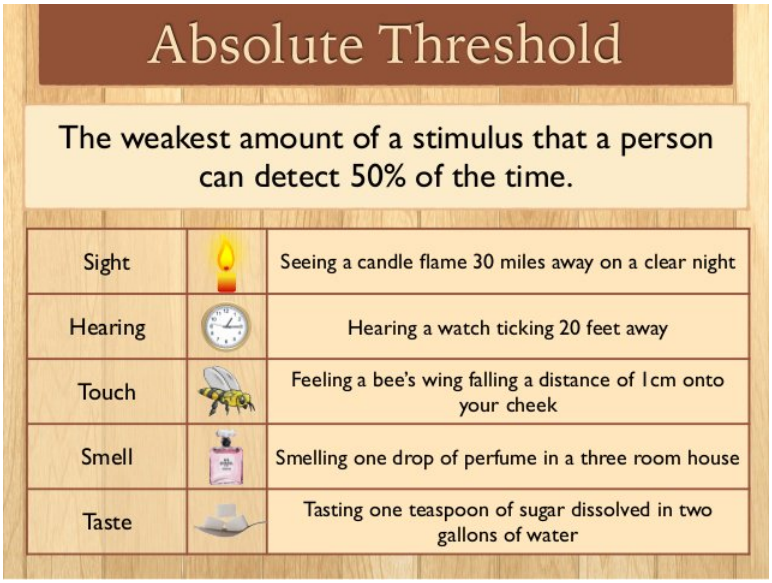


Fig 8.2.1 https://twitter.com/AP_Psychology/status/1094976965004705792/photo/1

It is also possible for us to get messages that are presented below the threshold for conscious awareness—these are called **subliminal messages**. A message below the absolute threshold is not strong enough to excite sensory receptors that send nerve impulses to the brain; when this occurs that threshold is said to be subliminal: We receive it, but we are not consciously aware of it. Over the years there has been a great deal of speculation about the use of **subliminal messages** in advertising, rock music, and self-help audio programs. Research evidence shows that in laboratory settings, people can process and respond to information outside of

awareness. But this does not mean that we obey these messages or are heavily influenced by these messages; in fact, hidden messages have little effect on behavior outside the laboratory (Kunst-Wilson & Zajonc, 1980; Rensink, 2004; Nelson, 2008; Radel, Sarrazin, Legrain, & Gobancé, 2009; Loersch, Durso, & Petty, 2013).

Methods for estimating thresholds

When we design experiments, we have to decide how we're going to approach a **threshold estimation**. Here are three common techniques:

- **Method of Limits**. The experimenter can increase the stimulus intensity (or intensity difference) until the observer detects the stimulus (or the change). For example, turn up the volume until the observer first detects the sound. This is intuitive, but it is also subject to bias — the estimated threshold is likely to be different, for example, if we start high and work down versus starting low and working up.
- **Method of Adjustment.** This is very much like the Method of Limits, except that the experimenter gives the observer control of the stimulus adjustment with the instructions to: "adjust the stimulus until it's clearly visible" or "adjust the color of the patch until it matches the test patch."
- **Method of Constant Stimuli**. This is the most reliable, but also the most time-consuming. You decide ahead of time what levels you are going to measure, do each one a fixed number of times, and record percent correct (or the number of detections) for each level. If you randomize the order, you can get rid of bias.

Absolute thresholds are generally measured under rigidly controlled conditions in situations that are optimal for sensitivity. Sometimes we are more interested in how much difference in stimuli is required to detect a difference between them. This is known as the **just noticeable difference** (jnd) or **difference threshold**. Unlike the **absolute threshold**, the **difference threshold** changes depending on the stimulus intensity. As an example, imagine yourself in a very dark movie theater. If an audience member were to receive a text message on their cell phone which caused their screen to light up, chances are that many people would notice the change in illumination in the theater. However, if the same thing happened in a brightly lit arena during a basketball game, very few people would notice. The cell phone brightness does not change, but its ability to be detected as a change in illumination varies dramatically between the two contexts. **Ernst Weber** proposed this theory of change in **difference threshold** in the 1830s, and it has become known as **Weber's law**.

Weber's law is approximately true for many aspects of our senses — for brightness perception, visual contrast perception, loudness perception, and visual distance estimation, our sensitivity to change decreases as the stimulus gets bigger or stronger. However, there are many senses for which the opposite is true: our sensitivity increases as the stimulus increases. With electric shock, for example, a small increase in the size of the shock is much more noticeable when the shock is large than when it is small. A psychophysical researcher named **Stanley Smith Stevens** asked people to estimate the magnitude of their sensations for many different kinds of stimuli at different intensities, and then tried to fit lines through the data to predict people's sensory experiences (Stevens, 1967). What he discovered was that most senses could be described by a power law of the form $\mathbf{P \propto S^n}$ (where P is the perceived magnitude, $\propto$ means "is proportional to", S is the physical stimulus magnitude, and n is a positive number). If n is greater than 1, then the slope (rate of change of perception) is getting larger as the stimulus gets larger, and sensitivity increases as stimulus intensity increases. A function like this is described as being expansive or supra-linear. If n is less than 1, then the slope decreases as the stimulus gets larger (the function "rolls over"). These sensations are described as being compressive. **Weber's Law** is only (approximately) true for compressive (sublinear) functions; **Stevens' Power Law** is useful for describing a wider range of senses see Fig 8.4.2 below.

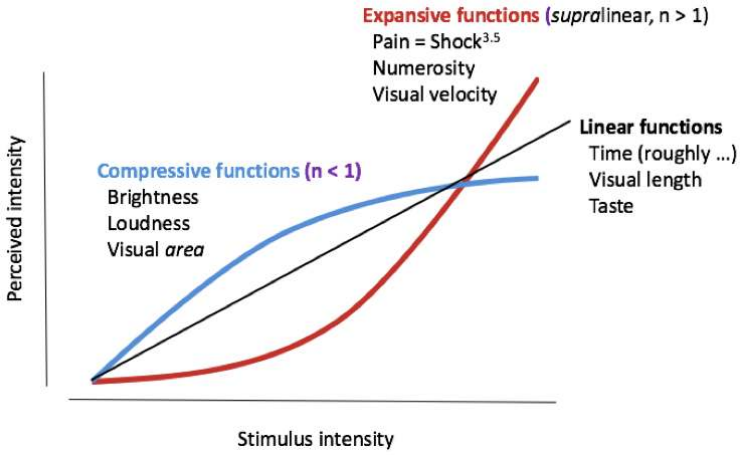


Fig 8.2.2 Different sensory systems exhibit different relationships between perceived magnitude and stimulus intensity. Sometimes, it makes the most sense to discount or ignore increases in stimulus intensity above a certain point; compressive sensory modalities with a power-law exponent less than 1 accomplish this. Other times, we need heightened sensitivity to stimuli with increased intensity; expansive sensory modalities, described by a power law with exponent greater than 1, accomplish this. Not all perception is non-linear, however; some senses are best described by a linear relationship between stimulus and perception.

Both **Stevens' Power Law** and **Weber's Law** are only approximately true. They are useful for describing, in broad strokes, how our perception of a stimulus depends on its intensity or size. They are rarely accurate for describing perception of stimuli that are near the absolute detection threshold. Still, they are useful for describing how people will expectably react to normal everyday stimuli.

Sensation and Perception Summary:

The world as we experience it is most often multimodal, involving combinations of our senses into one perceptual experience. The combination of senses that allow us to enjoy aspects of our everyday life requires our senses to be integrated. Interestingly, we actually respond more strongly to multimodal stimuli compared to the sum of each single modality together, an effect called the ***superadditive effect of multisensory integration****. This can explain how you're still able to understand what friends are saying to you at a loud concert, as long as you are able to get visual cues from watching them speak. If you were having a quiet conversation at a café, you likely wouldn't need these additional cues. Because we are able to process multimodal sensory stimuli, and the results of those processes are qualitatively different from those of unimodal stimuli, it's a fair assumption that the brain is doing something qualitatively different when they're being processed. There has been a growing body of evidence since the mid-1990's on the neural correlates of* ***multisensory integration****.*

8.2.1 Attribution

- Psychophysical Methods sections adapted by Isaias Hernandez from "Open Stax"; https://cnx.org/contents/Sr8Ev5Og@12.2:K-DZ-03P@12/5-1-Sensation-versus-Perception
- Psychophysical Methods sections adapted by Isaias Hernandez from "Wikipedia"; https://en.Wikipedia.org/wiki/Psychophysics#:~:text=Psychophysics%20has%20been%20described%20as,along%20one%20or%20more%20physical
- Psychophysical Methods sections adapted by Isaias Hernandez from "Psychophysical Methods"; https://pressbooks.umn.edu/sensationandperception/chapter/psychophysical-methods/

8.2.2 References

Galanter, E. (1962). Contemporary Psychophysics. In R. Brown, E.Galanter, E. H. Hess, & G. Mandler (Eds.), New directions in psychology. New York, NY: Holt, Rinehart & Winston.

Kunst-Wilson, W. R., & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. Science, 207, 557–558.

Nelson, M. R. (2008). The hidden persuaders: Then and now. Journal of Advertising, 37(1), 113–126.

Okawa, H., & Sampath, A. P. (2007). Optimization of single-photon response transmission at the rod-to-rod bipolar synapse. Physiology, 22, 279–286.

Radel, R., Sarrazin, P., Legrain, P., & Gobancé, L. (2009). Subliminal priming of motivational orientation in educational settings: Effect on academic performance moderated by mindfulness. Journal of Research in Personality, 43(4), 1–18.

Rensink, R. A. (2004). Visual sensing without seeing. Psychological Science, 15, 27–32.

Steingrimsson, R.; Luce, R. D. (2006). "Empirical evaluation of a model of global psychophysical judgments: III. A form for the psychophysical function and intensity filtering". *Journal of Mathematical Psychology*. **50**: 15–29. doi:10.1016/j.jmp.2005.11.005

Stevens, S. S. (1957). On the psychophysical law. Psychological Review 64(3):153—181. PMID 13441853

This page titled 8.2: Psychophysical Methods is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

8.3: Perception and Action

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Explain how perception guides action with an example
- Describe how action changes perception with an example
- Understand that acting on unconscious sensory information is possible

Brief Overview

What is the point of perceiving things if we cannot interact with them in some way? Perception is rarely a passive experience; it is almost always shaped by some kind of goal. Our perception is also shaped by the fact that it simply isn't feasible to process our entire sensory experience. We selectively attend to subsets of our experience and process that; much of our sensory world is not encoded. What we attend to depends on our goals at the moment, so our actions shape our perception and our perception shapes our action.

The relationship between perception and action is best described by a diagram with an arrow pointing both ways. Perception selects targets for action and helps us correct errors as we execute actions. Broadly speaking, there are two kinds of actions: navigation (moving around our environment) and intentional movements (such as reaching/grabbing) Fig 8.3.1.

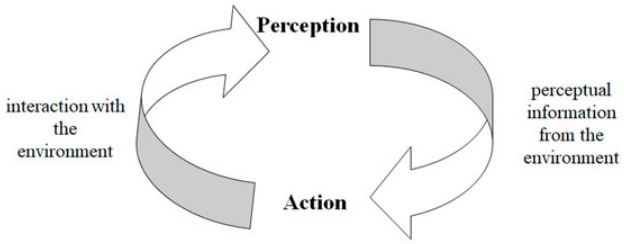


Fig 8.3.1 This figure shows the relationship between perception and action, which is described as a cycle of arrows pointing back and forth. An example of this cycle is when you reach out for a bottle sitting on a slanted surface: Banks et al (2000) showed that your sense of touch updates your visual estimates of the surface slant. (Provided by: MDPI. License: CC-BY 4.0)

Acting on unconscious sensory information is possible. There are a few dramatic examples of action without perception. For example a patient, DF, could mail letters through a slot but couldn't tell you if the slot had a horizontal or vertical orientation, or patient TN, who was completely blind but could navigate cluttered hallways. TN's case makes it obvious that visual information reaches the non-visual regions of the brain (such as areas of the **parietal cortex**) even when the **primary visual cortex** (V1) is not functioning typically.

The **Ebbinghaus size illusion** Fig 8.3.2is an optical illusion of relative size perception. In the best-known version of the illusion, two circles of identical size are placed near to each other, and one is surrounded by large circles while the other is surrounded by small circles. As a result of the juxtaposition of circles, the central circle surrounded by large circles appears smaller than the central circle surrounded by small circles.

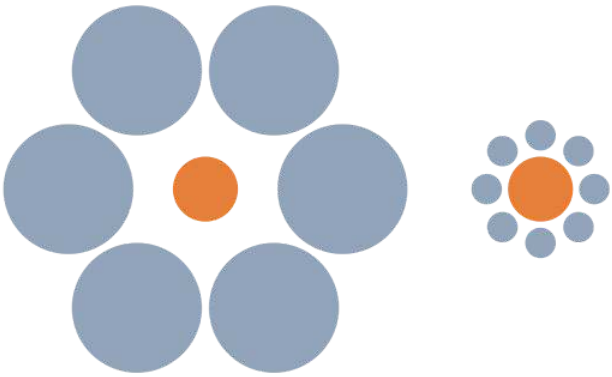


Fig 8.3.2 The two orange central circles are exactly the same size; however, the one on the right appears larger.

Similarly, healthy controls accurately grasp a center object affected by the rod-and-frame illusion, which is the phenomenon that the orientation of a line is altered by surrounding lines or grating (Dyde, 2000). These studies suggest that all of us maintain a "raw copy" of sensory information, separate from the version that we're consciously aware of (which has been shaped by inference), to guide action. However, there is contradictory research indicating that motor planning (Craje, 2008) and reaching behaviors (Dyde, 2000) can be impacted by illusions. So it remains an open and interesting question: When is our behavior influenced by illusions, and when does control of our motor system ignore illusions?

8.3.1 Attribution

- Perception and Action sections adapted by Isaias Hernandez from "Perception and Action"; https://pressbooks.umn.edu/sensationandperception/chapter/perception-action/
- Perception and Action sections adapted by Isaias Hernandez from "Psychology Wiki"; https://psychology.fandom.com/wiki/Visual_cortex
- Perception and Action sections adapted by Isaias Hernandez from "Wikipedia"; https://en.Wikipedia.org/wiki/Ebbinghaus_illusion

8.3.2 References

Crajé, C., van der Kamp, J., & Steenbergen, B. (2008). The effect of the "rod-and-frame" illusion on grip planning in a sequential object manipulation task. *Experimental brain research*, *185*(1), 53–62. https://doi.org/10.1007/s00221-007-1130-x

Dyde RT, Milner AD (2002) Two illusions of perceived orientations: one fools all of the people some of the time; the other fools all of the people all of the time. Exp Brain Res 144:518–527

Ernst, M., Banks, M. & Bülthoff, H. Touch can change visual slant perception. *Nat Neurosci* **3,** 69–73 (2000). https://doi.org/10.1038/71140

Haffenden AM, Goodale MA. The effect of pictorial illusion on prehension and perception. *J Cogn Neurosci*. 1998;10(1):122-136. doi:10.1162/089892998563824

This page titled 8.3: Perception and Action is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

8.4: Gestalt Principles of Perception

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

8.4.1 Gestalt Principles

Learning Objectives

- Explain the figure-ground relationship
- Define Gestalt principles of grouping
- Describe how perceptual set is influenced by an individual's characteristics and mental state
- Understand the influence stereotypes or prejudice can have on perceptual bias

Brief Overview

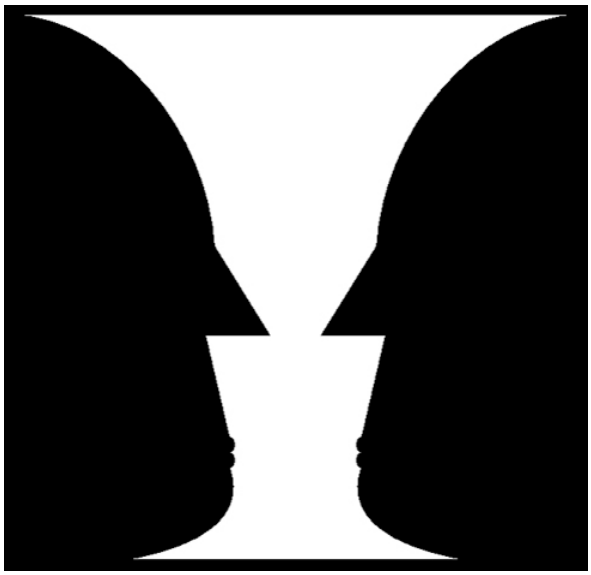
Gestalt theorists have been incredibly influential in the areas of sensation and perception. Gestalt principles such as figure-ground relationship, grouping by proximity or similarity, the law of good continuation, and closure are all used to help explain how we organize sensory information. Our perceptions are not infallible, and they can be influenced by bias, prejudice, and other factors.

8.4.2 Gestalt Psychology

In the early part of the 20th century, Max Wertheimer published a paper demonstrating that individuals perceived motion in rapidly flickering static images. Research conducted by Wertheimer and his colleagues proposed that perception involved more than simply combining sensory stimuli and led to a new field of psychology known as **Gestalt psychology**. The word **gestalt** literally means form or pattern, but its use reflects the idea that the whole is different from the sum of its parts. In other words, the brain creates a perception that is more than simply the sum of available sensory inputs, and it does so in predictable ways. Gestalt psychologists translated these predictable ways into principles by which we organize sensory information. As a result, Gestalt psychology has been extremely influential in the area of sensation and perception (Rock & Palmer, 1990).

8.4.2.1 Figure-ground relationship

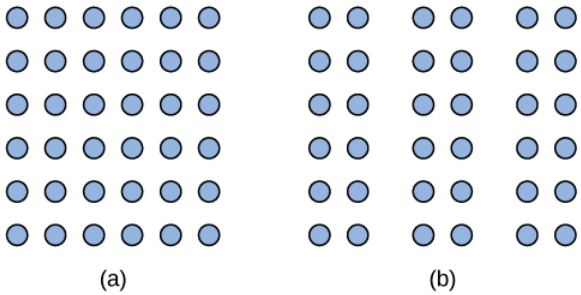
One Gestalt principle is the **figure-ground relationship**. According to this principle, we tend to segment our visual world into figure and ground. Figure is the object or person that is the focus of the visual field, while the ground is the background. As the figure below shows, our perception can vary tremendously, depending on what is perceived as figure and what is perceived as ground. Presumably, our ability to interpret sensory information depends on what we label as figure and what we label as ground in any particular case, although this assumption has been called into question (Peterson & Gibson, 1994; Vecera & O'Reilly, 1998). See figure 8.2.1 below.



8.4.1 The concept of figure-ground relationship explains why this image can be perceived either as a vase or as a pair of faces.

8.4.2.2 Principle of proximity

Another Gestalt principle for organizing sensory stimuli into meaningful perception is **proximity**. This principle asserts that things that are close to one another tend to be grouped together, as figure 8.2.2 below demonstrates.

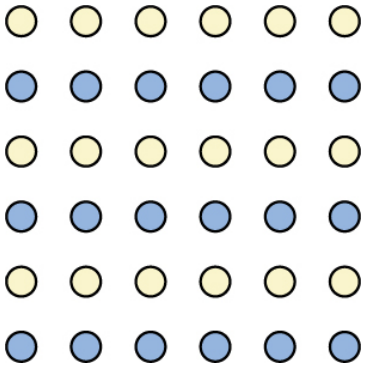


8.4.2 The Gestalt principle of proximity suggests that you see (a) one block of dots on the left side and (b) three columns on the right side.

How we read something provides another illustration of the **proximity concept**. For example, we read this sentence like this, notl iket hiso rt hat. We group the letters of a given word together because there are no spaces between the letters, and we perceive words because there are spaces between each word. Here are some more examples: Cany oum akes enseo ft hiss entence? What doth es e wor dsmea n?

8.4.2.3 Principle of similarity

We also often use the **principle of similarity** to group things in our visual fields. According to this principle, things that are alike tend to be grouped together as Figure 8.2.3 below illustrates.



8.4.3 When looking at this array of dots, we likely perceive alternating rows of colors. We are grouping these dots according to the principle of similarity.

Another example is when watching an American football game, we tend to group individuals based on the colors of their uniforms. While watching an offensive drive (a series of plays which help the offense score), we can get a sense of the two teams simply by grouping along this dimension. Two additional Gestalt principles are the **law of continuity** (or good continuation) and the **principle of closure**.

8.4.2.4 Law of continuity

The **law of continuity** suggests that we are more likely to perceive continuous, smooth flowing lines rather than jagged, broken lines as illustrated in the figure below.

parts (Fig 8.2.4)

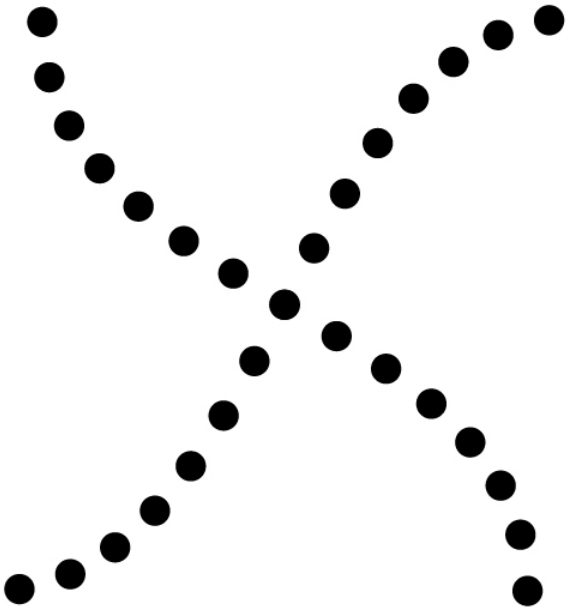


Fig 8.4.4 Good continuation would suggest that we are more likely to perceive this as two overlapping lines, rather than four lines meeting in the center.

8.4.2.5 Principle of closure

The **principle of closure** states that we organize our perceptions into complete objects rather than as a series of parts as illustrated in Fig 8.2.5



Fig 8.4.5 Closure suggests that we will perceive a complete circle and rectangle rather than a series of segments.

8.4.2.6 Pattern perception

According to Gestalt theorists, **pattern perception**, or our ability to discriminate among different figures and shapes, occurs by following the principles described above. You probably feel fairly certain that your perception accurately matches the real world, but this is not always the case. Our perceptions are based on **perceptual hypotheses**: educated guesses that we make while interpreting sensory information. These hypotheses are informed by a number of factors, including our personalities, experiences, and expectations. We use these hypotheses to generate our perceptual set. For instance, research has demonstrated that those who are given verbal priming produce a biased interpretation of complex ambiguous figures (Goolkasian & Woodbury, 2010).

8.4.3 Aspects of Bias, Prejudice, and Cultural Factors

In this chapter, you have learned that perception is a complex process. Built from sensations, but influenced by our own experiences, biases, prejudices, and cultures, perceptions can be very different from person to person. Research suggests that implicit racial prejudice and stereotypes affect perception. For instance, several studies have demonstrated that non-Black participants identify weapons faster and are more likely to identify non-weapons as weapons when the image of the weapon is paired with the image of a Black person (Payne, 2001; Payne, Shimizu, & Jacoby, 2005). Furthermore, White individuals' decisions to shoot an armed target in a video game is made more quickly when the target is Black (Correll, Park, Judd, & Wittenbrink, 2002; Correll, Urland, & Ito, 2006). This research is important, considering the number of very high-profile cases in the last few decades in which young Blacks were killed by people who claimed to believe that the unarmed individuals were armed and/or represented some threat to their personal safety. In the article *Stereotypes and Prejudice in the Perception of the "Other"* (Fedor, 2014), it is argued that **otherness** (that which is other than the concept being considered; often it means a person other than oneself) can lead to problematic interpersonal communication that often becomes permanent and may result in the prevention of community collaboration and development. The full article can be found here, https://www.sciencedirect.com/science/article/pii/S1877042814049702.

8.4.4 **Attribution**

- Gestalt Principles of Perception sections adapted by Isaias Hernandez from "OpenStax"; https://openstax.org/books/psychology/pages/5-6-gestalt-principles-of-perception

This page titled 8.4: Gestalt Principles of Perception is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

CHAPTER OVERVIEW

9: Movement

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

9.1 Overview

This chapter will explore some of the basic elements of the neural control of movement. A brief discussion of the various muscle systems of the body will be provided along with the role that each plays in our ability to navigate through our environment and keep our vital organs, such as the heart and lungs, operating. The amazing influence of microscopic, synaptic communication from neurons to muscle fibers will be illustrated with a look at the neuromuscular junction and the motor units that control muscle contraction. We will then zoom out a bit and review the multiple central and peripheral nervous system circuits involved with the control of both voluntary and involuntary motor functions. Lastly, the impact of specific disease states and neural damage to motor function will be examined.

9.1: Classification of Muscle Types and Functions

9.2: Synaptic Control of Movement- Neuromuscular Junction and Motor Units

9.3: Spinal Reflexes

9.4: Motor Circuitry- Neural Structures and Pathways

9.5: Impact of Disease States and Neural Damage on Motor Control

This page titled 9: Movement is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

9.1: Classification of Muscle Types and Functions

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Identify the three types of muscle tissue
- Compare and contrast the functions of each muscle tissue type
- Explain how muscle tissue can enable motion

9.1.1 Muscle Tissue: Key Properties and Classifications

Muscle tissue is generally characterized by properties that allow movement. A critical property is that muscles are ***excitable*** and are able to respond to a variety of stimuli. They are ***contractile***, meaning they can shorten and generate a pulling force. When attached between two movable objects, in other words, bones, contractions of the muscles cause the bones to move.

Some muscle movement is ***voluntary***, which means it is under conscious control. For example, a person decides to open a book and read a chapter on Psychology. Other movements are ***involuntary***, meaning they are *not* typically under conscious control, such as the contraction of your pupil in bright light or the rhythmic contraction of your heart muscles.

Muscle tissue used for voluntary and involuntary movement can be classified into three main types according to structure and function: **Skeletal, Cardiac, and Smooth**. Table 1 below illustrates the distinctions between these three muscle types.

9.1.1.1 Comparison of Structural and Functional Properties of Muscle Types

Table 1: Muscle Types - Structure, Function, and Location

Tissue	Histology	Function	Location
Skeletal	Long cylindrical fiber, striated, many peripherally located nuclei	Voluntary movement, produces heat, protects organs	Attached to bones and around entrance points to body (e.g., mouth, anus)
Cardiac	Short, branched, striated, single central nucleus	Contracts to pump blood	Heart
Smooth	Short, spindle-shaped, no evident striation, single nucleus in each fiber	Involuntary movement, moves food, involuntary control of respiration, moves secretions, regulates flow of blood in arteries by contraction	Walls of major organs and passageways

Skeletal muscle is attached to bones and its contraction makes possible locomotion (i.e. walking), facial expressions, maintaining posture, and other voluntary movements of the body. Skeletal muscles also generate heat as a byproduct of their contraction and thus participate in thermal regulation. Shivering is an involuntary contraction of skeletal muscles in response to perceived lower than normal body temperature.

Skeletal muscles act not only to produce movement but also to stop movement, such as resisting gravity to maintain posture. Small, constant adjustments of the skeletal muscles are needed to hold a body upright or balanced in any position. Muscles also prevent excess movement of the bones and joints, maintaining skeletal stability and preventing skeletal structure damage or deformation. Joints can become misaligned or dislocated entirely by pulling on the associated bones; muscles work to keep joints stable.

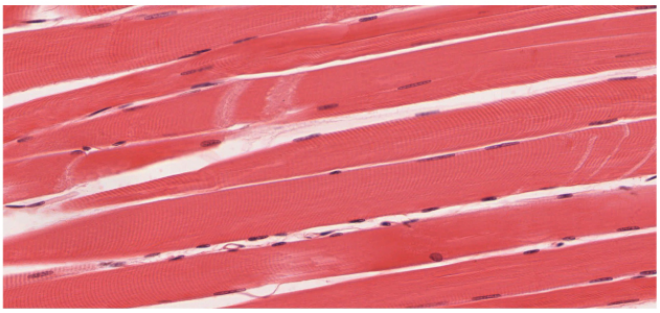
Skeletal muscles are also located throughout the body at the openings of internal tracts to control the movement of various substances. These muscles allow functions, such as swallowing, urination, and defecation, to be under voluntary control. Skeletal

muscles also protect internal organs (particularly abdominal and pelvic organs) by acting as an external barrier or shield to external trauma and by supporting the weight of the organs.

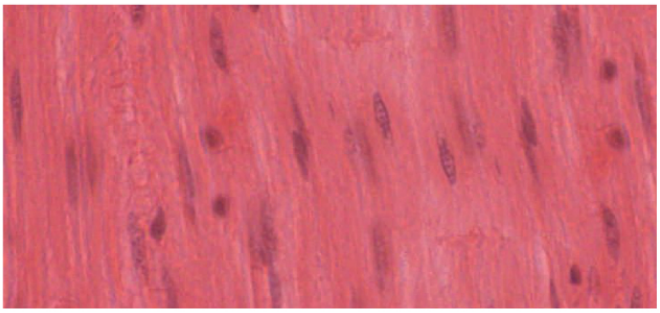
Skeletal muscle tissue is arranged in bundles surrounded by connective tissue. Under the light microscope, muscle cells appear striated (striped) with many nuclei squeezed along the membranes. The **striation** is due to the regular alternation of the contractile proteins actin and myosin, along with the structural proteins that couple the contractile proteins to connective tissues. The cells are multinucleated as a result of the fusion of many precursor cells to form each long muscle fiber.

Cardiac muscle forms the contractile walls of the heart. The cells of cardiac muscle, known as cardiomyocytes, also appear striated under the microscope. Unlike skeletal muscle fibers, cardiomyocytes are single cells typically with a single centrally located nucleus. A principal characteristic of cardiomyocytes is that they contract on their own intrinsic rhythms without any external stimulation. Cardiomyocytes attach to one another with specialized cell junctions called intercalated discs. Intercalated discs have both anchoring junctions and gap junctions. Attached cells form long, branching cardiac muscle fibers that are, essentially, a mechanical and electrochemical syncytium allowing the cells to synchronize their actions. The cardiac muscle pumps blood through the body and is under involuntary control. The attachment junctions hold adjacent cells together across the dynamic pressure changes of the cardiac cycle.

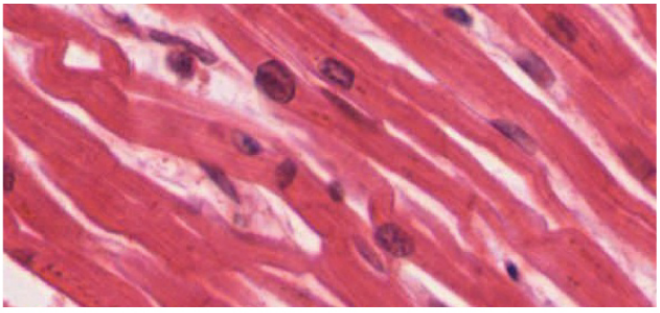
Smooth muscle tissue contraction is responsible for involuntary movements in the internal organs. It forms the contractile component of the digestive, urinary, and reproductive systems as well as the airways and arteries. Each cell is spindle shaped with a single nucleus and no visible striations (Figure 4.18).



(a)



(b)



(c)

Figure 9.1.1: **Muscle Tissue** (a) Skeletal muscle cells have prominent striation and nuclei on their periphery. (b) Smooth muscle cells have a single nucleus and no visible striations. (c) Cardiac muscle cells appear striated and have a single nucleus. (Micrographs provided by the Regents of University of Michigan Medical School © 2012)

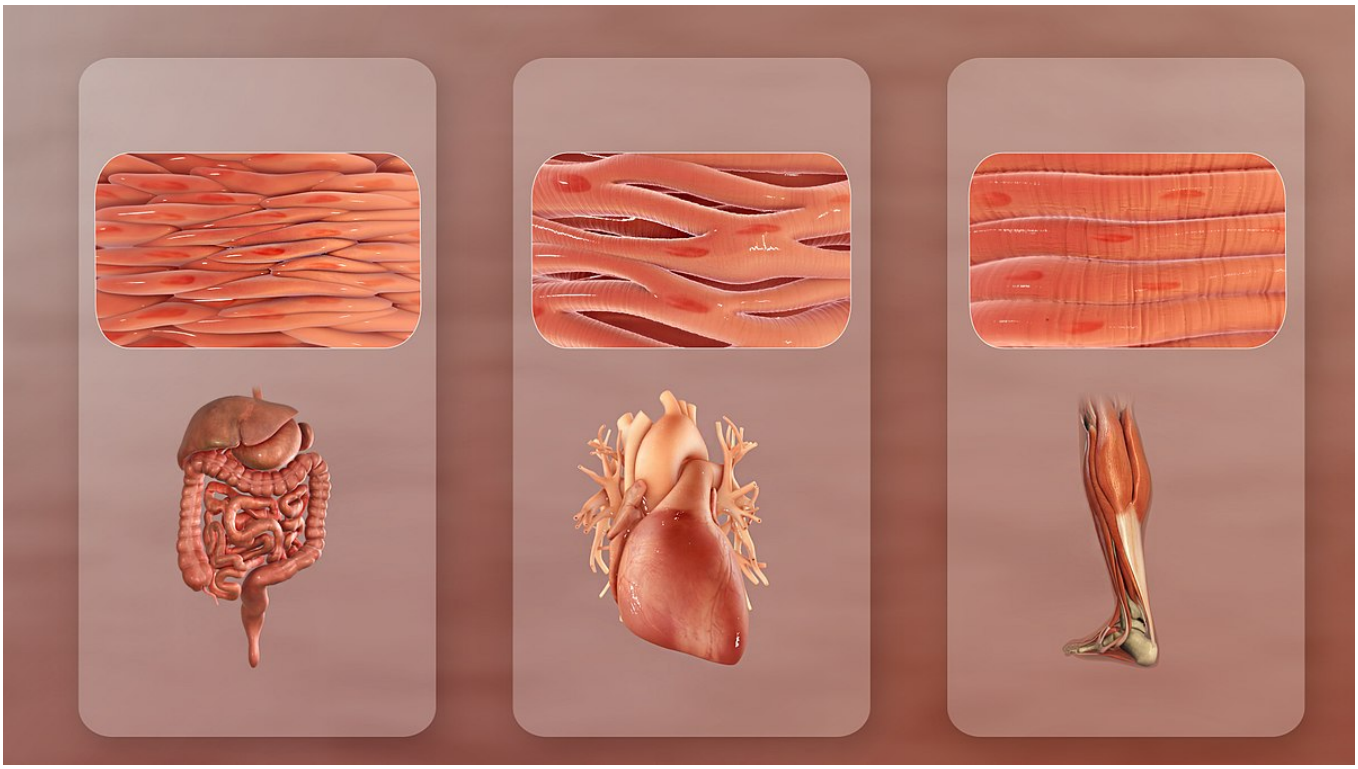


Figure **9.1.2**: **Types of Muscle Tissue** (Left) Smooth muscle cells (Center) Cardiac muscle cells (Right) Skeletal muscle cells. (Copyright; Wikimedia Commons, License: CC-BY-SA-4.0)

9.1.2

9.1.2.1 Slow and Fast Twitch Skeletal Muscles

Skeletal muscle fibers can be further subdivided into slow and fast-twitch subtypes depending on their metabolism and corresponding action. Most muscles are made up of combinations of these fibers, although the relative number varies substantially.

9.1.2.1.1 **Slow Twitch**

Slow-twitch fibers are designed for endurance activities that require long-term, repeated contractions, like maintaining posture or running a long distance. These activities require the delivery of large amounts of oxygen to the muscle, which can rapidly become rate-limiting if the respiratory and circulatory systems cannot keep up.

Due to their large oxygen requirements, slow-twitch fibers are associated with large numbers of blood vessels, mitochondria, and high concentrations of myoglobin, an oxygen-binding protein found in the blood that gives muscles their reddish color. One muscle with many slow-twitch fibers is the soleus muscle in the leg (~80% slow-twitch), which plays a key role in standing.

9.1.2.1.2 **Fast Twitch**

Fast-twitch fibers are good for rapid movements like jumping or sprinting that require fast muscle contractions of short duration. As fast-twitch fibers generally do not require oxygenation, they contain fewer blood vessels and mitochondria than slow-twitch fibers and less myoglobin, resulting in a paler color. Muscles controlling eye movements contain high numbers of fast-twitch fibers (~85% fast-twitch).

9.1.3 Attributions:

"Muscle Tissue: Key Properties and Classifications" and "Comparison of Structural and Functional Properties of Muscle Types" adapted by Alan Keys from J. Gordon Betts, Kelly A. Young, James A. Wise, Eddie Johnson, Brandon Poe, Dean H. Kruse, Oksana Korol, Jody E. Johnson, Mark Womble, Peter DeSaix, Anatomy and Physiology, OpenStax. **License:** CC BY 4.0

"Slow and Fast Twitch Skeletal Muscles" adapted from Anatomy and Physiology (Boundless) by LibreTexts. **License:** CC BY-SA.

This page titled 9.1: Classification of Muscle Types and Functions is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

9.2: Synaptic Control of Movement- Neuromuscular Junction and Motor Units

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Describe the structural and functional aspects of motor neurons and the neuromuscular junction
- Explain the key steps of electrical excitation of muscle fibers
- Apply knowledge of motor units and motor unit recruitment to functional changes in muscle strength

9.2.1 Motor Neurons

Although the neuroanatomy of the motor system will be covered in a later section, it is important to realize that a large portion of the Central and Peripheral Nervous Systems contain collections of neurons that influence motor action in some way. Many of these neurons in the CNS belong to so-called, "motor circuits" and are classified as ***motor*** neurons as opposed to sensory or interneurons. Motor neurons located in cortical and subcortical regions of the forebrain, such as motor and premotor cortex and the basal ganglia are referred to as **Upper Motor Neurons**, while those located in the medulla and spinal cord are the **Lower Motor Neurons**. The function of the Upper Motor Neurons will be addressed in later sections. The Lower Motor Neurons of the spinal cord play a direct role in both triggering voluntary movement and maintaining muscle tone. These are the **Alpha** and **Gamma** motor neurons, respectively.

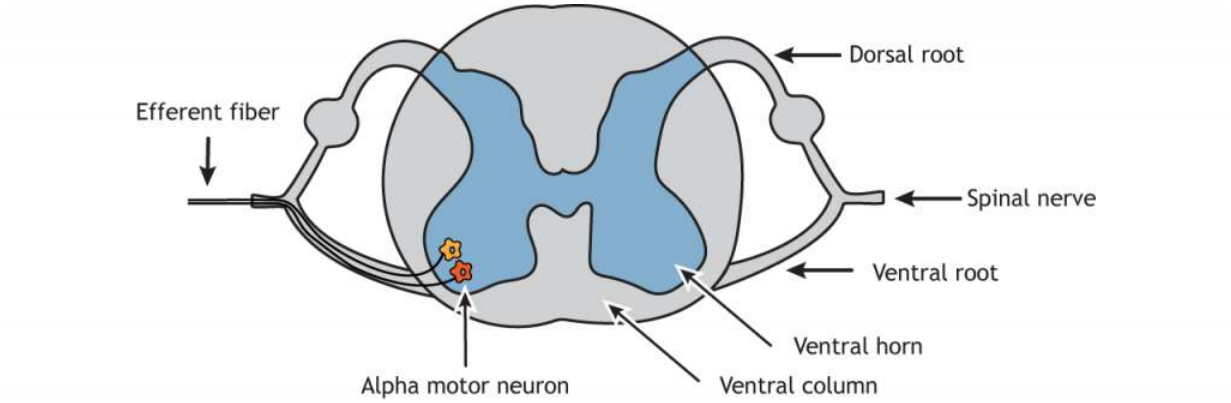


Figure **9.2.1**: Alpha motor neurons are located in the ventral (anterior) horn of spinal cord. Their axons, which are efferent fibers, travel to the muscles via spinal nerves. 'Alpha Motor Neurons' by Casey Henley. License: CC BY-NC-SA 4.0 International License.

The cell bodies of the Alpha motor neurons are located in the central nervous system in the ventral (i.e. anterior) horn of the spinal cord. Their axons leave the spinal cord via the ventral roots and travel primarily to skeletal muscle via efferent (outgoing) spinal nerves to cause muscle contraction. Gamma motor neurons function to keep muscle fibers prepared for action by stimulating them in a way to maintain "tautness" or muscle tone. This resting muscle tonus allows the muscle to respond much more effectively to the alpha motor neuron stimulation, thus alpha and gamma neurons work together to maintain a muscle's sensitivity for movement. Just imagine how difficult it would be to shoot a rubber band across a room if it wasn't stretched or taut (Gamma motor neuron influence) prior to letting it go (Alpha motor neuron influence).

9.2.2 The Neuromuscular Junction

A unique specialization of the skeletal muscle is the site where a motor neuron's axon terminal meets the muscle fiber—called the **neuromuscular junction (NMJ)**. This is where the muscle fiber first responds to signaling by the motor neuron. Every skeletal muscle fiber in every skeletal muscle is innervated by a motor neuron at the NMJ. Excitation signals from the neuron are the only way to functionally activate the fiber to contract. It is important to note that one motor neuron can activate multiple muscle fibers due to axonal branching (see the Motor Units section below for more details on this topic).

9.2.2.1 Excitability of Muscle: Excitation-Contraction Coupling

All living cells have membrane potentials, or electrical gradients across their membranes. For example, in neurons which are not being stimulated the membrane potential is approximately -70 mV inside of the cell relative to the outside. This is referred to as a neuron's resting potential.

Although living cells have a cellular membrane, only some, such as neurons and muscle cells are excitable. In other words, they can shift quickly from a resting state to an excited one. Thus, neurons and muscle cells can use their membrane potentials to generate electrical signals. They do this by controlling the movement of charged particles, called ions, across their membranes to create electrical currents. This is achieved by opening and closing specialized proteins in the membrane called ion channels. Although the currents generated by ions moving through these channel proteins are very small, they form the basis of both neural signaling and muscle contraction.

Both neurons and skeletal muscle cells are electrically excitable, meaning that they are able to generate action potentials. An action potential is a special type of electrical signal that can travel along a cell membrane as a wave. This allows a signal to be transmitted quickly and faithfully over long distances.

Although the term **excitation-contraction coupling** confuses or scares some students, it comes down to this: for a skeletal muscle fiber to contract, its membrane must first be "excited"—in other words, it must be stimulated to fire an action potential. The muscle fiber action potential, which sweeps along the muscle fiber as a wave, is "coupled" to the actual contraction through the release of calcium ions (Ca^{++}) from the muscle cells. Once released, the Ca^{++} interacts with the shielding proteins, forcing them to move aside so that the actin-binding sites are available for attachment by myosin heads. The myosin then pulls the actin filaments toward the center, shortening the muscle fiber.

In skeletal muscle, this sequence begins with signals from the somatic motor division of the nervous system. In other words, the "excitation" step in skeletal muscles is always triggered by signaling from the nervous system.

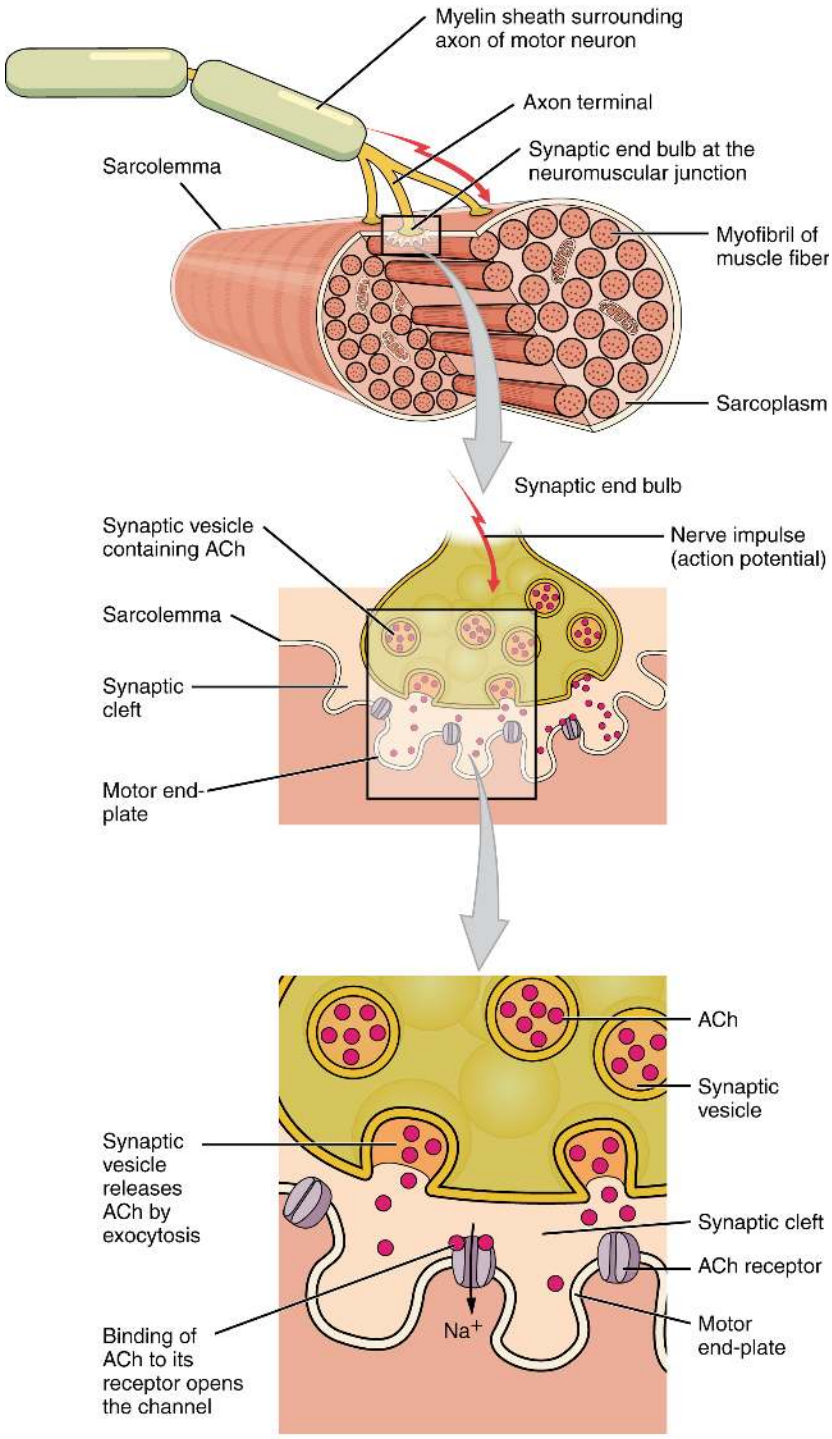


Figure **9.2.2**: Motor End-Plate and Innervation at the NMJ, the axon terminal releases the neurotransmitter, Acetylcholine (ACh). The motor end-plate is the location of the ACh-receptors in the muscle fiber sarcolemma. When ACh molecules are released, they diffuse across a minute space called the synaptic cleft and bind to the receptors.

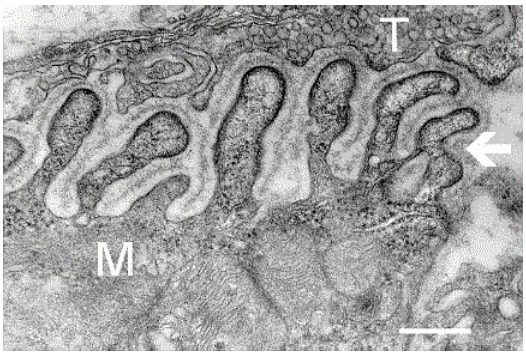


Figure **9.2.3**: Electron micrograph showing a cross-section through the neuromuscular junction. T is the axon terminal, M is the muscle fiber. The arrow shows junctional folds with basal lamina. Postsynaptic densities are visible on the tips between the folds. The scale is 0.3 µm. (Wikimedia Commons) License: *Public Domain: No Known Copyright*

Although a small number of motor neurons activating the skeletal muscles of the face, head, and neck are located in the brainstem, most motor neurons originate in the spinal cord, directing skeletal muscle fibers to contract throughout the rest of the body. These neurons have long processes, called axons, which are specialized to transmit action potentials long distances— in this case, all the way from the spinal cord to the muscle itself (which may be up to three feet away). The axons of multiple neurons bundle together to form nerves, like wires bundled together in a cable.

Signaling begins when a neuronal **action potential** travels along the axon of a motor neuron, and then along the individual branches to terminate at the NMJ. At the NMJ, the axon terminal releases a chemical messenger, or **neurotransmitter**, called **acetylcholine (ACh)**. The ACh molecules diffuse across a minute space called the **synaptic cleft** and bind to ACh receptors located within the **motor end-plate** of the sarcolemma on the other side of the synapse. Once ACh binds, a channel in the ACh receptor opens and positively charged ions can pass through into the muscle fiber, causing it to **depolarize**, meaning that the membrane potential of the muscle fiber becomes less negative (closer to zero).

As the membrane depolarizes, another set of ion channels called **voltage-gated sodium channels** are triggered to open. Sodium ions enter the muscle fiber, and an action potential rapidly spreads (or "fires") along the entire membrane to initiate excitation-contraction coupling.

Things happen very quickly in the world of excitable membranes (just think about how quickly you can snap your fingers as soon as you decide to do it). Immediately following depolarization of the membrane, it repolarizes, re-establishing the negative membrane potential. Meanwhile, the ACh in the synaptic cleft is degraded by the enzyme acetylcholinesterase (AChE) so that the ACh cannot rebind to a receptor and reopen its channel, which would cause unwanted extended muscle excitation and contraction.

Propagation of an action potential along the sarcolemma is the excitation portion of excitation-contraction coupling.This excitation actually triggers the release of Calcium ions (Ca^{++}) from their storage in the cell's sacroplasmic reticulum (SR). For the action potential to reach the membrane of the SR, there are periodic invaginations in the sarcolemma, called **T-tubules** ("T" stands for "transverse"). The diameter of a muscle fiber can be up to 100 μm, so these T-tubules ensure that the membrane can get close to the SR in the sarcoplasm. The arrangement of a T-tubule with the membranes of SR on either side is called a **triad**. The triad surrounds the cylindrical structure called a **myofibril**, which contains actin and myosin.

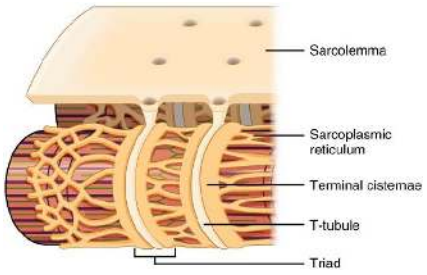


Figure **9.2.4**: The T-tubule: Narrow T-tubules permit the conduction of electrical impulses. The SR functions to regulate intracellular levels of calcium. Two terminal cisternae (where enlarged SR connects to the T-tubule) and one T-tubule comprise a triad—a "threesome" of membranes, with those of SR on two sides and the T-tubule sandwiched between them.

The T-tubules carry the action potential into the interior of the cell, which triggers the opening of calcium channels in the membrane of the adjacent SR, causing Ca^{++} to diffuse out of the SR and into the sarcoplasm. It is the arrival of Ca^{++} in the sarcoplasm that initiates contraction of the muscle fiber by its contractile units, or sarcomeres.

9.2.3 Motor Units

As described in the Neuromuscular Junction section above, every skeletal muscle fiber is innervated by an axon terminal of a motor neuron in order to contract. Each muscle fiber is innervated by only one motor neuron, but each motor neuron innervates a group of muscle fibers due to the axon's ability to branch. Recall that the NMJ is the point of connection between each axon terminal and it's corresponding muscle fiber. A motor neuron and the group of muscle fibers in a muscle that it innervates is called a **motor unit**. The size of a motor unit is variable depending on the nature of the muscle.

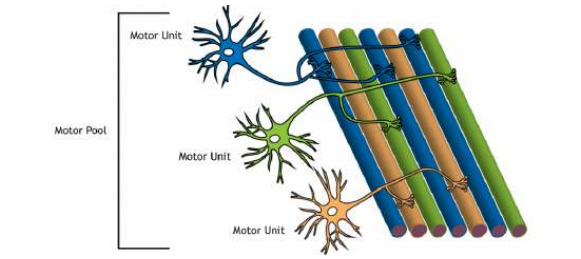


Figure **9.2.5**: Motor neurons can innervate more than one muscle fiber within a muscle. The motor neuron and the fibers it innervates are a motor unit. Three motor units are shown in the image: one blue, one green, one orange. Those three motor units innervate all the muscles fibers in the muscle and are the motor pool for that muscle. 'Motor Unit and Pool'by Casey Henley. License: CC BY-NC-SA 4.0 International License.

A small motor unit is an arrangement where a single motor neuron supplies a small number of muscle fibers in a muscle. Small motor units permit very fine motor control of the muscle. The best example in humans is the small motor units of the extraocular eye muscles that move the eyeballs. There are thousands of muscle fibers in each muscle, but every six or so fibers are supplied by a single motor neuron, as the axons branch to form synaptic connections at their individual NMJs. This allows for exquisite control of eye movements so that both eyes can quickly focus on the same object. Small motor units are also involved in the many fine movements of the fingers and thumb of the hand for grasping, texting, etc.

A large motor unit is an arrangement where a single motor neuron supplies a large number of muscle fibers in a muscle. Large motor units are concerned with simple, or "gross," movements, such as powerfully extending the knee joint. The best example is the large motor units of the thigh muscles or back muscles, where a single motor neuron will supply thousands of muscle fibers in a muscle, as its axon splits into thousands of branches.

There is a wide range of motor units within many skeletal muscles, which gives the nervous system a wide range of control over the muscle. The small motor units in the muscle will have smaller, lower-threshold motor neurons that are more excitable, firing first to their skeletal muscle fibers, which also tend to be the smallest. Activation of these smaller motor units results in a relatively small degree of contractile strength (tension) generated in the muscle. As more strength is needed, larger motor units, with bigger, higher-threshold motor neurons, are enlisted to activate larger muscle fibers. This increasing activation of motor units produces an increase in muscle contraction known as **recruitment**.

As more motor units are recruited, muscle contraction grows progressively stronger. In some muscles, the largest motor units may generate a contractile force of 50 times more than the smallest motor units in the muscle. This allows a feather to be picked up using the biceps brachii arm muscle with minimal force, and a heavy weight to be lifted by the same muscle by recruiting the largest motor units.

When necessary, the maximal number of motor units in a muscle can be recruited simultaneously, producing the maximum force of contraction for that muscle, but this cannot last for very long because of the energy requirements to sustain the contraction. To prevent complete muscle fatigue, motor units are generally not all simultaneously active, but instead some motor units rest while others are active, which allows for longer muscle contractions. The nervous system uses recruitment as a mechanism to efficiently utilize a skeletal muscle.

9.2.3.0.1 Attributions:

- Sections on Motor Neurons, Neuromuscular Junction, and Motor Units adapted from J. Gordon Betts, Kelly A. Young, James A. Wise, Eddie Johnson, Brandon Poe, Dean H. Kruse, Oksana Korol, Jody E. Johnson, Mark Womble, Peter DeSaix, Anatomy and Physiology, OpenStax. **License:** CC BY 4.0
- 'Alpha Motor Neurons' and 'Motor Unit and Pool' graphics by Casey Henley. **License:** CC BY-NC-SA 4.0 International License.

This page titled 9.2: Synaptic Control of Movement- Neuromuscular Junction and Motor Units is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

9.3: Spinal Reflexes

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Distinguish between the types of spinal reflexes

Spinal reflexes include the stretch reflex, the Golgi tendon reflex, the crossed extensor reflex, and the withdrawal reflex.

9.3.1 Stretch Reflex

The stretch reflex (myotatic reflex) is a muscle contraction in response to stretching within the muscle. This reflex has the shortest latency of all spinal reflexes. It is a monosynaptic reflex that provides automatic regulation of skeletal muscle length.

When a muscle lengthens, the muscle spindle is stretched and its nerve activity increases. This increases alpha motor neuron activity, causing the muscle fibers to contract and thus resist the stretching. A secondary set of neurons also causes the opposing muscle to relax. The reflex functions to maintain the muscle at a constant length.

9.3.2 Golgi Tendon Reflex

The Golgi tendon reflex is a normal component of the reflex arc of the peripheral nervous system. The tendon reflex operates as a feedback mechanism to control muscle tension by causing muscle relaxation before muscle force becomes so great that tendons might be torn.

Although the tendon reflex is less sensitive than the stretch reflex, it can override the stretch reflex when tension is great, making you drop a very heavy weight, for example. Like the stretch reflex, the tendon reflex is ipsilateral.

The sensory receptors for this reflex are called Golgi tendon receptors, and lie within a tendon near its junction with a muscle. In contrast to muscle spindles, which are sensitive to changes in muscle length, tendon organs detect and respond to changes in muscle tension that are caused by a passive stretch or muscular contraction.

9.3.3 Crossed Extensor Reflex



Jendrassik maneuver: The Jendrassik maneuver is a medical maneuver wherein the patient flexes both sets of fingers into a hook-like form and interlocks those sets of fingers together (note the hands of the patient in the chair). This maneuver is used often when testing the patellar reflex, as it forces the patient to concentrate on the interlocking of the fingers and prevents conscious inhibition or influence of the reflex.

The crossed extensor reflex is a withdrawal reflex. The reflex occurs when the flexors in the withdrawing limb contract and the extensors relax, while in the other limb, the opposite occurs. An example of this is when a person steps on a nail, the leg that is stepping on the nail pulls away, while the other leg takes the weight of the whole body.

The crossed extensor reflex is contralateral, meaning the reflex occurs on the opposite side of the body from the stimulus. To produce this reflex, branches of the afferent nerve fibers cross from the stimulated side of the body to the contralateral side of the spinal cord. There, they synapse with interneurons, which in turn, excite or inhibit alpha motor neurons to the muscles of the contralateral limb.

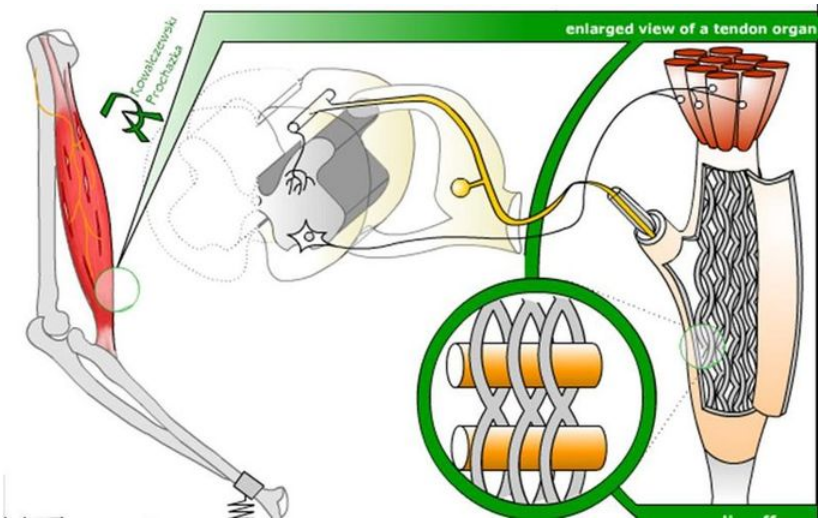
9.3.4 Withdrawal Reflex

The withdrawal reflex (nociceptive or flexor withdrawal reflex) is a spinal reflex intended to protect the body from damaging stimuli. It is polysynaptic, and causes the stimulation of sensory, association, and motor neurons.

When a person touches a hot object and withdraws his hand from it without thinking about it, the heat stimulates temperature and danger receptors in the skin, triggering a sensory impulse that travels to the central nervous system. The sensory neuron then synapses with interneurons that connect to motor neurons. Some of these send motor impulses to the flexors to allow withdrawal.

Some motor neurons send inhibitory impulses to the extensors so flexion is not inhibited—this is referred to as reciprocal innervation. Although this is a reflex, there are two interesting aspects to it:

1. The body can be trained to override that reflex.
2. An unconscious body (or even drunk or drugged bodies) will not exhibit the reflex.



Golgi tendon organ: The Golgi tendon organ, responsible for the Golgi tendon reflex, is diagrammed with its typical position in a muscle (left), neuronal connections in spinal cord (middle), and expanded schematic (right). The tendon organ is a stretch receptor that signals the amount of force on the muscle and protects the muscle from excessively heavy loads by causing the muscle to relax and drop the load.

9.3.5 Key Points

- The stretch reflex is a monosynaptic reflex that regulates muscle length through neuronal stimulation at the muscle spindle. The alpha motor neurons resist stretching by causing contraction, and the gamma motor neurons control the sensitivity of the reflex.
- The stretch and Golgi tendon reflexes work in tandem to control muscle length and tension. Both are examples of ipsilateral reflexes, meaning the reflex occurs on the same side of the body as the stimulus.

- The crossed extensor reflex is a contralateral reflex that allows the body to compensate on one side for a stimulus on the other. For example, when one foot steps on a nail, the crossed extensor reflex shifts the body's weight onto the other foot, protecting and withdrawing the foot on the nail.
- The withdrawal reflex and the more-specific pain withdrawal reflex involve withdrawal in response to a stimulus (or pain). When pain receptors, called nociceptors, are stimulated, reciprocal innervations stimulate the flexors to withdraw and inhibit the extensors to ensure they are unable to prevent flexion and withdrawal.

9.3.6 Key Terms

- **golgi tendon reflex**: A normal component of the reflex arc of the peripheral nervous system. In this reflex, a skeletal muscle contraction causes the agonist muscle to simultaneously lengthen and relax. This reflex is also called the inverse myotatic reflex because it is the inverse of the stretch reflex. Although muscle tension is increasing during the contraction, the alpha motor neurons in the spinal cord that supply the muscle are inhibited. However, antagonistic muscles are activated.
- **alpha motor neuron**: These are large, lower motor neurons of the brainstem and spinal cord. They innervate the extrafusal muscle fibers of skeletal muscle and are directly responsible for initiating their contraction. Alpha motor neurons are distinct from gamma motor neurons that innervate the intrafusal muscle fibers of muscle spindles.

9.3.7 ATTRIBUTIONS

CC LICENSED CONTENT, SHARED PREVIOUSLY

- Adapted from Anatomy and Physiology - Section 12.10C. **Authored and Provided by**: Boundless.com. **License**: ***CC BY-SA: Attribution-ShareAlike***

This page titled 9.3: Spinal Reflexes is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

9.4: Motor Circuitry- Neural Structures and Pathways

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Identify the main structures of neural motor circuits at the cortical, subcortical, and spinal levels
- Describe the main functional roles of the basal ganglia, premotor area, and motor cortex
- Apply knowledge of functional cortical anatomy to the control of fine and gross body movements

9.4.1 Neuroanatomy of Motor Systems

The motor system controls all of our skeletal muscle movement and more. There are multiple levels of control. Within the spinal cord, simple reflexes can function without higher input from the brain. Slightly more complex spinal control occurs when central pattern generators function during repetitive movements like walking. The motor, premotor and supplementary cortices in the brain are responsible for the planning and execution of voluntary movements. And finally, the basal ganglia and cerebellum modulate the responses of the neurons in the motor cortex to help with coordination, motor learning, and balance.

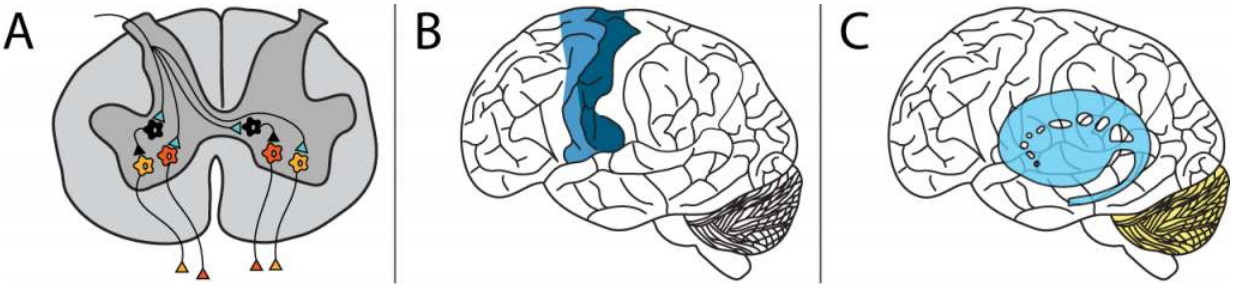


Figure **9.4.1**: Motor output is controlled at multiple levels: A. Spinal cord and spinal neurons, B. Motor (dark blue) and premotor/supplementary motor cortices (light blue with supplementary motor located in top portion), C. Basal ganglia (a subcortical structure shown in light blue) and cerebellum (yellow). 'Motor Control Levels' by Casey Henley is licensed under a CC BY-NC-SA 4.0 International License.

9.4.1.1 **Cortical Anatomy of Movement**

Much of the cortex is actually involved in the planning of voluntary movement. Sensory information, particularly the dorsal stream of the visual and somatosensory pathways, is processed in the posterior parietal cortex where visual, tactile, and proprioceptive information is integrated.

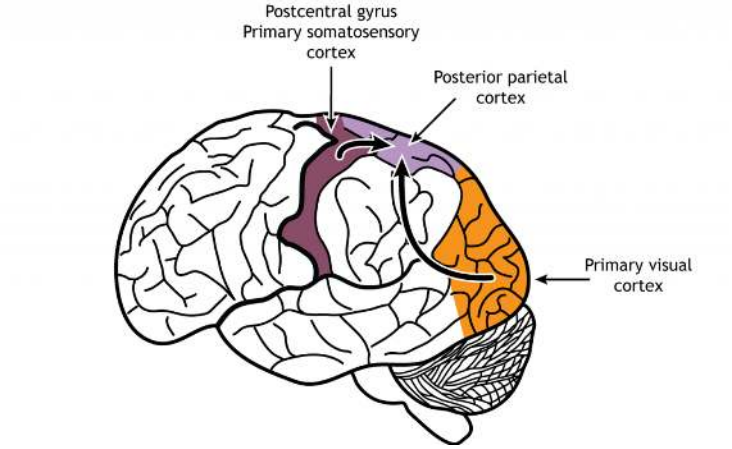


Figure **9.4.2**: Information from multiple sensory systems is processed in the posterior parietal cortex. Projections are sent from the primary somatosensory and primary visual cortices. 'Posterior Parietal by Casey Henley is licensed under a CC BY-NC-SA 4.0 International License.

Connections from the posterior parietal cortex are then sent to both the premotor regions and the prefrontal cortex. The prefrontal cortex, which is located in the front of the brain in the frontal lobe, plays an important role in higher level cognitive functions like planning, critical thinking, and understanding the consequences of our behaviors. The premotor area lies just anterior to the primary motor cortex. This region helps plan and organize movement and makes decisions about which actions should be used for a situation.

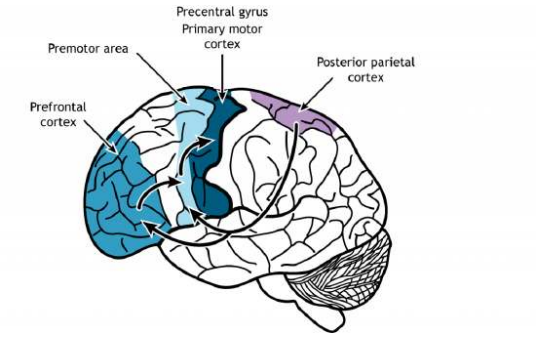


Figure **9.4.3**: Sensory information from the posterior parietal cortex is processed in the prefrontal cortex and premotor area, both located in the frontal cortex. These areas then send information to the primary motor cortex located in the precentral gyrus. 'Motor Regions' by Casey Henley is licensed under a CC BY-NC-SA 4.0 International License.

9.4.1.1.1 Role of Premotor Area

The premotor regions send some axons directly to lower motor neurons in the spinal cord using the same pathways as the motor cortex. However, the premotor cortex also plays an important role in the planning of movement. Two experimental designs have demonstrated this role. Monkeys were trained on a panel that had one set of lights in a row on top and one set of buttons that could also light up in a row on the bottom. The monkeys would watch for a top row light to turn on. This would indicate that within a few seconds, the button directly below would light up. When the button turned on, the monkeys were supposed to push the button.

Therefore, there were two light triggers in the experiment. The first required no motor movement from the monkey but did give the monkey information about where a motor movement would be needed in the near future. The second required the monkey to move to push the button. When brain activity was measured during this study, neurons in the premotor cortex became active when the first light trigger turned on, well before any movement actually took place (Weinrich and Wise, 1928).

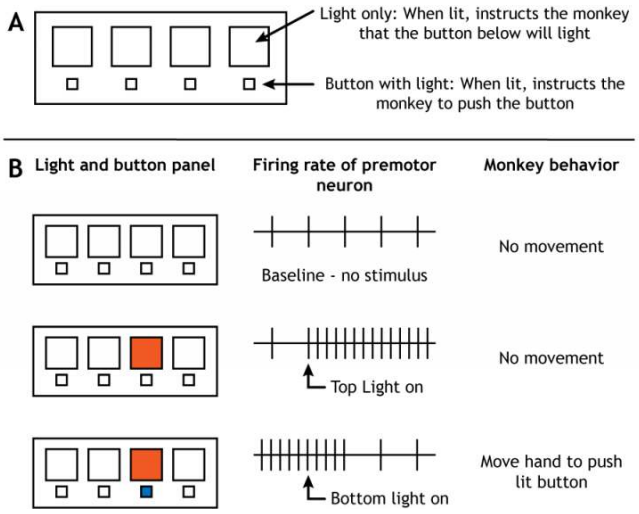


Figure 9.4.4: (A) Monkeys were trained to use a light panel. When lit, a light located in the top row informed the monkey that the button directly below would light up soon. Shortly after one of the top row lights turned on, the button directly below it would turn on, and the monkey would need to push that button for a reward. (B) When no lights are lit, premotor neurons fire at a baseline rate, and the monkey does not move. Neurons located in the premotor cortex increase action potential firing rate when the top row light turns on, even though the monkey makes no movement. The firing stops shortly after the monkey moves to push the bottom button after it turns on. (Based on Weinrich and Wise, 1928) 'Light Panel Experiment' by Casey Henley is licensed under a CC BY-NC-SA 4.0 International License.

In another experiment, people were trained to move their fingers in a specific pattern. Cerebral blood flow was then measured when they repeated the finger pattern and when they only imagined repeating the finger pattern. When the movement was only imagined and not actually executed, the premotor regions along with parts of the prefrontal cortex were activated (Roland, et al, 1980).

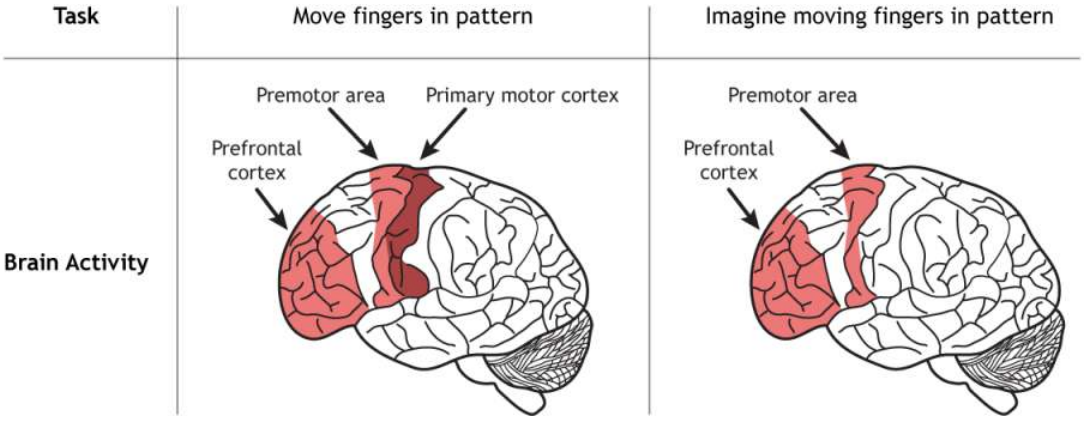


Figure 9.4.5: People were trained to move their fingers in a pattern. When the pattern was repeated, brain activity was seen in the primary motor cortex, along with the premotor area and prefrontal cortex. When the pattern was only imagined, and no finger movement took place, brain activity was seen in the premotor area and prefrontal cortex. (Based on Roland, et al, 1980) 'Finger Movement Experiment' by Casey Henley is licensed under a CC BY-NC-SA 4.0 International License.

These studies show that the premotor cortex is active prior to the execution of movement, indicating that it plays an important role in the planning of movement. The posterior parietal, prefrontal, and premotor regions, though, also communicate with a subcortical region called the basal ganglia to fully construct the movement plan. The basal ganglia are covered in the next subsection.

9.4.1.1.2 Role of Motor Cortex

Once the plan for movement has been created, the primary motor cortex is responsible for the execution of that action. The primary motor cortex lies just anterior to the primary somatosensory cortex in the precentral gyrus located in the frontal lobe.

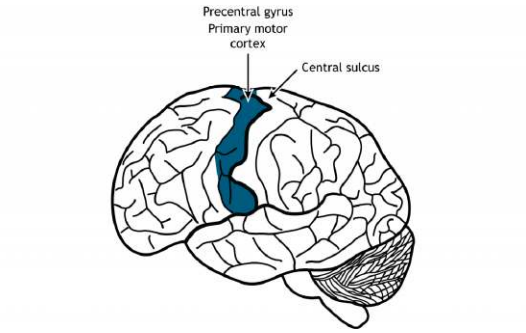


Figure 9.4.6: The primary motor cortex is located in the frontal lobe in the precentral gyrus, just anterior to the central sulcus. 'Primary Motor Cortex' by Casey Henley is licensed under a CC BY-NC-SA 4.0 International License.

Like the somatosensory cortex, the motor cortex is organized by a somatotopic map in that different areas of the body are controlled by distinct areas of the motor cortex. However, the motor cortex does not map onto the body in such an exact way as does the somatosensory system. It is believed that upper motor neurons in the motor cortex control multiple lower motor neurons in the spinal cord that innervate multiple muscles. This results in activation of an upper motor neuron causing excitation or inhibition in different neurons at once, indicating that the primary motor cortex is responsible for movements and not simply activation of one muscle. Stimulation of motor neurons in monkeys can lead to complex motions like bringing the hand to the mouth or moving into a defensive position (Graziano et al, 2005).

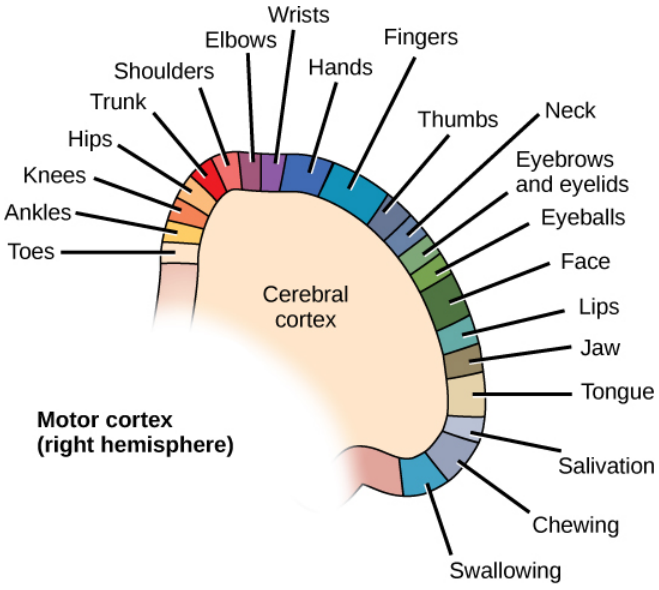


Figure 9.4.7: Different regions of the motor cortex control different muscle groups and their neighboring regions control neighboring muscle groups. Although the right hemisphere is shown, a mirror-image exists in the left hemisphere as well. Note that all areas of the body that have voluntary motor control are represented on the primary motor cortex. License: CC BY 4.0, via Wikimedia Commons

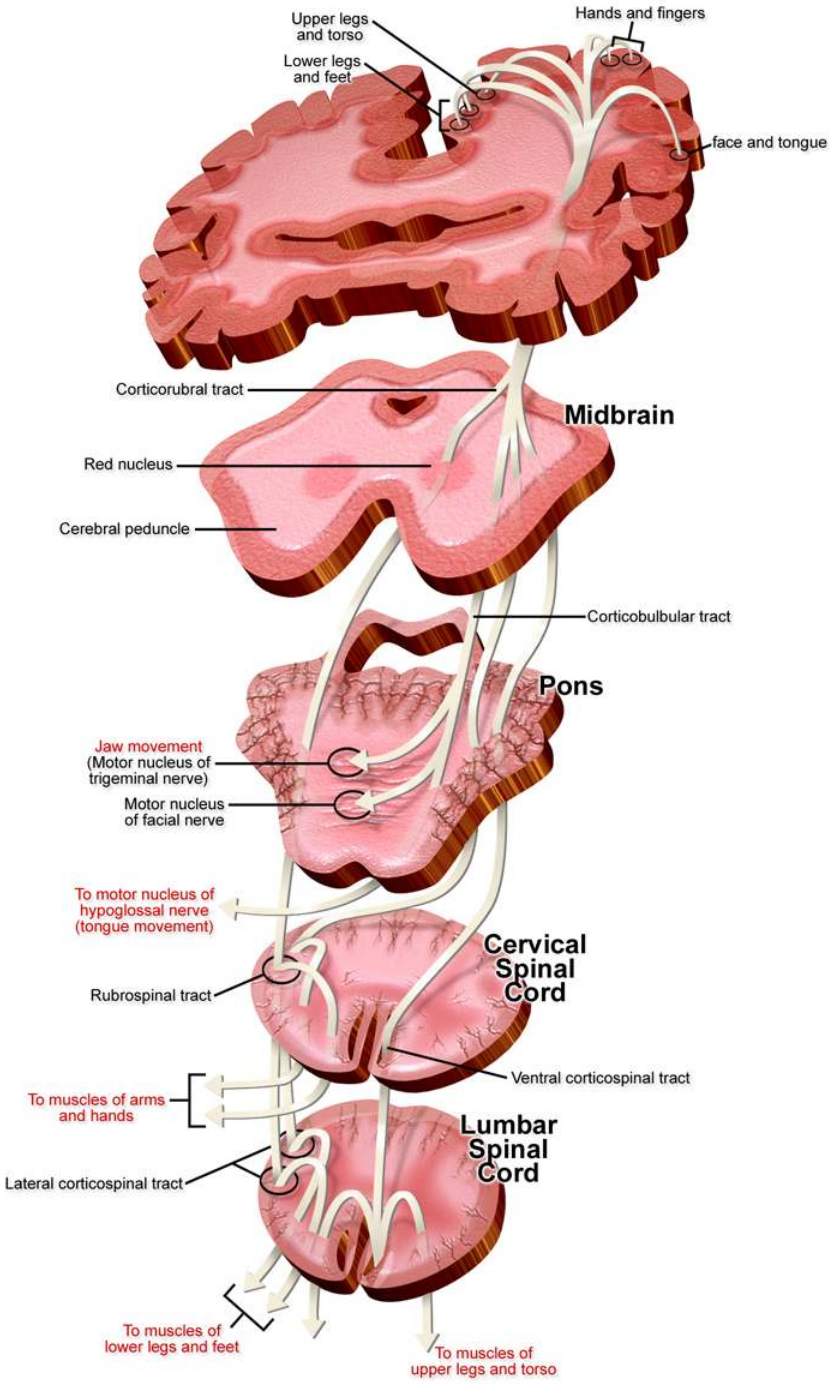


Figure 9.4.8: Descending Motor Pathways from Primary Motor Cortex to Spinal Cord. By BruceBlaus CC BY-SA 4.0, via Wikimedia Commons

9.4.1.2 Basal Ganglia

The basal ganglia are a group of subcortical nuclei, meaning groups of neurons that lie below the cerebral cortex. The basal ganglia is comprised of several nuclei including the caudate nucleus, putamen, and the globus pallidus. Another nucleus typically associated with the basal ganglia is the subthalamic nucleus. The substantia nigra of the midbrain provides dopaminergic input to these areas.

The basal ganglia are primarily associated with motor control, since motor disorders, such as Parkinson's or Huntington's diseases stem from dysfunction of neurons within the basal ganglia. For voluntary motor behavior, the basal ganglia are involved in the

initiation or suppression of behavior and can regulate movement through modulating activity in the thalamus and cortex. In addition to motor control, the basal ganglia also communicate with non-motor regions of the cerebral cortex and play a role in other behaviors such as emotional and cognitive processing.

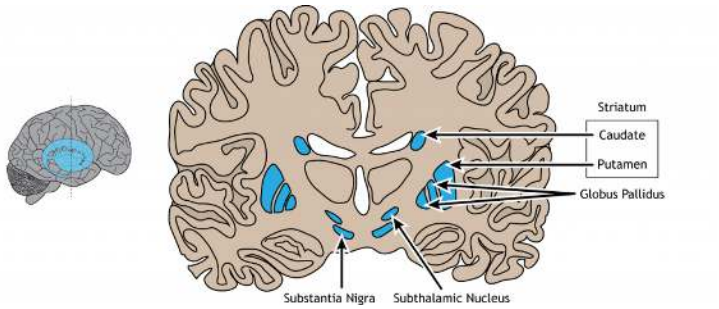


Figure 9.4.9: The basal ganglia are subcortical structures located at the base of the forebrain. They are comprised of the caudate and putamen, which both make up the striatum, as well as the globus pallidus, substantia nigra, and subthalamic nucleus. 'Basal Ganglia' by Casey Henley is licensed under a CC BY-NC-SA 4.0 International License.

9.4.1.2.3 Basal Ganglia Input

The majority of information processed by the basal ganglia enters through the striatum. The principal source of input to the basal ganglia is from the cerebral cortex. This input is glutamatergic (i.e. uses glutamate as its neurotransmitter) and therefore, excitatory. The substantia nigra is also a region with critical projections to the striatum and is the main source of dopaminergic input. Dopamine plays an important role in basal ganglia function. Parkinson's disease results when dopamine neurons in the substantia nigra degenerate and no longer send appropriate inputs to the striatum. Dopamine projections can have either excitatory or inhibitory effects in the striatum, depending on the type of metabotropic dopamine receptor the striatal neuron expresses. Dopamine action at a neuron that expresses the D1 receptor is excitatory. Dopamine action at a neuron that expresses the D2 receptor is inhibitory.

9.4.1.2.4 Basal Ganglia Output

The primary output region of the basal ganglia is the internal segment of the globus pallidus. This region sends inhibitory GABAergic projections to nuclei in the thalamus. This inhibitory output has a tonic, constant firing rate, which allows the basal ganglia output to both increase and decrease depending on the situation. The thalamus then projects back out to the cerebral cortex, primarily to motor areas.

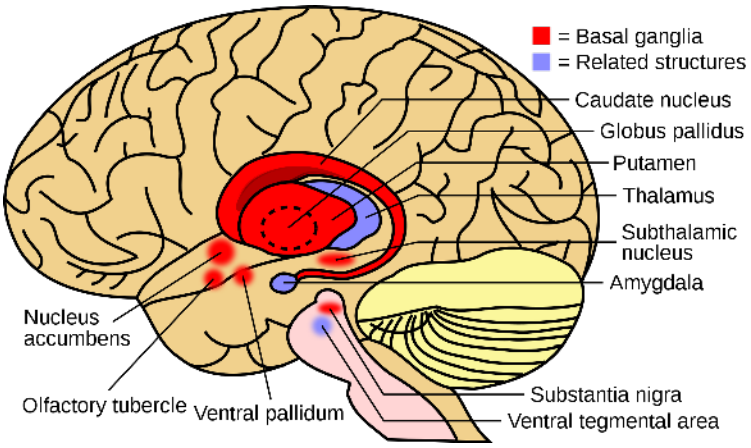


Figure 9.4.10: Basal Ganglia and Associated Structures. CC BY-SA 4.0, via Wikimedia Commons

9.4.2 Attributions:

- Adapted from Foundations of Neuroscience - Chapter 26. Authored by Casey Henley is licensed under a **License:** (CC BY-NC-SA) 4.0 International License. https://openbooks.lib.msu.edu/neuros...g-of-movement/

This page titled 9.4: Motor Circuitry- Neural Structures and Pathways is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

9.5: Impact of Disease States and Neural Damage on Motor Control

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Describe the general symptoms of the major motor disorders of the nervous system
- Categorize motor disorders based on their underlying neuroantomical or neurochemical mechanisms
- Explain some of the treatments for some of the major disorders

9.5.1 Motor Disorders and Their Neuroanatomical/Neurochemical Mechanisms

Given that motor control involves such an extensive collection of structures and circuits throughout both the central and peripheral nervous systems, it is amazing how well and, often effortlessly, these systems function every second of our lives. However, the comprehensive nature of motor systems also allows for multiple avenues of potential damage or degeneration which can lead to profound consequences in one's ability to move or even survive.

Damage or disease can occur in regions of the forebrain associated with Upper Motor Neurons in the cortex or the basal ganglia, and/or Lower Motor Neuron regions of the medulla or spinal cord. These anatomical distinctions can be critical in the diagnosis and treatment of specific motor disorders.

9.5.1.1 **Multiple Sclerosis**

Multiple sclerosis (MS) is a progressive motor disease that attacks neurons in both the central and peripheral nervous systems. The main mode of this attack is the destruction of the myelin coating around the axon which ultimately will slow down action potential conductance. As axons are demyelinated, inflammatory patches called lesions are formed. As the disease progresses, the myelin-producing oligodendrocytes and, ultimately, the axons themselves are destroyed. There is compelling evidence that the destruction is caused by selective activation of the cellular immune system and inflammatory molecules, thus supporting the idea that MS is an autoimmune disease.

Findings from MRI studies indicate that an individual with MS may have abnormalities around sites like the lateral ventricle, optic nerve, brainstem, spinal cord, cerebellum and other areas (see Figure 9.5.1 below).

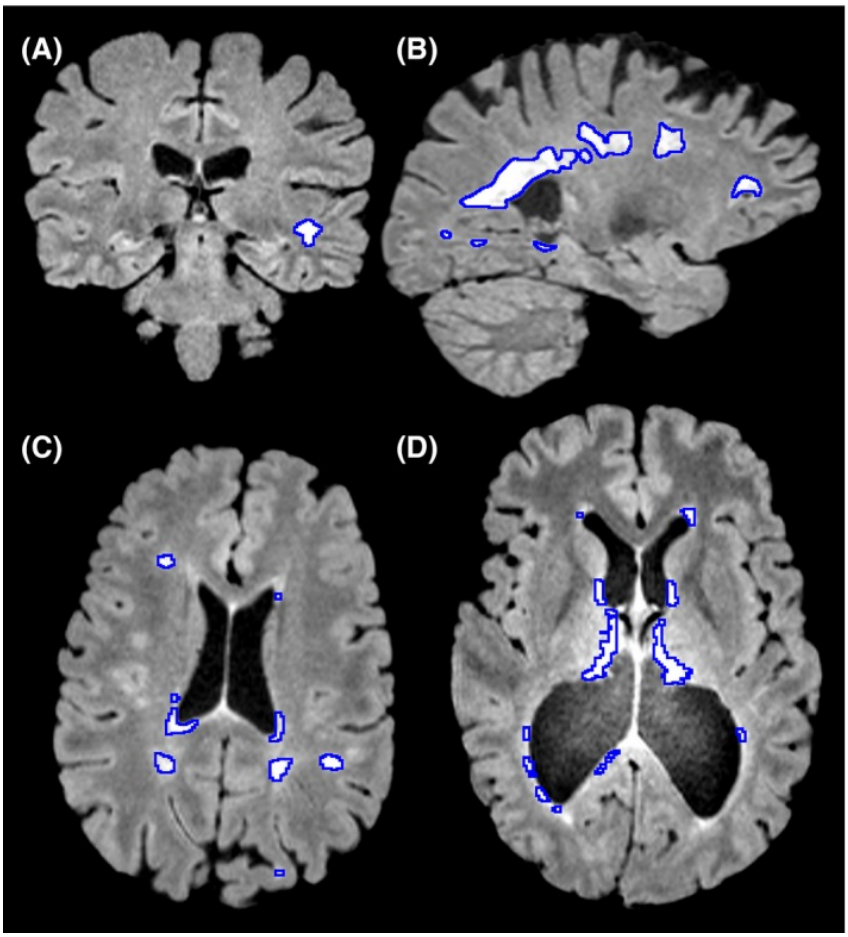


Figure **9.5.1**: MRI images showing white hyperintensities in different areas of the brain of an individual experiencing symptoms of multiple sclerosis. (A) A coronal slice showing a lesion in peripheral white matter. (B) A sagittal slice showing large periventricular lesions. (C) An axial slice showing both periventricular and peripheral white matter lesions. (D) A case where severe atrophy caused midline false positives to not be removed, as they were further from midline than expected. Image from National Library of Medicine. License: CC BY 4.0.

9.5.1.1.1 Symptoms of MS

Multiple sclerosis also presents itself differently in acute and chronic phases. During the acute phase, the condition is associated with intermittent symptoms, whereas the chronic phase is associated with progressive forms of the disease and increased severity of symptoms.

Most people experience their first symptoms of MS between the ages of 20 and 40 with initial blurred or double vision, red-green color distortion, or even blindness in one eye. Most MS patients also experience muscle weakness in their extremities and difficulty with coordination and balance. These symptoms may be severe enough to impair walking or even standing. In the worst cases, MS can produce partial or complete paralysis.

Most people with MS also exhibit paresthesias, transitory abnormal sensory feelings such as numbness, prickling, or "pins and needles" sensations. Some may also experience pain. Speech impediments, tremors, and dizziness are other frequent complaints. Occasionally, people with MS have hearing loss. Approximately half of all people with MS experience cognitive impairments such as difficulties with concentration, attention, memory, and poor judgment, but such symptoms are usually mild and are frequently overlooked. Depression is another common feature of MS.

9.5.1.1.2 **Potential Causes of MS**

Currently, there is no known cause for MS although many different hypotheses have been put forward. Surprisingly, there are no known associative genes that have been implicated in causing MS, and scientists believe there might be a complex interaction

between genes and environmental factors that produce demyelination. The most common theories about the cause(s) of multiple sclerosis include:

1. Viral infection resulting in an autoimmune reaction
2. Genetic factors: inherited predisposition possibly through the immune system, although a mutation in a gene has not been discovered
3. Environmental factor(s) which might work with genetics such as low vitamin D levels or smoking.

Although this is a well-characterized disorder, we still don't know much about its causes.

9.5.1.2 **Parkinson's Disease**

Parkinson's disease (PD) belongs to a group of conditions called motor system disorders, which cause unintended or uncontrollable movements of the body. The precise cause of PD is unknown, but some cases are hereditary while others are thought to occur from a combination of genetics and environmental factors that trigger the disease. In PD, brain cells become damaged or die in the substantia nigra which produces the neurotransmitter dopamine--a chemical needed to produce smooth, purposeful movement.

9.5.1.2.1 Symptoms of PD

The four primary symptoms of PD are:

- tremor--shaking that has a characteristic rhythmic back and forth motion
- rigidity--muscle stiffness or a resistance to movement, where muscles remain constantly tense and contracted
- bradykinesia--slowing of spontaneous and automatic movement that can make it difficult to perform simple tasks or rapidly perform routine movements
- postural instability--impaired balance and changes in posture that can increase the risk of falls.

Other symptoms may include difficulty swallowing, chewing, or speaking; emotional changes; urinary problems or constipation; dementia or other cognitive problems; fatigue; and problems sleeping.

PD usually affects people around the age of 70 years but can occur earlier. PD affects men more than women. Currently there are no specific tests that diagnose sporadic PD.

9.5.1.2.2 **Potential causes of PD**

The current model for most neurodegenerative conditions is that neurons die via apoptosis, a programmed form of cell death. Although this type of cell death occurs during normal development, neurodegenerative diseases such as Parkinson's Disease and Alzheimer's Disease are believed to use apoptotic pathways. This suggests that there is a component protein that is aberrantly folded to produce apoptosis. This has very important clinical implications as turning on or turning off specific genes/proteins is a method to stop or possibly reverse neurodegeneration.

Most cases of this disorder appear to be sporadic (i.e. non-genetic in origin) although a very few cases seem to have a genetic origin. Parkinson's Disease is most often associated with the loss of "pigmented" nuclei in the brain and typically involves the loss of a group of neurons found in the substantia nigra (see Figure **9.5.2**). The substantia nigra neurons are dopaminergic and are pigmented because they contain the protein *melanin*.

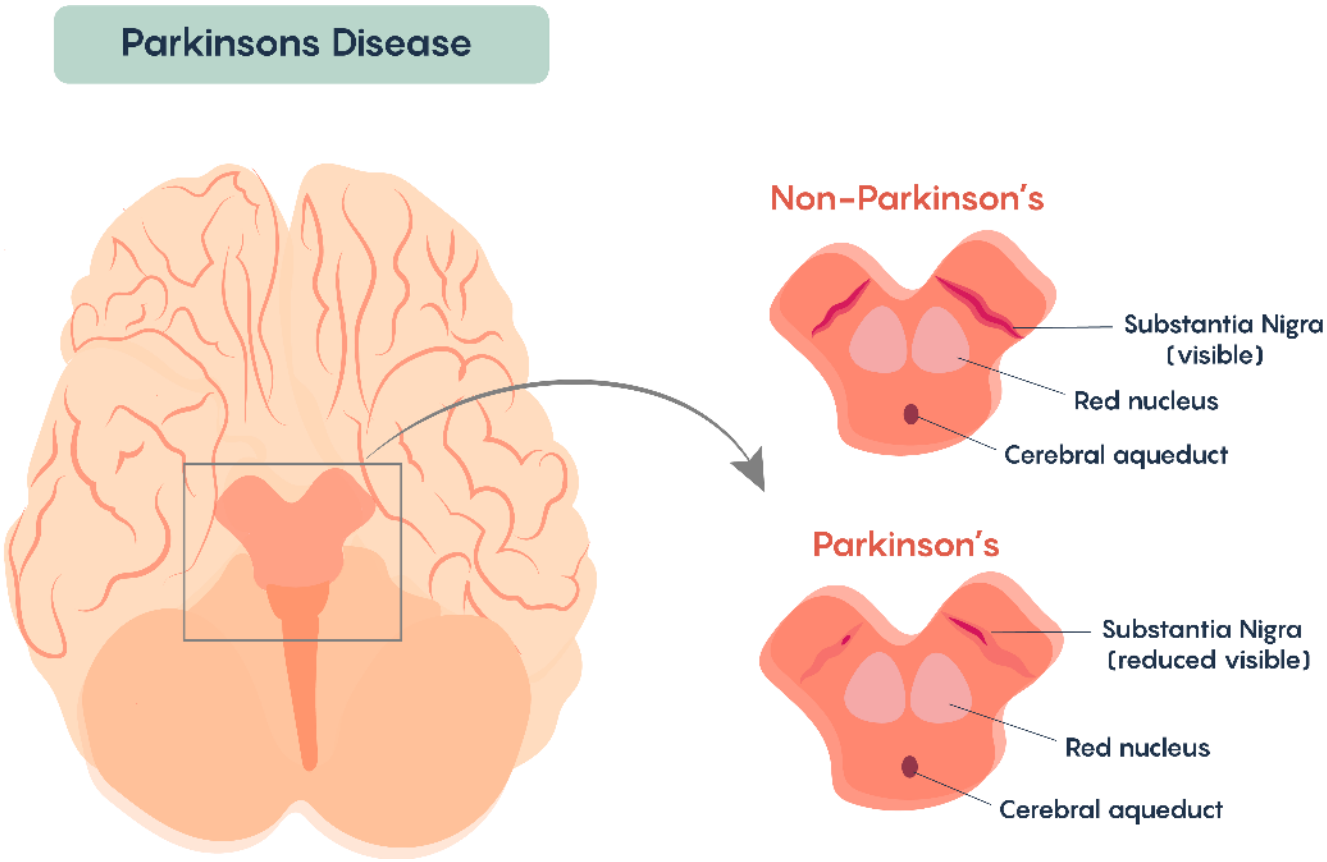


Figure **9.5.2**: Loss of the pigmented dopaminergic neurons within the Substantia Nigra of individuals with Parkinson's Disease.

Loss of the substantia nigra cells affects the processing and execution of voluntary movement in individuals with Parkinson's Disease. Similar to MS, once diagnosed, the symptoms become continuous and progressive – i.e. the symptoms worsen over time. Again, there are many similarities with MS, and there is no known cure for Parkinson's Disease. The disease remains idiopathic although there are some known causes of Parkinson's Disease including loss of movement following cerebral atherosclerosis, viral encephalitis, and as the result of side effects from drugs such as phenothiazides and reserpine.

Alpha-synuclein is a naturally occurring protein within neurons. Mutations in the PARK1 and PARK4 genes which normally encodes for alpha-synuclein have been associated with Parkinson's Disease. As such, many animal models of Parkinson's Disease look for the production of the fibrillary form of alpha-synuclein as it misfolds and then accumulates within the substantia nigral neurons known as Lewy bodies (see Figure **9.5.3**). The misfolding and accumulation of alpha-synuclein has been hypothesized to be the reason that neurons undergo apoptosis, although the exact mechanism for how this occurs remains to be elucidated.

Lewy bodies and alpha-synuclein – hallmark features of Parkinson's Disease

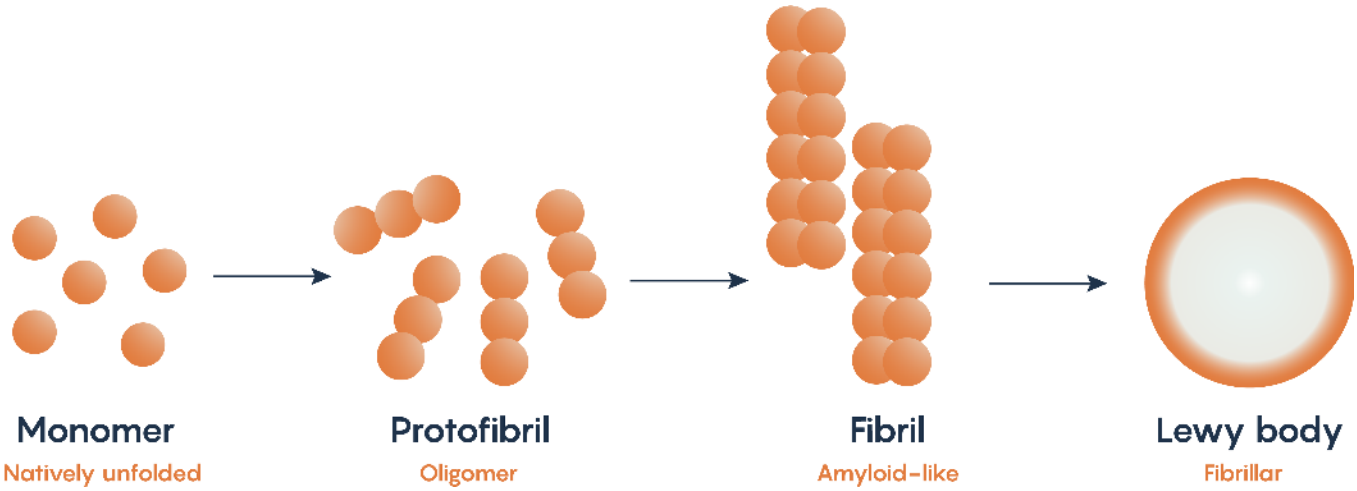


Figure 9.5.3:*Misfolding and subsequent accumulation of alpha-synuclein.*

9.5.1.2.3 Treatment of PD

At present, there is no cure for PD, but a variety of medications provide dramatic relief from the symptoms. Usually, affected individuals are given levodopa (l-dopa) combined with carbidopa. Carbidopa delays the conversion of levodopa into dopamine until it reaches the brain. Nerve cells can use levodopa to make dopamine and replenish the brain's dwindling supply. Although levodopa helps most people with PD, not everyone responds equally to the drug. The symptoms of bradykinesia and rigidity respond best, while tremor may be only marginally reduced. Problems with balance and other symptoms may not be alleviated at all.

There are a large variety of other drugs that affect the dopamine system and treat symptoms of PD. Some mimic the role of dopamine in the brain, causing the nerve cells to react as they would to dopamine, while others prolong the effects of levodopa by preventing the breakdown of dopamine in the brain. In addition to these dopaminergic drugs, anticholinergic drugs have been shown to help control symptoms of tremor and rigidity.

In some cases, surgery may be appropriate if the disease doesn't respond to drugs. One option is deep brain stimulation (DBS), in which electrodes are implanted into the brain and connected to a small electrical device called a pulse generator to painlessly stimulate the brain to block signals that cause many of the motor symptoms of PD. DBS is generally appropriate for people with levodopa-responsive PD who have developed dyskinesias or other disabling "off" symptoms despite drug therapy. However, DBS does not stop PD from progressing and some problems may gradually return.

9.5.1.3 Huntingtons Disease

Huntington's disease (HD) is an inherited disorder which causes the death of select populations of neurons. The affected neurons are located in various areas of the brain, including those in the basal ganglia which help to control voluntary (intentional) movement.

9.5.1.3.1 Symptoms of HD

Symptoms of the disease, which gets progressively worse, include uncontrolled movements (chorea), abnormal body postures, and changes in behavior, emotion, judgment, and cognition. People with HD also develop impaired coordination, slurred speech, and difficulty feeding and swallowing.

HD typically begins between ages 30 and 50. An earlier onset form called juvenile HD occurs under age 20. Its symptoms differ somewhat from adult onset HD and include rigidity, slowness, difficulty at school, rapid involuntary muscle jerks called myoclonus, and seizures. More than 30,000 Americans have HD.

Huntington’s disease causes disability that gets worse over time. Currently no treatment is available to slow, stop, or reverse the course of HD. People with HD usually die within 10 to 30 years following symptom onset, most commonly from infections (most often pneumonia) and injuries related to falls.

9.5.1.3.2 Causes of HD

Huntington's disease is caused by a mutation in the gene for a protein called huntingtin. The mutated gene includes an increased number of repeats of a select portion of it's normal genetic code. How these increased repeats lead to the disorder is still unclear. Each child of a parent with HD has a 50-50 chance of inheriting the mutated gene. A child who does not inherit the HD gene will not develop the disease and cannot pass it to subsequent generations. Since this is an autosomal dominant trait, a person who inherits the HD gene will eventually develop the disease. HD is generally diagnosed based on a genetic test, medical history, brain imaging, and neurological and laboratory tests.

9.5.1.3.3 Treatment of HD

There is no treatment that can stop or reverse the course of HD. The drugs, tetrabenazine and deuterabenazine can treat chorea associated with HD. Antipsychotic drugs may ease chorea and help to control hallucinations, delusions, and violent outbursts. Drugs may be prescribed to treat depression and anxiety. Side effects of drugs used to treat the symptoms of HD may include fatigue, sedation, decreased concentration, restlessness, or hyperexcitability, and should only be used when symptoms create problems for the individual.

9.5.1.4 Amyotrophic lateral sclerosis (Formally known as Lou Gehrig's Disease)

Amyotrophic lateral sclerosis (ALS) is a rare neurological disease that primarily affects the neurons responsible for controlling voluntary muscle movement. Voluntary muscles produce movements like chewing, walking, and talking. The disease is progressive, meaning the symptoms get worse over time. Currently, there is no cure for ALS and no effective treatment to halt or reverse the progression of the disease.

ALS belongs to a wider group of disorders known as motor neuron diseases, which are caused by gradual deterioration (degeneration) and death of motor neurons. As motor neurons degenerate, they stop sending messages to the muscles and the muscles gradually weaken, start to twitch, and waste away (atrophy). Eventually, the brain loses its ability to initiate and control voluntary movements.

Early symptoms of ALS usually include muscle weakness or stiffness. Gradually all voluntary muscles are affected, and individuals lose their strength and the ability to speak, eat, move, and even breathe. Most people with ALS die from respiratory failure, usually within three to five years from when the symptoms first appear. However, about 10 percent of people with ALS survive for 10 or more years.

Because people with ALS usually can perform higher mental processes such as reasoning, remembering, understanding, and problem solving, they are aware of their progressive loss of function and may become anxious and depressed. A small percentage of individuals may experience problems with language or decision-making, and there is growing evidence that some may even develop a form of dementia over time.

9.5.1.4.1 Possible Causes of ALS

The cause of ALS is not known, and scientists do not yet know why ALS strikes some people and not others.

Motor neuronal hyperexcitability potentially contributes to motor neuron death in ALS. Many previous studies support this hypothesis, including clinical trials of riluzole, a glutamate antagonist that delays ALS progression (Miller et al., 2012).

In addition to the possible role of motor neuron dysfunction, scientific evidence suggests that both genetics and environment play a role in motor neuron degeneration and the development of ALS.

In 1993, scientists supported by the National Institute of Neurological Disorders and Stroke (NINDS) discovered that mutations in the *SOD1* gene were associated with some cases of familial ALS. Since then, more than a dozen additional genetic mutations have been identified, many through NINDS-supported research.

Research on certain gene mutations suggests that changes in the processing of RNA molecules may lead to ALS-related motor neuron degeneration. RNA molecules are involved with the production of molecules in the cell and with gene activity.

Other gene mutations indicate there may be defects in protein recycling—a naturally occurring process in which malfunctioning proteins are broken down and used to build new working ones. Still others point to possible defects in the structure and shape of motor neurons, as well as increased susceptibility to environmental toxins.

Researchers are studying the impact of environmental factors, such as exposure to toxic or infectious agents, viruses, physical trauma, diet, and behavioral and occupational factors. For example, exposure to toxins during warfare, or strenuous physical activity, are possible reasons for why some veterans and athletes may be at increased risk of developing ALS. Ongoing research may show that some factors are involved in the development or progression of the disease.

9.5.2 Attributions:

Selections of Parkinsons Disease and Multiple Sclerosis sections adapted from Neuroscience Canadian, 2nd Edition. **License:** CC-BY 4.0.

Amyotrophic Lateral Sclerosis section adapted from National Institute of Neurological Disorders and Stroke (NINDS). **License:** ***Public Domain: No Known Copyright***

Miller RG, Mitchell JD, Moore DH. (2012) Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). Cochrane Database Syst Rev., Mar 14;2012(3):CD001447. doi: 10.1002/14651858.CD001447.pub3.

This page titled 9.5: Impact of Disease States and Neural Damage on Motor Control is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

CHAPTER OVERVIEW

10: Learning and Memory

> This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

10.1: Learning, Genes, and Adaptation
10.2: Types of Learning and Biological Adaptation
10.3: Introduction to Memory and the Brain
10.4: Synaptic Mechanisms of Learning and Memory
10.5: Brain Mechanisms of Memory Disorders
10.6: Localization of Types and Stages of Memory
10.7: Hormones, Learning, and Memory

This page titled 10: Learning and Memory is shared under a mixed license and was authored, remixed, and/or curated by Kenenth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

10.1: Learning, Genes, and Adaptation

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe several definitions of learning which vary depending on theoretical perspective.
2. Discuss the claim that behavior is not random, but orderly and lawful.
3. Describe the fundamental law that governs the organization of behavior and the origin of that natural law.
4. Explain the similarities and differences between information in genes and learned information.
5. Discuss the nature-nurture continuum.
6. Describe examples of "biological preparedness" or biological constraints on learning.

10.1.1 Overview

In this section, we examine learning and how it contributes to the adaptive organization of mind and behavior. To understand learning within its broader biological context, it is important to compare it with genetic information, the other major source of information that organizes behavior into adaptive patterns. Learning refers to behavioral change as a result of experience, but learning is not a single, unitary phenomenon. Instead there are many types of learning which are specialized to adapt to various features of the environment and to solve a variety of adaptive problems. In other words, learning evolved in many forms to serve a range of adaptive functions. Learned and genetic sources of information interact to organize behavior into adaptive patterns. The relative contribution of each varies with the species and the behavior in question.

10.1.2 Learned and Genetic Sources of Information Shape Adaptive Behavior

Learning is defined differently by different psychologists depending upon the theoretical emphasis of the psychologist. Strict behaviorists rejected explanations of behavior that involved reference to internal mental processes. Instead, they emphasized the measurable relations between observable behaviors and observable stimuli. Consequently, behaviorists usually define learning as a change in behavior as a result of experience (excluding changes due to fatigue or other special circumstances). Cognitive psychologists, partially in reaction against strict behaviorism, are more likely to define learning as acquisition of information as a result of experience. From an evolutionary perspective, learning can be thought of as the acquisition of information during the lifetime of the individual animal. This last phrase differentiates learned information from genetic information which is acquired over the evolutionary history of the species.

Behavior is not random. Instead, like all of nature, behavior follows natural laws. An evolutionary perspective applied to learning emphasizes the role of learning as a means of achieving successful adaptation. Certainly, the ability to learn new behaviors in response to new situations encountered in the environment is valuable to survival and therefore increases the chances for reproduction and the transmission of one's genes into future generations, including genes for learning itself. This consideration is important because it suggests how and why learning evolved. It also suggest why behavior is orderly and lawful. Behavior is organized to promote survival and reproduction. This outcome is an inevitable product of the forces of evolution, most notably, natural selection. Thus, learning itself is an evolved property of nervous systems. The properties and laws of learning have evolved and are consistent with the primary law of evolution, natural selection. Learning reflects this law so that learned information organizes behavior in ways that serve an animal's biological fitness in the "struggle for existence" (to the extent that this is not true for any individual animal or human, we recognize some form of behavioral or psychological pathology). In section 10.2, we discuss in more detail how different forms of learning contribute to adaptation and improve biological fitness to particular features of the environment.



Figure 10.1.1: Dusky Langur infant and mother. In the first few months the infant is constantly breast-fed. The mother must teach her infant to eat leaves, as shown in this photo. (Image and caption adapted from Wikimedia Commons; File:Dusky Langur infant learning to eat leaves.jpg; https://commons.wikimedia.org/wiki/F...eat_leaves.jpg; by: Roughdiamond21; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

Behavior, and other traits of organisms, ultimately are organized by information. Just like a builder needs information in the form of structural plans to construct a building, information is required to build and operate an organism, including the guidance of its movement, its behavior. In short, behavior is not random, but orderly and lawful, because it is organized by information selected and filtered by two processes. Learning is one source of the information that organizes behavior into adaptive patterns. The laws of learning tend to organize learned behavior in ways that serve adaptation to the environment. This is because the laws of learning themselves are evolved properties of brains, and natural selection shaped these laws into adaptive form so that learned behavior tends to increase survival and reproduction. The second informational source for the adaptive organization of living things is in genes, already discussed briefly above, and in some detail in Chapter 3 on evolution and genetics. From earlier discussion, we know that genetic information is filtered over countless generations by natural selection and other processes of evolution (see Chapter 3) so that the information transmitted across generations in genes generates traits that, in general, contribute to survival and reproduction (genetic abnormalities are, of course, an exception). In short, information is required to organize living things and their functioning. This information comes from two sources: 1) genes and genetic evolution; and 2) learning and laws of learning (in those organisms that are capable of learning).

Let's compare these two sources of information which organize behavior into forms which help an animal adapt to environmental demands (see Table 10.1.1 below).

First, genetic information is available to all living organisms on earth, but learned information is available only to species whose brains are equipped with properties that make learning possible. Do flies learn? Do cockroaches? Do frogs? If they do, how much do they learn, how readily do they learn? How important is learning in the life of a frog compared to how important learning is in the life of a dog, a chimpanzee, or a human? In general, learning is more characteristic of the brains of the more complex animals, especially the mammals. Nevertheless, learning does occur even in many invertebrates (animals without backbones, such as insects or squid and octopus; I once visited the lab of a psychologist at the University of Hawaii who was studying learning in bees). Usually, however, the more complex the animal's brain, the more likely it is that the animal has well developed capacities for learning--capacities for the acquisition of information during its individual lifetime.

Secondly, genetic information has been acquired and transmitted generation after generation over the entire genetic history of the species. For example, think of flies, or pelicans. As you know from your own personal observation, flies are very good at flying. And so are pelicans. They ride the air currents perfectly and skim the ocean only inches above it, their paths rising and falling with the tops of breaking waves. As the wave rises they rise just enough to stay inches above the wave while flying at perhaps 30 miles per hour or more. When the pelican dives to catch a fish, it tucks its wings just at the right moment, and it rarely misses its target. How do these birds (and flies) know how to fly at all, and to fly and dive so well?



Figure 10.1.2: Brown pelican tucking its wings as it dives for a fish. Genetic information controls the complex, precisely timed movements involved (Image from Wikimedia Commons; File:Brown pelican (Pelecanus occidentalis occidentalis) diving.jpg; https://commons.wikimedia.org/wiki/F...is)_diving.jpg; By Charles J. Sharp, Sharp Photography; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

Over the evolutionary history of these diverse species (one an insect, the other a large bird), natural selection has preserved the genetic information that allows flies and pelicans to fly (they don't learn it); and it is just the right information needed to organize the complex movements of flight in these two very different species.

How did the "right" information, information that so perfectly organizes the movements required for flight, get into fly DNA and pelican DNA? Of course, the answer is natural selection. The wrong information has been weeded out over thousands of generations by natural selection, leaving for reproduction to future generations only the "correct" information, the information that guides with precision the complex movements required for flight in these two species. Note that this involves an information acquisition process, but it is species-wide and takes place over the thousands and thousands of generations that make up the genetic and evolutionary history of each species. The primary mechanism for acquisition of genetic information, the "correct" (i.e., adaptive) genetic information, is natural selection. Its method of transmission is heredity.

By contrast, learned information is acquired not over the genetic history of the species, but during the short lifetime of the individual animal. This is useful because some changes in the environment that may impact chances for survival and reproduction (successful adaptation) may be temporary or highly specific or even unique to the experience of only one or a few individuals within a species. Such changes are far too rapid and too specific to one or a few individuals for the relatively slow processes of natural selection to operate. For example, learning where the local water holes are would certainly be very useful to a lion during the dry season. Dependence upon genetics for this type of information isn't likely to work very well. Learning mechanisms are required.



Figure 10.1.3: Lioness drinking from small water hole on the African savannah. Information about location of water on the savannah changes too rapidly to be incorporated into genes and therefore must be learned (Image from Wikimedia Commons; File:Lioness drinking.jpg; https://commons.wikimedia.org/wiki/F...s_drinking.jpg; By James Wagner; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

In sum, the mechanisms of information acquisition are different for genetic information (the law of natural selection and the laws of heredity) and for learning (the laws of learning; see below), but both result in information acquisition. In the healthy organism this information facilitates adaptation to the environment (with some exceptions as when learning goes awry; an example would be behavioral pathologies such as drug addiction where drugs such as amphetamine, by virtue of their chemical structure, activate pleasure circuits in the brain when they shouldn't be activated, thereby reinforcing harmful drug-taking behaviors; see chapter on psychoactive drugs).

Thirdly, the mode of coding and storage is different for genetic information compared to learned information. Remember, DNA forms a molecular code for coding and storage of genetic information. Learned information is coded and stored in memory systems in brains. We don't yet know all the details of the coding and storage mechanisms in learning. Early theories proposed a molecular storage mechanism in brain proteins. This led to memory transfer experiments in the 1970's wherein an animal was taught something, like a conditioned response to a signal, and then killed, and its brain ground up in a malt mixer; the liquid brain was centrifuged to extract brain proteins or RNA molecules, which were then injected into the brains of untrained recipient animals. These animals were then tested to see if they showed the conditioned response to the signal, that is, to see if they had memory for something they had not experienced themselves. Some positive results were reported by some labs; but most labs could not get the memory transfer effect and this line of research disappeared after a few years. Current research suggests that changes in existing synapses, making them more or less responsive, and/or formation of new synapses, stores learned information in memory. We will cover the details of synaptic change in section 10.4.

Fourth, genetic information is transmitted across generations by genetic transmission (heredity). By contrast, learned information, in a few species including us, can also be transmitted across generations--not by genetic transmission, but instead by what is known as cultural transmission (tradition, imitation of the old by the younger members of the social group, books, teaching and learning, storytelling, and so forth). Cultural transmission occurs in relatively few species. We can do it; chimpanzees can do it to a limited degree; Japanese macaque (ma-cack) monkeys can do it to a limited degree. For example, one clever macaque female invented an efficient way to clean sand from wheat grains; other macaques in the group watched and learned, and the learned behavior spread throughout the group and then across generations--cultural transmission (Schofield, et al., 2018). However, rats, and most other

species, can't do cultural transmission, and certainly no other species comes close to using cultural transmission to the extent that humans do.



Figure 10.1.4: Macaque monkey feeding. Learned behaviors can be transmitted across generations by cultural transmission by humans and a limited number of other species including macaques (Image from Wikimedia Commons; File:Artis Dining Japanese macaque (6807832054).jpg; https://commons.wikimedia.org/wiki/F...807832054).jpg; By Kitty Terwolbeck; licensed under the Creative Commons Attribution 2.0 Generic license).

Cultural transmission takes a special kind of brain but most species don't have the kind of brain circuit organization required for cultural transmission (the transmission of learned information from generation to generation). Language in humans, of course, plays a crucial role in human cultural transmission, but so do technological and social inventions such as writing, the printing press, film, video, formal education, research institutions, computers, and the internet. All have magnified human cultural transmission and played an enormous role in recent human adaptation (Koenigshofer, 2011). Just as swimming can be considered a defining adaptation of fish, cultural transmission is an adaptive specialty of humans.

	Period of Acquisition	Laws of Acquisition	Mechanisms of Encoding/Storage	Transmission Mechanism
Genetic information	acquired over generations (slow information acquisition process), producing species-wide, innate adaptations	acquired by the laws of evolution, primarily by natural selection, over the evolutionary history of the species	encoded and stored in a molecular code in DNA in the genes and chromosomes of the nucleus of cells	transmitted across generations by genetic transmission (heredity, inheritance)
Learned information	acquired during an individual's lifetime (fast information acquisition process) producing rapid and individualized adaptation	acquired by the laws of learning such as the law of effect and association by occurrence of events close in time	encoded and stored in changes in neural circuits probably involving changes in synaptic connections between neurons	transmitted across generations by cultural transmission

Table 10.1.1. Comparisons between genetic and learned information organizing behavior into adaptive patterns. Adaptations, including behavioral adaptations, are not random, but highly structured. This structure requires information. Two sources of information organize behavioral adaptations--genetics and learning, nature and nurture, in interaction (table and caption by Kenneth A. Koenigshofer, PhD; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

10.1.3 Types of Biological Adaptation

At this point it is useful to note that biological adaptations can be loosely categorized into three (overlapping) general types:

1. Anatomical adaptations (structural features of the organism such as having fur, or wings, or fins, or hands, or bones, or a liver, or a large brain with a well developed cerebral cortex). These are organized by genetic information and change (evolve) only relatively slowly (although small changes, for example, in beak size and thickness of finches in the Galapagos, can occur much more rapidly; Grant & Grant, 1993).
2. Physiological adaptations (internal dynamic processes of the organism such as photosynthesis in plants, digestive processes in animals, the immune system's operations, shivering in response to cold, sweating in response to excess heat, the circulatory system's operations, etc.). These are organized by genetic information and change (evolve) only relatively slowly, although the immune system shows significant plasticity during the individual lifetime to deal with newly encountered pathogens.
3. Behavioral adaptations(the goal-directed movements of the organism and the mental processes such as thoughts, plans, emotions, perceptions, reasoning processes, imagination, etc. that underlie and control these movements). As implied above, behavioral adaptations can be organized by genetic information (genetically preprogrammed, "instinct" and reflexes; flying in flies; swimming in fish; feeding in frogs; sex drive in humans) or by learned information (acquired during the lifetime of the animal), or by a combination of both of these sources. It is important to be aware that *the abilities for learning* in any species are themselves genetically evolved. Species that can learn can do so only because learning, and the properties of the nervous system that make learning possible, evolved in those species, including humans, as a result of natural selection.



Figure **10.1.5**: Human sex drive is inborn. It can be thought of as a psychological adaptation involving intense emotions leading to reproductive behavior within the context of a pair-bond, increasing likelihood of surviving offspring. Learned cultural practices influence courtship practices and the expression of innate sex drive. (Image from Wikipedia Commons; File:Family love wiki008.jpg; https://commons.wikimedia.org/wiki/F...ve_wiki008.jpg; By Shagil Kannur; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

Behavioral adaptations can be categorized, at least conceptually, into 2 major types based on the source of the information that controls them: 1) learned adaptive patterns of behavior, often called "adaptive adjustments" by biologists (organized primarily by learned information, that is, information acquired during the lifetime of the individual animal, such as agricultural practices in humans) and 2) genetic or innate behavioral adaptations (organized by hereditary information, information stored in DNA, acquired over the evolutionary history of the species). However, many behavioral adaptations (perhaps most, in the more complex animal species especially) are a combination of both sources of information. Learning always involves genetic information to some degree because even "general" forms of learning such as conditioning incorporate and interact with information acquired by genetic evolution. Ability to learn is genetically evolved and learning processes follow laws of learning evolved by natural selection. Even conditioning mechanisms incorporate genetically encoded information that guides learning (Chiappe & McDonald, 2005; Gallistel, 1992, 2000; Koenigshofer, 2017).

Some authors refer to learned behaviors as "adjustments," rather than "adaptations." However, here we sometimes may use "learned behavioral adaptations" to emphasize that learning capacities themselves are genetically evolved, and that, generally (in healthy animals), learning and learned behaviors typically contribute to improved adaptation because of the evolved adaptive organization of pleasure and pain circuitry in the brain. For example, reward mechanisms in the brain that reinforce learned voluntary behaviors involve genetically evolved reward circuits, and punishments that inhibit voluntary behaviors involve activation of genetically evolved pain and fear circuits (reduced activity in reward circuits can also be "punishing"). In addition, behavioral change in general is a consequence of other evolved properties of brains at the synaptic level (see Section 10.4). The behavioral guidance provided by these circuits improves adaptation, increasing biological fitness (chances of survival and reproduction).

The behaviors of various animal species can be thought of as falling on a **nature-nurture continuum**, with some behaviors in some species (flies, roaches) being almost completely at the nature (innate) end of the continuum (i.e. behavior determined by genes and genetic evolution), while at the other end, the nurture end of the continuum, are behaviors which are dependent primarily on information acquired during the lifetime of the individual (i.e. learned information), such as termite fishing by a chimpanzee, or in the case of humans, how to make a flint spearhead, how to make a light bulb, do calculus, dig a well, or engage in modern agricultural practices (Koenigshofer, 2011, 2016). See figure 10.1.6 below.

10.1.4 Long-Term, Across-Generation Categorical Information vs. Short-Term Specific Information

One way to look at learning and genetics involves a general principle governing these two sources of information. Recall from Module 3.1 of this text that natural selection can only act on long-term, recurrent, across-generation environmental conditions in order to create complex adaptations. To illustrate, let's take an example from anatomy. Bones of sufficient strength to support the bodies of land animals against the downward pull of gravity evolved because the force of gravity has been present and stable over countless generations. To take a behavioral example, humans and many other animals evolved neural circuits for thirst as a motivating force causing them to seek out and consume water. However, natural selection could not have created the neural mechanisms for thirst if cellular need for water had not been consistently present over countless generations.

The general principle is this: natural selection can create genetic adaptations only to situations, adaptive problems, or other conditions of the environment that are regularly present generation after generation, because natural selection requires stable selection criteria over long periods of time to evolve complex adaptations. But in the case of behavioral/psychological adaptations, genetics can only give rather general direction to behavior because the environmental situations or conditions that led to the genetic evolution of the adaptation are themselves rather general categories of events; furthermore, these events are variable in their details but constant in their more abstract common features (Koenigshofer, 2017).

For example, the psychological adaptation, thirst, can guide and motivate the search for water and can motivate its consumption when found, but natural selection cannot code the location of water because the location of water in the environment is too variable from individual to individual and over time, and is therefore too unstable for natural selection to genetically encode the location of water in the environment. This is where learning comes in. Thirst, an innate adaptation, motivates the search for water, but once it is found, the specifics of the location and the quality of the water source must be learned and remembered. In this way, learning supplements genetic information. Learning fills in the detailed information needed to solve a particular adaptive problem or exploit a particular environmental opportunity to adaptive advantage within the specific environment of the individual animal. Genetic information is widespread or even universal throughout a species, whereas learned information is often unique, at least in its details, to the individual.

Genetic information acquired over generations of natural selection is more general and categorical (i.e. more abstract)--e.g. the feeling of thirst means "find and consume water," but doesn't provide the critical information about details such as the location of water. Thus, there arises a general principle: we can expect that whenever information acquired through evolution by natural selection is insufficient to specify a solution to an adaptive problem because more detailed and specific or particular information is needed, then learning mechanisms will evolve that are specific to the problem category (Koenigshofer, 2017).

For example, food selection is important in omnivores such as rats, coyotes, and humans (because they eat a wide variety of potential foods; unlike Koala bears that only eat eucalyptus leaves). In omnivores a specialized form of learning, called taste aversion learning, has evolved. Research shows that learned associations between taste and gastrointestinal illness are readily formed by omnivores, including us, but not in other species such as baleen whales that consume only one kind of food. Have you ever eaten some particular food and then gotten sick later and now you can't stand that food or taste? Genetic evolution equipped omnivores like us with genetic information that allows us to quickly learn taste-illness associations (but not associations between the sight of the food and illness; this is an example of a "biological (i.e. genetic) constraint" on learning). Genetic evolution biologically prepared us to learn taste-illness associations, but it could not tell us which specific tastes we should associate with illness. That detail is left to learning, specific to the experience of the individual during the individual's lifetime.

Another example is that genetic information acquired by natural selection predisposes us, as a species, to learn cause-effect relationships in the environment (see Chapter 14 on Intelligence and Cognition) but doesn't tell us which *specific* things in one's own particular environment are causally related--again, that detail is left to learning and the learning is guided by genetically evolved predispositions to search for and learn causal relations in the environment (Koenigshofer, 2017).



Figure **10.1.6**: Web construction by spiders is a complex set of movements directed by circuits in the spider nervous system constructed by information in the spider's DNA. The spider's behavior is so stereotyped that experts who study spiders can identify the species of spider from the structure of its web alone, even if the spider is not present. No learned information is required by the spider to construct its species-typical web. By contrast, human construction of structures is highly dependent upon learned information that has been culturally transmitted across generations. The human brain evolved mechanisms for learning, cultural transmission, comprehension of three-dimensional space, and the ability to visualize in imagination forms like the one depicted here. (Images from Wikipedia Commons; Source of image of spider and web: File:Spider weaving it's web.jpg; https://commons.wikimedia.org/wiki/F...it%27s_web.jpg; by Varun V Vasista; licensed under the Creative Commons Attribution-Share Alike 4.0 International license. Source of image of palace in Spain: File:Dawn Charles V Palace Alhambra Granada Andalusia Spain.jpg; https://commons.wikimedia.org/wiki/F...usia_Spain.jpg; by Jebulon; made available under the Creative Commons CC0 1.0 Universal Public Domain Dedication license).

10.1.5 Slow vs. Fast Mechanisms of Behavioral Change

Genetically organized behavioral adaptations (such as web-building in spiders), like anatomical and physiological adaptations, are organized by genetic information and therefore they change (evolve) only relatively slowly, over many generations by natural selection and other mechanisms of evolution (see Chapter 3). However, for many genes in the organism's phenotype, the gene's *expression* can be affected by multiple environmental factors. The study of this kind of gene-environment interaction is called epigenetics (see section on epigenetics in Chapter 3).

By contrast, learned behavioral adaptations ("adaptive adjustments") can change moment to moment and, as discussed above, may be transmitted (in some species) to future generations by cultural transmission to the benefit of those future generations. For example, the agricultural practices--learned behavioral adaptations or "adjustments"--upon which we have come to depend for our food supply, were invented by people who died generations ago. However, the death of those who invented these behaviors did not result in the loss of the successful behavioral adaptations upon which we depend for our food supply. Instead, these learned behavioral adaptations have been passed on, and refined, over many generations, to our current generation to its adaptive advantage--not by genetic transmission, but by cultural transmission. Our great capacity for cultural transmission is the single thing that makes us most different from other species, and accounts more than anything else for what we think of as being most characteristically human---technology, art, agriculture, governments, economies, medicine, and science. All would impossible were we, humans, not so powerfully equipped by our brain evolution for efficient cultural transmission of learned behavioral adaptations across generations. By this means, the learning of prior generations is not lost, but remains over generations for each new generation to build upon. Just like flying is a specialty of birds, or swimming is a specialty of fish, cultural transmission is the specialty of the human species.

10.1.6 Summary

In short, learning describes processes whereby information is acquired during the lifetime of an individual animal. Learned information, along with genetic information, helps organize the animal's behavior into adaptive patterns, especially in response to short-term environmental changes where specific, frequently changing details are adaptively significant and therefore must be captured by the organism and put to adaptive use. For example, learning and remembering where a temporary water hole is located

on the open savannah is essential to survival for innumerable species that live on the African plains. Because many short-term event details do not regularly recur over generations, such non-recurrent environmental details cannot drive natural selection for genetically evolved instinctual or reflexive adaptations. Instead adaptation to such novel idiosyncratic event details favors the evolution of learning mechanisms (Koenigshofer, 2017). Usually this learned information supplements hereditary (genetic) sources of information in the organization of successful behavioral adaptations--behavioral solutions to the problems of survival and reproduction. As noted above, behavior is one way that organisms (at least animals) adapt. Behavior becomes organized into adaptive patterns by information (i.e. behavior is not random, but guided by laws of nature which govern chances of survival and reproduction--laws of evolution and the laws of learning). The information that organizes behavior comes from the genes after having been perfected by eons of genetic evolution by natural selection. That organizing information can also come from learning by an individual animal during its lifetime (laws which govern learning are also organized to enhance survival and reproduction--see sections below).

Adaptive behaviors can be transmitted to future generations. If the behavior is organized by information in the genes, then that behavior can be transmitted by genetic transmission (heredity). If the information for a particular behavior comes from learning (and is stored in memory), then the behavior can be transmitted to future generations, not by genetic transmission, but instead by cultural transmission (in some species, as noted above). Thus behavioral adaptations in species capable of cultural transmission can undergo not only genetic evolution (true for behaviors organized by genetic information contained in DNA, and also true of anatomical and physiological adaptations) but also cultural evolution (examples of cultural evolution are the development of human technology, modern medical practices, agricultural practices, economies, science, governmental systems and so on, leading over time to generally improved human adaptation). Cultural transmission and the resulting cultural evolution gives our species great survival advantage. Cultural transmission accounts for the success of the human species more than any other single factor.

10.1.7 References

Chiappe, D., and MacDonald, K. (2005). The evolution of domain-general mechanisms in intelligence and learning. *Journal of General Psychology*, *132* (1), 5-40.

Gallistel, CR. (1992). Classical conditioning as an adaptive specialization: A computational model. In D.L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 35-67). San Diego: Academic Press.

Gallistel, C. R. (2000). The replacement of general-purpose learning models with adaptively specialized learning modules. *The Cognitive Neurosciences*, *2*, 1179-1191.

Garcia, J., and Koelling R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science, 4*, 123-124.

Grant, B. R., & Grant, P. R. (1993). Evolution of Darwin's finches caused by a rare climatic event. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *251* (1331), 111-117.

Kandel, E. (1976). *Cellular Basis of Behavior*. San Francisco. W.H. Freeman and Company.

Koenigshofer, K.A. (2011). *Mind Design: The Adaptive Organization of Human Nature, Minds, and Behavior*. Pearson Education. Boston.

Koenigshofer, K.A. (2016). *Mind Design: The Adaptive Organization of Human Nature, Minds, and Behavior*. Revised Edition. Amazon e-book.

Koenigshofer, K. A. (2017). General Intelligence: Adaptation to Evolutionarily Familiar Abstract Relational Invariants, Not to Environmental or Evolutionary Novelty. *The Journal of Mind and Behavior*, 119-153.

Seligman, M. (1971). Phobias and preparedness. *Behavior Therapy*, 2, 307–321.

Tolman, E. C., and Brunswik, E. (1935). The organism and the causal texture of the environment. *Psychological review*, *42* (1), 43.

10.1.8 Attributions

Section 10.1, "Learning, Genes, and Adaptation" is original material written by Kenneth A. Koenigshofer, PhD. and is licensed under CC BY 4.0.

Images from Wikimedia Commons.

This page titled 10.1: Learning, Genes, and Adaptation is shared under a mixed license and was authored, remixed, and/or curated by Kenenth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

10.2: Types of Learning and Biological Adaptation

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

10.2.1 Learning Objectives

1. Describe the various types of learning and how each contributes to adaptation
2. Discuss in what ways both classical and operant conditioning involve the learning of predictive relations between different types of events in the world
3. Discuss the learning of predictive relationships among events in the environment and the navigation of "the causal texture of the world"
4. Describe specialized forms of learning and how each contributes to solution of a particular adaptive problem (problem domain) in a particular species
5. Explain classical (Pavlovian) conditioning and instrumental (operant) conditioning
6. Describe similarities and differences between the two types of conditioning
7. Explain the concept of adaptively specialized learning and give examples
8. Discuss the four aspects of observational learning according to Social Learning Theory.
9. Describe how habituation conserves resources
10. Describe research that suggests cognitive learning in animals

Overview

Learning brings to mind school, memorization, tests, and study. But these associations are related to a single type of learning only, human verbal learning. Actually, learning, as you may already know, is a much broader phenomenon. In this module, we examine learning in its various forms and the ways in which it advances adaptation to the environment. First of all, learning of various sorts occurs in a wide range of animal species. For example, experiments have shown that even an insect species, the honey bee, can learn (for example, where a food source is located). Furthermore, across species, a wide range of behaviors can be acquired or modified by learning, giving behavior in many species a great deal of plasticity. In addition, we now know that learning is not a single unitary capacity, but instead there are many different, but related, processes, ranging from imprinting to cognitive learning. As discussed previously, behavior is a type of adaptation. It is not random, but quite orderly, and as such it is dependent upon information for its organization. Remember that the information that organizes behavioral adaptations can come from genes, from learning, or from a combination of both. In this module, preliminary to discussion of the neural mechanisms of learning, we examine several types of learning and how they interact with genetic information to serve adaptation. Imprinting is a clear example of learning that is biologically (genetically) "prepared" (controlled and facilitated by innate factors as a consequence of evolution). Both forms of conditioning, classical and operant, involve learning the predictive relationships between events in the environment and both involve innate genetic facilitation (biological preparedness) and biological constraints just like imprinting does. However, in conditioning, genetic influences on learning involve much more abstract and relational features of the world than in more clearly specialized forms of learning such as imprinting and learned taste aversion. In every kind of learning, learning fills in informational details that are too variable, short-term, and individually experienced to be captured by natural selection and thereby encoded into genes. By contrast, more abstract and general features, common to particular learning situations or problem types across generations, *are* captured by natural selection, genetically encoded, and provide innate, genetic information about the problem type (problem domain) that guides and facilitates the learning.

10.2.2 Habituation and Adaptation

Habituation is a simple form of learning that produces a decrease in response to a repeated stimulus that has no adaptive significance. In other words, as an adaptively unimportant stimulus is repeatedly presented to an animal, it will gradually cease responding to the stimulus as it learns that the stimulus holds no information that might impact its biological fitness (survival and reproduction). Prairie dogs typically sound an alarm call when threatened by a predator, but they become habituated to the sound of

human footsteps when no harm is associated with this sound; therefore, they no longer respond to the sound of human footsteps with an alarm call, nor do they run and hide, but instead they spend their time and energy on more productive behaviors. Habituation is a form of non-associative learning given that the stimulus is not associated with any punishment or reward.

Habituation is highly adaptive. It increases the efficient use of an animal's limited resources. Imagine birds feeding in a field along a highway. The first time a bird feeds in this situation, it will probably fly away when a car speeds past. However, over time, as more cars pass by and nothing harmful or otherwise adaptively significant happens to the bird, it learns to ignore the passing traffic. Instead of flying away, it simply continues to feed, saving time, metabolic energy, and the limited processing capacity of its brain. Habituation conserves these important biological resources of the organism, resources that would otherwise be wasted by responding to adaptively irrelevant stimuli. Without habituation, behavior would lose biological efficiency, decreasing biological fitness in the struggle for survival and reproduction.

Habituation occurs when your brain has already extracted all the adaptive information that a stimulus or an event holds. Once habituation has occurred, as long as the habituated stimulus or event is unchanging, it makes good adaptive sense for the animal to ignore it. Habituation has done its job by conserving the animal's biological resources. But habituation is only half the story. What happens when, after habituation, a change in the stimulus or situation occurs? Stimulus change might signal that something important has happened. If so, the animal must reengage the processing capacity of its brain, its attentional resources, to evaluate whether the stimulus change might signal something that is adaptively important that may require a behavioral response. This recovery of responding to a habituated stimulus is called dishabituation (the undoing of habituation as the organism now responds to the stimulus situation it had previously stopped responding to). For example, suppose there is a loud backfire from one of the passing cars. The birds fly off, at least temporarily.



Figure **10.2.1**: Habituation is adaptive. Birds feeding along a road eventually habituate to passing traffic preventing them from wasting time and energy by flying away (escape behavior) unnecessarily in the absence of real threat, thereby increasing time and energy available for feeding (Image from Wikimedia Commons; File:Birds feeding on newly ploughed land with Ballymagreehan Hill in the background - geograph.org.uk - 2202208.jpg; https://commons.wikimedia.org/wiki/F..._-_2202208.jpg; by Eric Jones; licensed under the Creative Commons Attribution-Share Alike 2.0 Generic license.).

Another example comes from a former student of mine who had been in the Navy submarine corps. He said that when he was assigned to his first sub, he couldn't sleep for several nights because of the loud metallic noise coming from the engines. However, after several nights of hearing the racket, he finally "got used to it", he habituated to a repeated stimulus that had no adaptive

significance, and he was then able to sleep soundly. Can you guess what would now wake him up? It was the *absence* of the noise, when the engines were shut down, that would immediately awaken him. Maybe the sub is under attack, or the engines are damaged and they are sinking, or perhaps the sub has docked in port and there could be some adaptive consequences for the sailor like a potential romantic opportunity on shore, or other adaptive resources such as preferred foods. Habituation is undone by stimulus change. Habituation and dishabituation are constantly at work efficiently conserving or deploying the organism's biological resources (metabolic energy, time, and brain processing capacity) as required by environmental circumstances. The net effect is efficient and adaptive distribution of the organism's biological resources, thereby increasing biological fitness.



Figure 10.2.2: After habituation to the loud sound of their submarine's engines, sailors sleep through the noise, but are awakened by the silence if the engines stop. Stimulus change reverses habituation. This is called dishabituation and is highly adaptive; see text (Image from Wikimedia Commons; File:Vladivostok Submarine S-56 Forward torpedo room P8050522 2475.jpg; https://commons.wikimedia.org/wiki/F...50522_2475.jpg; by Alexxx1979; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

One lesson is clear: that stimulus change is very potent in causing the brain to become alert and responsive. This implies that the brain must hold an ongoing representation or neural model of a current situation and must respond to any mismatch between that ongoing memory and the current stimulus situation (see chapter on Intelligence, Cognition, and Language). This is a very adaptive property of the brain's functioning. For example, imagine you are studying psychology late at night. Your window is open and a breeze gently bangs the blinds against the window frame as you read. You will probably habituate to the sound, and thus pay no attention to it, to the point where you don't even hear it as you continue to study. But then, if the lights in your house suddenly go out, you probably will dishabituate and pay attention to every little sound, including the sounds coming from the blinds. Your brain is programmed to respond to stimulus change, to novelty, because in stimulus change there may be information important for survival and reproduction, demanding an adaptive response. Could those sounds in the dark be an intruder who might harm you or your family (who, by the way, carry a portion of your genes; see discussion of kin selection in the chapter on evolution and genetics)?

If there is no danger or other adaptively significant event associated with the stimulus change, then habituation will again occur as the animal or human learns that the stimulus change can be safely ignored. As noted above, habituation prevents the animal or human from wasting time, energy, and processing capacity on things that have no adaptive importance, thereby increasing behavioral efficiency. Habituation is an elegant mechanism for conserving biological resources and thus contributes greatly to adaptation and biological fitness (survival and reproduction). Interestingly, habituation mechanisms appear to be highly conserved across a wide range of species, emphasizing the importance of habituation for survival (see Schmid et al., 2010).

10.2.3 Conditioning and Biological Adaptation

Basic principles of learning are always operating and always influencing human and animal behavior. This section continues by discussing two fundamental forms of associative learning: classical (Pavlovian) and operant (instrumental) conditioning. Through them, we respectively learn to associate 1) stimuli in the environment, or 2) our own behaviors, with adaptively significant events,

such as rewards and punishments or other stimuli. The two types of learning have been intensively studied because they have powerful effects on behavior, and because they provide methods that allow scientists to rigorously analyze learning processes in detail, an endeavor important to biological psychologists searching for the physical basis of learning and memory in the brain. This module describes how both classical and operant conditioning involve the learning of predictive relationships between events (if event A occurs, then event B is likely to follow), and how this contributes to adaptation. The module concludes by discussion of adaptively specialized forms of learning and observational learning, which are forms of learning that are largely distinct from classical and operant conditioning.

10.2.4 Classical Conditioning

Many people are familiar with the classic study of “Pavlov’s dog,” but rarely do they understand the significance of Pavlov's discovery. In fact, Pavlov’s work helps explain why some people get anxious just looking at a crowded bus, why the sound of a morning alarm is so hated, and even why we swear off certain foods we’ve only tried once. Classical (or Pavlovian) conditioning is one of the fundamental ways we learn about the world, specifically its predictive relationships between events. This involves learning *what leads to what* in an organism's environment. This is extremely valuable adaptive information that animals and humans appear to incorporate into brain-mediated cognitive models of how the world works--information which allows prediction and therefore improved organization of behavior into adaptive patterns (a topic to be discussed more in the chapter on Intelligence, Cognition, and Language). But classical conditioning is far more than just a theory of learning; it is also arguably a theory of identity. Your favorite music, clothes, even political candidate, might all be a result of the same process that makes a dog drool at the sound of bell.

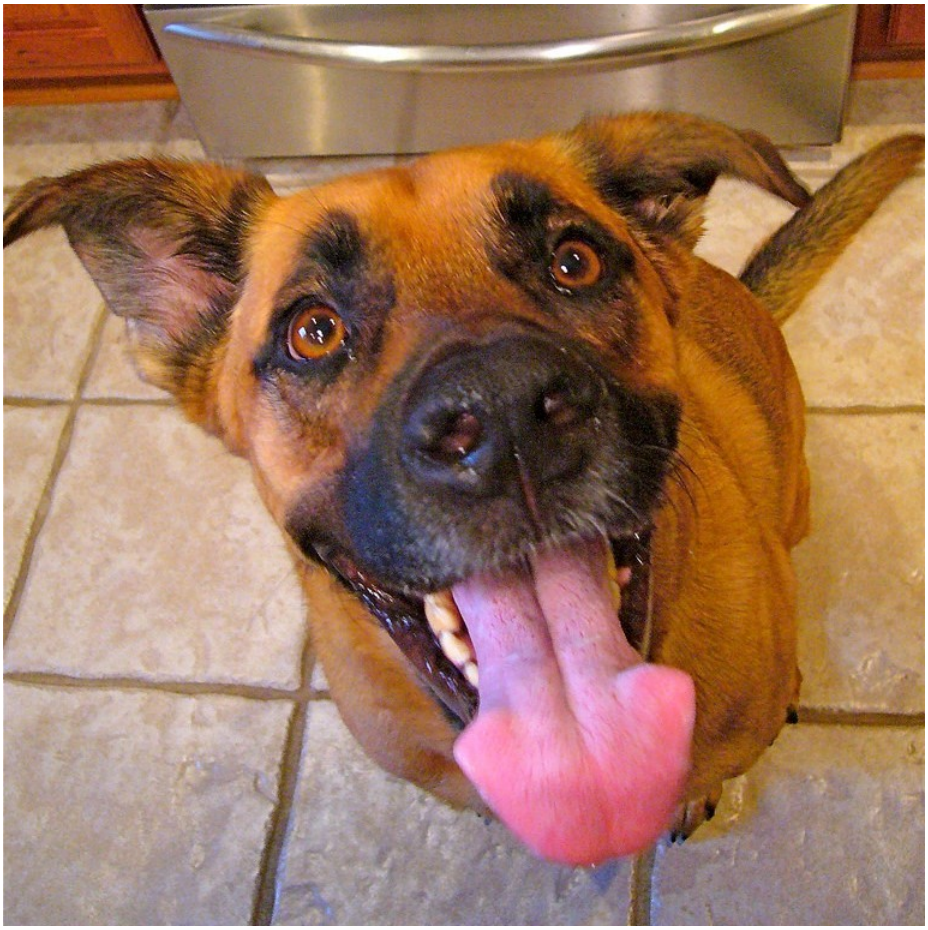


Figure **10.2.3**: Does your dog learn to beg for food because you reinforce her by feeding her from the table? Classical conditioning signals to your dog when food reward may be available and operant conditioning (see below) occurs when begging is reinforced by a food reward [Image: David Mease, https://nobaproject.com/modules/cond...g-and-learning, CC BY-NC 2.0]

In his famous experiment, Pavlov rang a bell and then gave a dog some food. After repeating this pairing multiple times, the dog eventually treated the bell as a signal for food, and began salivating in anticipation of the treat. This kind of result has been reproduced in the lab using a wide range of signals (e.g., tones, light, tastes, settings) paired with many different events besides food (e.g., drugs, shocks, illness; see below).

We now believe that this same learning process, **classical conditioning**, is engaged, for example, when humans associate a drug they've taken with the environment in which they've taken it; when they associate a stimulus (e.g., a symbol for vacation, like a big beach towel) with an emotional event (like a burst of happiness); and when a cat associates the sound of an electric can opener with feeding time. Classical conditioning is strongest if the conditioned stimulus (CS) and unconditioned stimulus (US) are intense or salient. It is also best if the CS and US are relatively new and the organism hasn't been frequently exposed to them before. And it is especially strong if the organism's biology (its genetic evolution) has prepared it to associate a particular CS and US. For example, rats, coyotes, and humans are naturally inclined by natural selection and the resulting evolution of their brain circuitry to associate an illness with a flavor, rather than with a light or tone. Although classical conditioning may seem "old" or "too simple" a theory, it is still widely studied today because it is a straightforward test of associative learning that can be used to study other, more complex behaviors, and biological psychologists can use it to study how at least some forms of learning occur in the brain.

10.2.5 Conditioning Involves Learning Predictive Relations

Pavlov was studying the salivation reflex, reflexive drooling in response to food placed in the mouth. A **reflex** is an innate, adaptive, genetically built-in stimulus-response (S-R) relationship; in this case, the stimulus is food in the mouth, the unconditioned or **unconditional stimulus** (a stimulus unconditional, i.e. not dependent upon prior learning; **US**), and the **unconditional response** (unconditional, not dependent, upon prior learning; **UR**) is salivation to food in the mouth which lubricates the food and starts to break it down, facilitating mastication, swallowing, and digestion--the adaptive function of the salivation reflex (US--UR).

Recall that Pavlov found that if he rang a bell just before feeding his dogs, the dogs came to associate the sound of the bell with the coming presentation of food. Thus, after this classical conditioning had occurred, the bell alone (the conditioned or **conditional stimulus**--conditional upon prior learning; **CS**) caused the dog to salivate, before the presentation of food. The conditional stimulus (Pavlov's original terminology which was mistranslated from Russian as "conditioned") is a signal that has no importance to the organism until it is paired with something that does have adaptive significance, in this case, food. Reliable pairing of CS and US close together in time (**temporal contiguity**) is important to the processes of classical conditioning. However, temporal contiguity alone is not sufficient for classical conditioning. The pairing of stimuli must be reliable so that a **predictive relationship** is maintained between CS and US. Predictiveness between CS and US determines whether or not an association is formed. If the CS does not predict occurrence of the US, no conditioning occurs (Gallistel, et al., 1991).

Evidence that classical conditioning involves the **learning of predictive relations** between stimuli comes from a phenomenon known as **blocking** (see Kamin, 1969). In blocking, if a CS already predicts the US, and if a new CS is added, no association is formed between the *new* CS and the US, no matter how many times or how closely in time they are paired together. The reason? The *original* CS already predicts the US, so no new or additional predictive information about occurrence of the US is added by the new CS. Thus, the new CS is ignored as a source of predictive information and is therefore "blocked" from becoming a signal for the US, and no conditioning between the new CS and US occurs. Furthermore, other studies show that if the predictive relation between CS and US is "diluted" or weakened by any presentations of the US without the CS preceding it, conditioning between CS and US is impeded. This is because if the US can occur without the CS preceding the US, then this fact reduces the predictive power of the CS and violates the contingency between CS and US, impairing conditioning. This shows that a **predictive contingency** between CS and US is what is learned in classical conditioning and that contingency is the necessary condition for conditioning to take place (Rescorla, 1966, 1968).

Both of these important research findings emphasize that conditioning is about *learning to predict what leads to what* in the environment--**classical conditioning** is the **learning of predictive relations between stimuli**, leading to the learned emission of responses that **prepare for** the coming **US**. Blocking and other related effects indicate that the learning process tends to take in the most valid **predictors of adaptively significant events** and ignore the less useful ones. This is an exceedingly important process that allows the animal or human to learn important adaptive information about its specific environment.

For example, as a result of classical conditioning, Pavlov's dogs learned to anticipate, to predict, one event (US) signaled by earlier occurrence of another (CS). In a sense, the dogs use the bell as a signal to predict that food is on its way; therefore they salivate

when the bell is rung, because they now expect that food is coming next. After conditioning, when Pavlov presented the bell, the dogs salivated to the bell alone (**conditional response, CR**), whereas before conditioning they did not. Thus, we have a change in behavior as a result of experience--learning. The animal has acquired information about its environment--i.e. predictive information. The world is full of predictive and causal relations. If the organism is to effectively organize its behavior to maximize adaptation, it must be able to learn these relations and put that information to adaptive use. Classical conditioning is a mechanism for learning these **predictive (sometimes causal) relationships** among things in its specific environment.

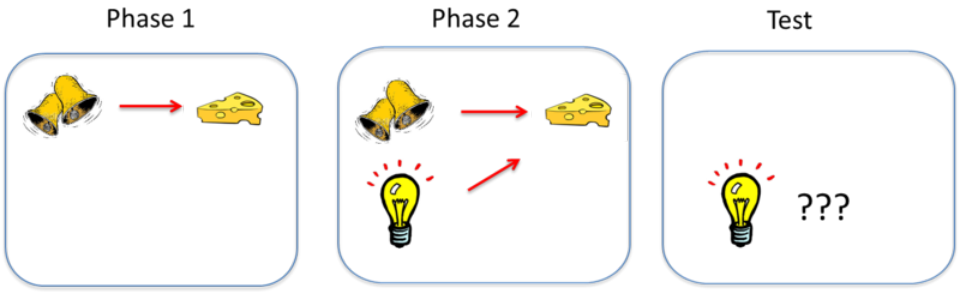


Figure 10.2.4: Illustration of blocking. Phase 1: Conditioning where bell is CS, food is US. Phase 2: after Phase 1 conditioning has been completed a second CS, light, is added. Test: When light CS is tested, no CR occurs to it; no conditioning to the light CS occurs. Reason: Bell CS already predicts food; therefore light does not add any additional predictive information so no conditioning to light CS occurs. Conclusion: Classical conditioning involves learning of predictive relations between stimuli. An animal first learns to associate one CS—call it stimulus A—with a US. In the illustration above, the sound of a bell (stimulus A) is paired with the presentation of food. Once this association is learned, in a second phase, a second stimulus—stimulus B—is presented alongside stimulus A, such that the two stimuli are paired with the US together. In the illustration, a light is added and turned on at the same time the bell is rung. However, because the animal has already learned the association between stimulus A (the bell) and the food, the animal doesn't learn an association between stimulus B (the light) and the food. That is, the conditioned response only occurs during the presentation of stimulus A, because the earlier conditioning of A "blocks" the conditioning of B when B is added to A. (Image from M. Boulton, NOBA, https://nobaproject.com/modules/cond...g-and-learning; courtesy of Bernard W. Balleine).

Classical conditioning is anticipatory, and preparatory for future appearance of the US. A classical CS (e.g., the bell) does not merely elicit a simple, unitary reflex. Pavlov emphasized salivation because that was the only response he measured. But his bell almost certainly elicited a whole *system* of responses that functioned to get the organism ready for the upcoming US (food). For example, in addition to salivation, CSs (such as the bell) that signal that food is coming also elicit the secretion of gastric acid, pancreatic enzymes, and insulin (which gets blood glucose into cells). All of these responses anticipate coming food and prepare the body for efficient digestion, improving adaptation.

Classical conditioning is also involved in other aspects of eating. Flavors associated with certain nutrients (such as sugar or fat) can become preferred without arousing any awareness of the pairing (an example of implicit or unconscious learning). For example, protein is a US that your body automatically craves more of once you start to consume it (UR): since proteins are highly concentrated in meat, the flavor of meat becomes a CS (or cue, that proteins are on the way), which perpetuates the cycle of craving for yet more meat (this automatic bodily reaction now a CR).

In a general way, classical conditioning occurs whenever neutral stimuli are associated with adaptively significant events. Classical conditioning is of great adaptive importance in a wide range of circumstances across species. Significantly, Gallistel (1992) points out that classical conditioning permits an animal to map what Tolman and Brunswik (1935) called the "causal texture of the environment." Understanding *what causes what* in the environment is highly adaptive because it allows successful, accurate prediction about events and makes effective manipulation of the environment possible. In a way, in classical conditioning, the animal learns an **if-then contingency** or **if-then predictive relationship** between two events in its environment. If event A occurs, then event B is likely to follow. If bell rings, then food is likely coming next. Learning *what leads to what* in the world permits prediction and therefore much more effective organization of behavioral adaptation than would otherwise be possible. In the evolution of animal brains, we can speculate that animals who had the capacity in their brains for classical conditioning would certainly have had an adaptive advantage, and thus a selective (as in natural selection) advantage, over other members of their

species which lacked that ability or possessed it to a lesser degree (Koenigshofer, 2011, 2016). If you can predict that something is coming, then you can better prepare for it and increase the chances of an adaptive outcome for you (and your genes). Classical conditioning has apparently been conserved during evolution given that classical conditioning is found in an enormous variety of animals from humans to sea slugs.

Modern studies of classical conditioning use a very wide range of CSs and USs and measure a wide range of conditioned responses including emotional responses. If an experimenter sounds a tone just before applying a mild shock to a rat's feet, the tone will elicit fear or anxiety after one or two pairings. Similar fear conditioning plays a role in creating many anxiety disorders in humans, such as phobias and panic disorders, where people associate cues (such as closed spaces, or a shopping mall) with panic or other emotional trauma (Mineka & Zinbarg, 2006). Here, rather than a response like salivation, the CS triggers an emotion. Have you experienced conditioned emotional responses to formerly emotionally neutral stimuli? How about the emotional response you might have to a particular song, or a particular place, that once was your and your ex-partner's favorite song or the place you would go to meet one another? Or after you break up with someone, you seem to see their car (or cars like theirs) everywhere and you have a brief moment of anticipation that you might see them. Classical conditioning plays a large role in our emotional lives, and in the emotional lives of other animals as well.



Figure **10.2.5**: Intense emotions can be classically conditioned to originally neutral stimuli such as places or songs associated with a special person. (Image from Wikimedia Commons; File:A couple looking at the sea.jpg; https://commons.wikimedia.org/wiki/F...at_the_sea.jpg; by Joydip dutt; https://commons.wikimedia.org/wiki/F...at_the_sea.jpg; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

Where classical conditioning takes place in the nervous systems varies with the nature of the stimuli involved. For example, clearly an auditory CS such as a bell will involve auditory pathways including the auditory system's medial geniculate nucleus of the thalamus (Fanselow & Poulos, 2005) and auditory cortex in the temporal lobe, while a visual CS will involve visual pathways including the visual system's lateral geniculate nucleus of the thalamus and visual areas of cortex. A US such as food will involve taste pathways, whereas presentation of a painful shock US will involve pain and fear pathways. Researchers have identified a number of brain areas that become active during fear conditioning. In response to a painful shock during conditioning, there is increased neural activity in the **red nucleus**, **amygdala**, **dorsal striatum**, the **brainstem**, the **insula**, and parts of the **anterior cingulate cortices**, whereas **anticipation** of a shock that was not delivered increased neural activity in only the **red nucleus**, the **anterior insular** and **dorsal anterior cingulate cortices** (Linnman, et al., 2011). Other biological psychologists found that the **cerebellum** has a special role in simpler forms such as the conditioning of the eye blink reflex in rabbits (Thompson & Steinmetz, 2009), whereas more complex conditioning involves the **hippocampus** and **hippocampal-cerebellar interactions** (Schmajuk & DiCarlo, 1991). Research by scientists such as Eric Kandell (1976) on the neural mechanisms of learning suggests that at the level

of the synapse the mechanisms for learning may be very similar for all types of learning, across species, involving changes in synaptic conductance. We will consider mechanisms of learning at the synaptic level later in this chapter.

Pavlov discovered a number of phenomena of classical conditioning. He found that once his dogs were classically conditioned to one bell (one CS), they would also respond (produce a CR, salivation to a bell in this case) to other similar stimuli (other bells). You may recall from Introductory Psychology that this is called **stimulus generalization**. The more similar a new stimulus was to the original CS, the stronger the CR to the new stimulus. This is called **stimulus generalization gradient**. For example, if you get bitten by a German Shepherd dog, you will probably be a little afraid of all dogs (stimulus generalization) and you will be *more* afraid of all big dogs and *even more* afraid of all German Shepherd dogs (stimulus generalization gradient). Clearly, stimulus generalization is very adaptive, because it **permits you to make predictions** not only about the original stimulus but also about **similar stimuli--stimuli of the same category**. Inferences and predictions based on generalization from one member of a category to other members of the same category is one prominent feature of cognition in intelligent creatures, including humans (see chapter on Intelligence, Cognition, and Language).

Pavlov also discovered **extinction**. Recall from Introductory Psychology that if the CS is not followed by the US at least some of the time, then eventually the CR will no longer occur in response to the presentation of the CS. When the animal stops producing the CR in response to the CS, we say that extinction of the CR has occurred. After conditioning of the salivation response, when the dog salivates to the bell, it is really salivating in *anticipation* of coming food. Even though the food may not now occur, the dog will still salivate to the bell for awhile, but eventually it learns to stop responding to the bell (extinction of the conditioned response) once the bell no longer reliably predicts food. In extinction, the animal hasn't forgotten its previous experience of a predictive relationship between stimuli, it simply learns that the predictive relation no longer holds, so it uses this new information and stops responding to the old contingency between previously conditioned events. This is demonstrated by the fact that a single pairing of the old predictive contingency between CS and US is often sufficient to reinstate the extinguished response. In conditioning the dog learned the rule that bell means coming food; in extinction the dog learns that the rule has changed and bell no longer predicts food, so naturally it stops salivating to the bell--that's extinction of the CR.

This makes good adaptive sense. Remember that in classical conditioning, the animal is learning a predictive relationship between two events in its environment (i.e. If CS occurs, then US is likely to follow, so the animal responds to the CS in anticipation of the expected US). However, if in fact the predictive relationship between those events in the environment no longer exists (that is, occurrence of CS no longer reliably predicts the future occurrence of the US), then it makes good (adaptive) sense that the animal will no longer respond to the CS since the CS no longer signals that the US will follow.

In fact, what is happening during conditioning and extinction is that the animal is **tracking (or adjusting its mental model of) the predictive relationships** among events in its world as those relationships dynamically change. Conditioning is really part of an animal's neural representation of its environment (not necessarily explicit or conscious), including the things in that environment and the relationships among them. Any individual instance of **classical conditioning is just a slice in time** in a continuous; ongoing, dynamic process, within the animal's brain, of **modeling or representing the world and its changing contingencies**--contingencies or predictive relations between stimuli which the animal must successfully "navigate" in order to successfully adapt. Extinction, like acquisition of conditioned responses, is just one component in the animal's ability to **track the predictive relationships** between events in its environment and then to put that information to work guiding behavior. Similar dynamics apply to the second major type of conditioning, operant (or instrumental) conditioning.

10.2.6 Operant Conditioning

Operant conditioning, as you probably recall from your course in Introductory Psychology, is a second kind of conditioning in which the organism actively operates on the environment. Like classical conditioning, it also involves learning a predictive relationship between events, but the events are different than in classical conditioning. In operant or instrumental conditioning, the animal or human learns a **predictive relationship between its own voluntary behavior and the outcome or effect of that behavior**. In the best-known example, a rat in a laboratory learns to press a lever or a bird pecks a display in a cage (called a “Skinner box”) to receive food.

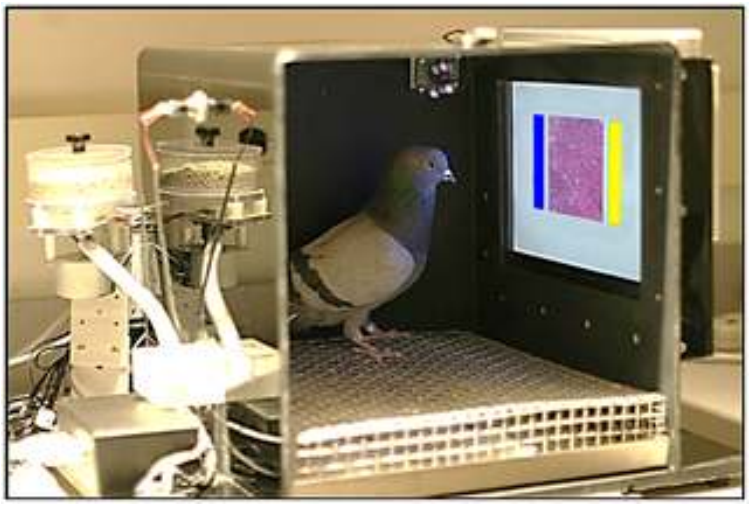


Figure 10.2.6: An operant conditioning chamber. Pecking on the correct color will deliver food reinforcement to the pigeon (Image from Wikimedia Commons; File:The pigeons' training environment.png; https://commons.wikimedia.org/wiki/F...nvironment.png; Creative Commons CC0 1.0 Universal Public Domain Dedication).

Operant conditioning research studies how the effects of a behavior influence the probability that it will occur again. For example, the effects of the rat's lever-pressing behavior (i.e., receiving a food pellet) influences the probability that it will keep pressing the lever. According to Thorndike's **law of effect**, when a behavior has a positive (satisfying) effect or consequence, it is likely to be repeated in the future. However, when a behavior has a negative (painful/unpleasant) consequence, it is *less* likely to be repeated in the future. In other words, the **effect of a response, determines its future probability**. Effects that increase the frequency of behaviors are referred to as **reinforcers**, and effects that decrease their frequency are referred to as **punishers**.

In general, operant conditioning involves an animal or human **tracking the reinforcement contingencies** or dependencies in its environment and exploiting them to its advantage. Clearly, operant conditioning is highly adaptive. Operant conditioning shapes the voluntary behavior of the organism to maximize reinforcement and minimize punishment, just as we would expect from the law of effect. The law of effect in turn depends upon **reward circuits** in the **mesolimbic system** (see Chapter on Psychoactive Drugs) and on circuits for **pain and the emotional response to pain** which are distributed in many regions of the brain including the **somatosensory cortex, insula, amygdala, anterior cingulate cortex, and the prefrontal cortex** (see chapter on Sensory processes). **Reinforcers activate mesolimbic pleasure circuitry** in the brain or reduce activity in pain circuits. This gives the organism feedback about its actions. Reinforcers and activation of pleasure circuits tend to be **associated with enhanced adaptation** (food, a reinforcer to an animal deprived of food, enhances chances of survival). **Punishers activate circuitry for physical or emotional pain** and tend to be **associated with reduced adaptation and biological fitness** (i.e. reduced chances of survival and reproduction; physical pain is associated with potential tissue damage, while emotional pain is often associated with loss of things or persons upon which one depends or highly values, including romantic partners, financial or social status, etc.). Voluntary behaviors by the animal which lead to positive consequences (and increased pleasure; reinforcement) for the animal tend to be repeated (increase in probability in the future). Voluntary behaviors which the animal produces that lead to nothing or lead to negative outcomes for the animal (a reduction in pleasure or the occurrence of pain) tend not to be repeated, but avoided by the animal in the future. In this way, by the law of effect, in psychologically healthy individuals, voluntary behavior becomes molded into more and more adaptive patterns so that animals and humans spend most of their time engaged in activities that improve adaptation and avoid activities that reduce it. By this means, learned voluntary behavior tends to serve successful behavioral adaptation.



Figure **10.2.7**: Receiving a reward can condition you toward certain behaviors. For example, when you were a child, your mother may have offered you this deal: "Don't make a fuss when we're in the supermarket and you'll get a candy treat on the way out." [Image: Oliver Hammond, https://nobaproject.com/modules/cond...g-and-learning; CC BY-NC-SA 2.0]

This is a wonderful mechanism for assuring the adaptive organization of behavior in species which have behavioral capacities beyond rigidly genetically programmed behavior. In species such as fish, reptiles, amphibians, and many invertebrate species, the larger portion of behavioral adaptation is organized by information in the genes, honed over millions of years of evolution by natural selection. These reflexes and "instincts" generally do not rely very much, if at all, upon information acquired during the lifetime of the individual animal (i.e. learning), but instead upon information acquired over the evolutionary history of the species and stored in DNA.

Learned voluntary behavior is flexible but must also be directed into adaptive patterns by some principle, and that principle is the law of effect. The law of effect allows behavioral flexibility but also provides a mechanism for assuring that the animal (or human) learns adaptive behavior, behavior good for it and its genes, most of the time. "Voluntary" behaviors are not rigidly pre-formed by genetic information, but are organized in a more general way which allows for their modification by information gathered from the animal's current environment (experienced day to day and even moment to moment). This information gathering, moment to moment, continuously modifying "voluntary" behavior into more and more adaptive patterns is the essence of operant conditioning and its evolutionary significance. Without the law of effect, or some similar principle provided by natural selection, voluntary behavior would be chaotic, without direction, and could not be an instrument for adaptation to the environment. Behavior would degenerate into maladaptive form leading to the rapid demise of such creatures with brains so ill-equipped for survival. The law of effect, built into the circuitry of human and animal brains by eons of natural selection, permits behavioral flexibility with an overall adaptive direction--an exquisite evolved mechanism for regulation of behavior that is not under rigid genetic control.

Like classical conditioning, operant conditioning is part of the way an animal forms a **mental model or neural representation** of its environment and the predictive relationships among events in that environment, but, as noted above, the events are different than in the case of classical conditioning. In classical conditioning, it is the predictive relationships between two stimulus events that is learned. In operant conditioning, it is the predictive relationships between a voluntary response and its outcome or consequence that is learned. Prediction allows preparation for future, even if the future is moments away, and preparation for what is coming next improves chances of survival and reproduction (for example, classical conditioning of sexual reflexes gives a reproductive

advantage in male chimpanzees competing for mating opportunities; operant conditioning of courtship behaviors in humans can improve reproductive success).

When we examine operant conditioning, we see that it bears some similarities to evolution by natural selection. In operant conditioning, voluntary responses by an animal that are successful usually are reproduced (repeated) and those that are unsuccessful (not reinforced) get weeded out (they are not repeated). In evolution by natural selection, genetic alternatives that are successful (lead to better adaptation) get reproduced (and appear in future generations) and those that are unsuccessful (do not lead to improved adaptation or are maladaptive) get weeded out (they are not replicated in future generations). In each case, a selection mechanism (natural selection or the law of effect) preserves some alternatives (genetic or behavioral, respectively) into the future, while eliminating others. In both cases, the result is improved adaptation.

10.2.7 Instrumental Responses Come Under Stimulus Control

As you know, the classic operant response in the laboratory is lever-pressing in rats, reinforced by food. However, things can be arranged so that lever-pressing only produces pellets when a particular stimulus is present. For example, lever-pressing can be reinforced only when a light in the Skinner box is turned on; when the light is off, no food is released from lever-pressing. The rat soon learns to discriminate between the light-on and light-off conditions, and presses the lever only in the presence of the light (responses in light-off are extinguished). In everyday life, think about waiting in the turn lane at a traffic light. Although you know that green means go, only when you have the green *arrow* do you turn. In this regard, the operant behavior is now said to be under **stimulus control**. And, as is the case with the traffic light, in the real world, stimulus control is probably the rule. We constantly monitor the environment for signals that tell us that a certain voluntary behavior is now called for, when at other times it is not. Will approaching someone to ask for a date be successful? We look for signs telling us whether the voluntary response of asking for a date is likely or unlikely to lead to reinforcement. If positive signs are not present, we may look for a better opportunity.

The stimulus controlling the operant response is called a **discriminative stimulus**. The stimulus can "set the occasion" for the operant response: It sets the occasion for the response-reinforcer relationship. For example, a person who is reinforced for drinking alcohol or eating excessively learns these behaviors in the presence of certain stimuli—a pub, a set of friends, a restaurant, or possibly the couch in front of the TV. These stimuli can be associated with the reinforcer. In this way, classical and operant conditioning are always intertwined.

Stimulus-control techniques are widely used in the laboratory to study perception and other psychological processes in animals. For example, the rat would not be able to respond appropriately to light-on and light-off conditions if it could not see the light. Following this logic, experiments using stimulus-control methods have tested how well animals see colors, hear ultrasounds, and detect magnetic fields. That is, researchers pair these discriminative stimuli with responses they know the animals already understand (such as pressing the lever). In this way, the researchers can test if the animals can learn to press the lever only when an ultrasound is played, for example.

These methods can also be used to study "higher" cognitive processes. For example, pigeons can learn to peck at different buttons in a Skinner box when pictures of flowers, cars, chairs, or people are shown on a miniature TV screen (Wasserman, 1995). Pecking button 1 (and no other) is reinforced in the presence of a flower image, button 2 in the presence of a chair image, and so on. Pigeons can learn the discrimination readily, and, under the right conditions, will even peck the correct buttons associated with pictures of *new* flowers, cars, chairs, and people they have never seen before. The birds have learned to **categorize** the sets of stimuli. Stimulus-control methods can be used to study how such categorization is learned, and for biological psychologists these methods can be used along with specific brain lesions to investigate the brain areas involved in categorization in animals.

10.2.8 Operant Conditioning Involves Choice

Another thing to know about operant conditioning is that the response always requires choosing one behavior over others. The student who goes to the bar on Thursday night chooses to drink instead of staying at home and studying. The rat chooses to press the lever instead of sleeping or scratching its ear in the back of the box. The alternative behaviors are each associated with their own reinforcers. And the tendency to perform a particular action depends on both the reinforcers earned for it and the reinforcers earned for its alternatives.

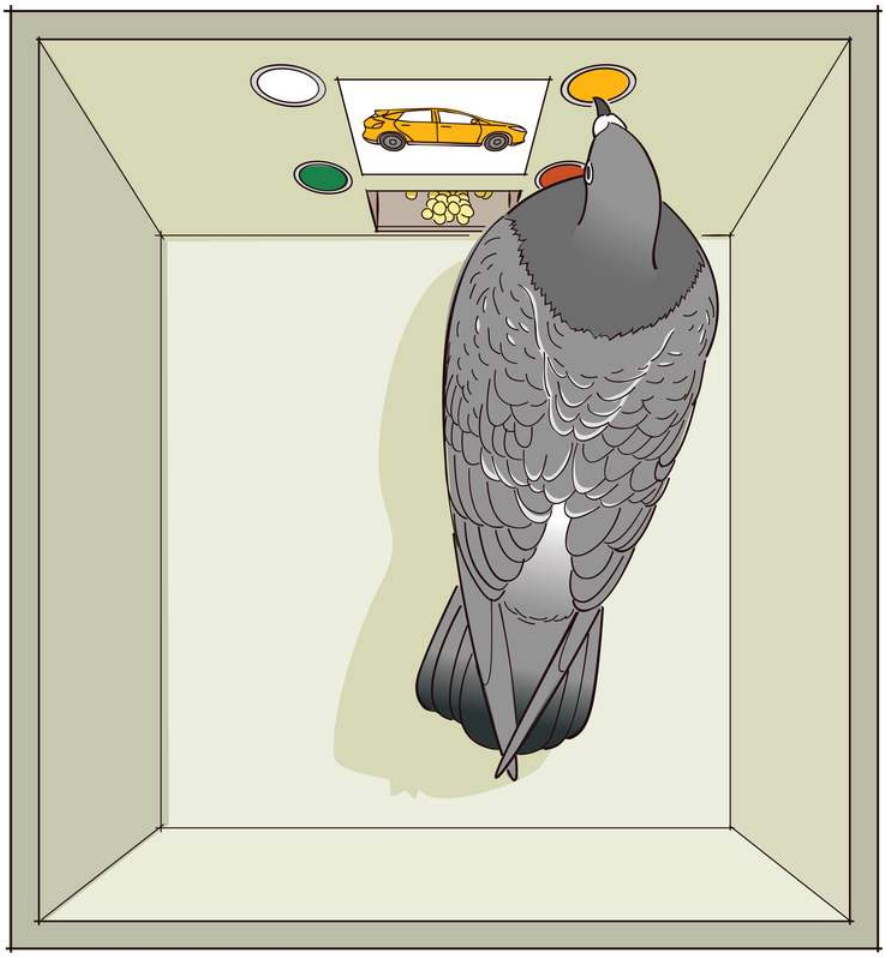


Figure 10.2.8: Pigeon in Skinner Box.

To investigate this idea, choice has been studied in the Skinner box by making two levers available for the rat (or two buttons available for the pigeon), each of which has its own reinforcement or payoff rate. A thorough study of choice in situations like this has led to a rule called the **quantitative law of effect** (Herrnstein, 1970), which can be understood without going into quantitative detail: The law acknowledges the fact that the effects of reinforcing one behavior depend crucially on how much reinforcement is earned for the behavior's alternatives. For example, if a pigeon learns that pecking one light will reward two food pellets, whereas the other light only rewards one, the pigeon will only peck the first light. However, what happens if the first light is more strenuous to reach than the second one? Will the cost of energy outweigh the bonus of food? Or will the extra food be worth the work? In general, a given reinforcer will be less reinforcing if there are many alternative reinforcers in the environment. For this reason, alcohol, sex, or drugs may be less powerful reinforcers if the person's environment is full of other sources of reinforcement, such as achievement at work or love from family members. (Image: from NOBA, Conditioning and Learning; https://nobaproject.com/modules/cond...g-and-learning)

An important distinction of operant conditioning is that it provides a method for studying how consequences influence "voluntary" behavior. As discussed above, the rat's decision to press the lever is voluntary, in the sense that the rat is free to make and repeat that response whenever it wants. Classical conditioning, on the other hand, is just the opposite—depending instead on "involuntary" behavior (e.g., the dog doesn't choose to drool; it just does). So, whereas the rat must actively participate and perform some kind of behavior to attain its reward, the dog in Pavlov's experiment is a passive participant. One of the lessons of operant conditioning research, then, is that voluntary behavior is strongly influenced by its consequences (recall the law of effect).

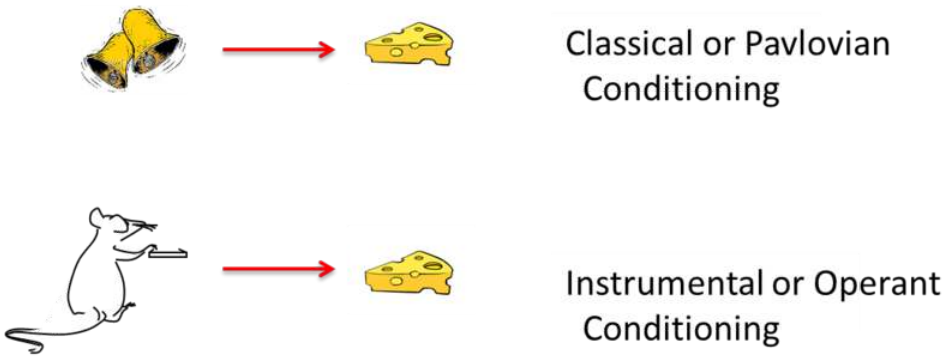


Figure **10.2.9**: Basic elements of classical and instrumental conditioning. The two types of learning differ in *what* is learned. In classical conditioning, the animal has learned to associate a *stimulus* with an adaptively significant event, food in this case. In operant conditioning, the animal has learned to associate a voluntary *behavior*, pressing the lever, with an adaptively significant event, food. (Image and caption from M. Boulton, NOBA, Conditioning and Learning; courtesy of Bernard W. Balleine; https://nobaproject.com/modules/cond...g-and-learning).

The two types of conditioning occur continuously throughout our lives. It has been said that “much like the laws of gravity, the laws of learning are always in effect” (Spreat & Spreat, 1982).

10.2.9 Cognition in Instrumental Learning

Modern research also indicates that reinforcers do more than merely strengthen or “stamp in” the behaviors they are a consequence of, as was Thorndike’s original view. Instead, animals learn about the specific consequences of each behavior, and will perform a behavior depending on how much they currently want—or “value”—its consequence. This idea is best illustrated by a phenomenon called the **reinforcer devaluation effect** (see Colwill & Rescorla, 1986). A rat is first trained to perform two instrumental actions (e.g., pressing a lever on the left, and on the right), each paired with a different reinforcer (e.g., a sweet sucrose solution, and a food pellet). At the end of this training, the rat tends to press both levers, alternating between the sucrose solution and the food pellet. In a second phase, one of the reinforcers (e.g., the sucrose) is then separately paired with illness. This conditions a taste aversion to the sucrose. In a final test, the rat is returned to the Skinner box and allowed to press either lever freely. No reinforcers are presented during this test (i.e., no sucrose or food comes from pressing the levers), so behavior during testing can only result from the rat’s memory of what it has learned earlier. Importantly here, the rat chooses *not* to perform the response that once produced the reinforcer that it now has an aversion to (e.g., it won’t press the sucrose lever). This means that the rat has learned and remembered the reinforcer associated with each response, and can combine that knowledge with the knowledge that the reinforcer is now “bad.” Reinforcers do not merely stamp in responses; response varies with how much the rat wants/doesn’t want a reinforcer. As described above, in operant conditioning, the animal tracks the changing reinforcement and punishment contingencies in its environment, as part of a dynamic mental model or neural representation of its world, and it adjusts its behavior accordingly.

Habituation, classical conditioning, and operant conditioning are just three types of learning. Each contributes to adaptation and increases biological fitness (chances of solving the problems associated with survival and reproduction). There are many other types of learning as well, often quite specialized to perform a particular biological function. These specialized forms of learning, also known as adaptive specializations of learning, have been studied mostly by ethologists and behavioral biologists, but biological psychologists are becoming increasingly interested in such forms of learning and their importance. For instance, psychologists have studied one of these specialized forms of learning, taste aversion learning, extensively. In addition to this form, we will now also examine adaptive specializations of learning involved in bird navigation by the stars during migration, bee navigation by the sun, and acquisition of bird song, which some researchers have compared to human language acquisition.

10.2.10 Specialized Forms of Learning (Adaptive Specializations of Learning)

> "Biological mechanisms are adapted to the exigencies of the functions they serve. The function of memory is to carry information forward in time. The function of learning is to extract from experience properties of the environment likely to be useful in the determination of future behavior." (Gallistel, 2003, p. 259).

> "One cannot use a hemoglobin molecule as the first stage in light transduction and one cannot use a rhodopsin molecule as an oxygen carrier, any more than one can see with an ear or hear with an eye. Adaptive specialization of mechanism is so ubiquitous and so obvious in biology, . . . it is odd but true that most past and contemporary theorizing about learning does not assume that learning mechanisms are adaptively specialized for the solution of particular kinds of problems. Most theorizing assumes that there is a general purpose learning process in the brain, a process adapted only to solving the problem of learning. . . . , this is equivalent to assuming that there is a general purpose sensory organ, which solves the problem of sensing." (Gallistel, 2000, p. 1179).

As the quotes above imply, learning, the acquisition of information during the lifetime of the individual animal, comes in many different forms. Many specialized forms of learning are highly specific for the solution of specific adaptive problems (problem domains) often found in only one or a few species. These types of learning probably involve specialized neural circuits, organized for their particular specialized form of learning by natural selection.

A familiar sight is ducklings walking or swimming after their mothers. Hatchling ducks recognize the first adult they see, usually their mother, and make a bond with her which induces them to follow her. This type of non-associative learning is known as imprinting. Imprinting is a form of learning occurring at a particular age or a particular life stage that is very important in the maturation process of these animals as it encourages them to stay near their mother in order to be protected, greatly increasing their chances of survival. Imprinting provides a powerful example of **biologically prepared learning** in response to particular genetically determined cues. In the case of imprinting, the duckling becomes imprinted on the first moving object larger than itself that it sees after hatching. Because of this, if newborn ducks see a human before they see their mother, they will imprint on the human and follow him or her in just the same manner as they would follow their real mother. Because this form of learning is biologically prepared, some cues that trigger the duckling to learn to follow its mother (or a person) are innately programmed into the duckling's genetically controlled brain circuitry while the details of the imprinting object (usually their real mother) are learned from the duckling's initial experience with a moving object larger than itself (again, most likely its real mother). Though this learning is very rapid because it is genetically (biologically) facilitated, it is also very resilient, its effects on behavior lasting well into adulthood. This form of learning illustrates well the general principle that all learning relies upon an underpinning of genetic information about some features of the learning situation while details are filled in by learning through experience. For example, genetic information programs the duckling's brain to follow the first moving thing it sees after hatching that is larger than itself, the details about what that object looks like are added into the bird's memory by learning during imprinting.

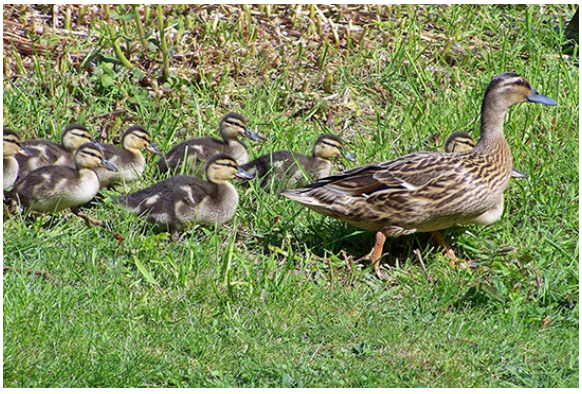


Figure **10.2.10**: Imprinting is the learned attachment of ducklings to their mother or, under rare circumstances, to the first moving object larger than themselves that they see after hatching (Image from LibreTexts, **Book: General Biology (Boundless)**, Learned Animal Behavior 45.7A; caption by Kenneth A. Koenigshofer, Ph.D.; licensed under

10.2.11 Taste Aversion Learning

Taste aversion learning is a specialized form of learning that helps omnivorous animals (those that eat a wide range of foods) to quickly learn to avoid eating substances that might be poisonous. In rats, coyotes, and humans, for example, eating a new food that is later followed by sickness causes avoidance of that food in the future. With food poisoning, although having fish for dinner may not normally be something to be concerned about (i.e., a “neutral stimulus”), if it causes you to get sick, you will now likely associate that neutral stimulus (the fish) with the adaptively significant event of getting sick.

This specialized form of learning is genetically or **"biologically prepared"** so that taste-illness associations are easily formed by a single pairing of a taste with illness, even if taste and illness are separated by extended periods of time from 10-15 minutes up to several hours. Colors of food, or sounds present when the food is consumed, cannot be associated with illness, only taste and illness can be associated. This is known as "belongingness," an example of "biological preparedness," in which learning has specialized properties as a result of genetic evolution--in this case, only taste and illness can be readily associated, not visual or auditory stimuli and illness (Garcia and Koelling, 1966; Seligman, 1971). A second difference is that taste aversion learning requires only a single pairing of taste and illness, whereas classical conditioning usually requires many pairings of the CS and US. The usual requirements for multiple pairings and close temporal contiguity between stimuli don't apply in learned taste aversion. This makes adaptive sense, because in the wild, sickness from a new food which is toxic won't occur immediately, but only after it has had time to pass into the digestive system and be absorbed. And when it comes to poisons, you may not get a second chance to learn to avoid that substance in the future.

This genetically prepared form of learning evolved over generations of natural selection in omnivorous species which consume a large variety of foods. Species that are very specialized feeders, such as Koala bears which eat eucalyptus leaves exclusively or baleen whales which filter ocean water for microscopic organisms, have not evolved taste aversion learning. They simply don't need it because they never experience novel foods that could possibly pose a threat.

10.2.12 Adaptively Specialized Learning for Navigation and Song Acquisition

Another example of a specialized form of learning evolved for the solution of a specific adaptive problem is the **learning of the night sky by** a species of bird called **Indigo Buntings.** Have you ever noticed that the stars appear to rotate throughout the night around an area of the sky, the celestial pole, close to the north star? Indigo buntings migrate south for the winter and then return north when temperatures warm. Experiments have shown that they use the stars, specifically the Big Dipper containing the pole star, Polaris, and the circumpolar stars within about 35 degrees of the center of rotation of the night sky to guide them on their journey of several thousand kilometers. But there is a special problem presented by the Earth's environment. Because of a wobble of the Earth's axis, the celestial pole and the north star change every few thousand years. A different star and star pattern will then indicate north and become the beacon that guides the migration of the Indigo Buntings.

These birds are genetically programmed to migrate and to use the night sky to guide them. But the specific constellation and star patterns that guide them must be learned because celestial north and the stars that mark it change too frequently in evolutionary time to be genetically encoded into their brains by evolution. The north star, whether Polaris, or another star thousands of years from now, is the only star that appears approximately stationary throughout the night, while all the other stars appear to turn around it as the night progresses. Indigo Bunting nestlings essentially sit up at night and watch the night sky from their nests. Because of a genetically programmed set of brain circuits, they note the star which barely moves, picking it out from all the others, which appear to move throughout the night (the movement of course is just apparent; the stars are relatively fixed, what's moving is the Earth rotating under them). By learning which star is essentially stationary in the night sky and the patterns of stars around it, they learn the correct star and star patterns to use for navigation when it comes time to migrate. The time for buntings to learn the night sky is limited. If deprived of exposure to the night sky until later in life they cannot learn the night sky as adults (Gallistel, et al., 1991, p. 16).

Two other examples of specialized forms of learning are of interest here. **Bees learn to navigate by** the position of the **sun** which changes with the date, time, and place on the Earth's surface. When moved to a new location, they learn to update their navigation to compensate for the change in location. This learning is strongly guided by innate, genetically evolved information stored in the bee brain (Dyer & Dickson, 1994; Towne, 2008).

One other example is of special interest, **song learning in song birds**. Birds of a specific species can only learn the song of their own species illustrating genetic constraints on what songs they can learn (Gallistel, et al., 1991, p. 21). White-crowned sparrows show variations in their song depending upon their geographical location, akin to dialects in human language. Experiments have shown that young white-crowned sparrows learn the specific dialect by exposure to it during a critical period for song learning in the species. A critical period during which learning must occur indicates another genetic constraint on learning and is similar to critical periods evident in imprinting, learning the night sky by Indigo Buntings, and a sensitive period in humans prior to adolescence for language acquisition.

The learning involved in the above examples is not classical or operant conditioning, but each is a very specialized form of learning in one particular species for solution to a specific adaptive problem (a specific domain). Note that the learned information supplements and interacts with information in the genes, guiding the animal to attend to and to readily learn highly specific information. In the case of the buntings, in a sense, the birds "imprint" on the correct stars for navigation later in life.

These examples illustrate a general principle that learning does not occur without genetically evolved information guiding and facilitating the learning in various ways. Note that in each of the examples above, genetically internalized, implicit “knowledge” about invariants in the learning situation pre-structures the learning, facilitating the capture of the problem-relevant details too variable and short-term to be captured directly by natural selection in genetic mechanisms (Koenigshofer, 2017).

This principle applies not only to highly specialized forms of learning such as that in Indigo buntings or the other examples above, but even in more general forms of learning such as classical conditioning (Chiappe & MacDonald, 2005; Gallistel, 1992; Koenigshofer, 2017) and causal learning in children (Koenigshofer, 2017; Walker and Gopnik, 2014), each of which involves innate knowledge about general features of each learning situation. For example, in the cause of causal learning, children have an innate predisposition to understand cause-effect as a general property of the world that guides their learning of specific cause-effect relations in their particular environment (Koenigshofer, 2017). Furthermore, several genetically internalized “default assumptions” built into conditioning and causal learning mechanisms by natural selection are that “causes are reliable predictors of their effects, that causes precede their effects, . . . that in general, causes tend to occur in close temporal proximity to their effects (Revulsky, 1985; Staddon, 1988) . . . and the temporal contiguity of cause and effect is a general feature of the world” (Chiappe and MacDonald, 2005).

In addition, from this perspective, the fact that conditioning (except in taste aversion learning) generally requires multiple pairings of CS and US, or of operant response and its effect, is *not* a shortcoming or a “weakness” of conditioning processes but rather may be an evolved adaptive feature of conditioning fashioned by natural selection to prevent formation of potentially spurious (and therefore, maladaptive) associations (Koenigshofer, 2017).

Gallistel (1992) argues that even classical conditioning is a specialized form of learning that performs important biological functions. According to Gallistel, classical conditioning was shaped by natural selection to discover "what predicts what" in the environment--in mathematical terms, according to Gallistel, a problem in multivariate non-stationary time series analysis. Time series because what is being learned is the temporal dependence or contingency of one event on another; multivariate time series because many events/variables may or may not predict the US; and non-stationary because the contingencies between CS and US often change over time (Gallistel, 2000, p. 1186).

10.2.13 Observation learning

Observation learning is learning by watching the behavior of others. It is obviously an extremely important form of learning in us, but it is also "an ability common to primates, birds, rodents, and insects" (Dawson et al., 2013). It plays a crucial role in human social learning. Imagine a child walking up to a group of children playing a game on the playground. The game looks fun, but it is new and unfamiliar. Rather than joining the game immediately, the child opts to sit back and watch the other children play a round or two. Observing the others, the child takes note of the ways in which they behave while playing the game. By watching the behavior of the other kids, the child can figure out the rules of the game and even some strategies for doing well at the game.



Figure **10.2.11**: Children observing a social model (an experienced chess player) to learn the rules and strategies of the game of chess. [Image: David R. Tribble, https://nobaproject.com/modules/cond...g-and-learning, NOBA, Conditioning and Learning; CC BY-SA 3.0]

Observational learning is a component of Albert Bandura's **Social Learning Theory** (Bandura, 1977), which posits that individuals can learn novel responses via observation of key others' behaviors. Observational learning does not necessarily require reinforcement, but instead hinges on the presence of others, referred to as **social models**. Social models in humans are typically of higher status or authority compared to the observer, examples of which include parents, teachers, and older siblings. In the example above, the children who already know how to play the game could be thought of as being authorities—and are therefore social models—even though they are the same age as the observer. By observing how the social models behave, an individual is able to learn how to act in a certain situation. Other examples of observational learning might include a child learning to place her napkin in her lap by watching her parents at the dinner table, or a customer learning where to find the ketchup and mustard after observing other customers at a hot dog stand.

Bandura theorizes that the observational learning process consists of four parts. The first is *attention*—as, quite simply, one must pay attention to what s/he is observing in order to learn. The second part is *retention*: to learn one must be able to retain the behavior s/he is observing in memory. The third part of observational learning, *initiation*, acknowledges that the learner must be able to execute (or initiate) the learned behavior. Lastly, the observer must possess the *motivation* to engage in observational learning. In our vignette, the child must want to learn how to play the game in order to properly engage in observational learning. Bandura, Ross, & Ross (1963) demonstrated that children who observed aggression in adults showed less aggressive behavior if they witnessed the adult model receive punishment for their aggression. Bandura referred to this process as **vicarious reinforcement**, as the children did not experience the reinforcement or punishment directly, yet were still influenced by observing it.

10.2.14 Observation Learning and Cultural Transmission Improve Biological Fitness in Non-human Animals

Orangutans in protected preserves have been seen copying humans washing clothes in a river. After watching humans engage in this behavior, one of the animals took pieces of clothing from a pile of clothes to be washed and engaged in clothes washing behavior in the river, imitating behavior it had recently observed in humans. Orangutans also use observation learning to copy behaviors of other orangutans. Observation learning has also been reported in wild and captive chimpanzees and in other primates such as Japanese macaque monkeys. One of the most thoroughly studied examples of observation learning in animals is in Japanese macaques.

There is a large troop of macaques (Old World monkeys; see Chapter 3) which lives near the beaches on an island in Japan, Koshima Island. Researchers were interested to see how these animals would respond if the researchers scattered novel foods such as wheat grain on the sand. The animals meticulously picked out the grain from the sand one grain at a time, laboriously cleaning it a grain at a time. However, researchers reported that after a while one Japanese macaque in the troop invented an efficient method for cleaning the grain by scooping up handfuls of wheat grain and sand and throwing the mixture into the water. The wheat grains floated while the sand sank. This macaque monkey then scooped up quantities of clean grain floating on the surface of the water and ate its fill, repeating its novel grain cleaning behavior again and again. Although this showed impressive intelligence and inventiveness on the part of this monkey, just as significant was the fact that **other members of the troop observed this behavior and copied it**. By observation learning, most of the troop learned this innovative method of cleaning and separating grain from sand. As time passed, youngsters observed older members engaging in this **learned behavior** and copied it so that the behavior was **passed over several generations** (Schofield et al., 2018). This example of observation learning illustrates one of its most important biological functions--observation learning is a primary mechanism of **cultural transmission of learned behavior across generations**, not only in animals like the macaques, but even more so in humans. The effect of the cultural transmission of learned behavioral adaptations from generation to generation produces "cumulative culture . . . characterized as a ‘ratchet,’ yielding progressive innovation and improvement over generations (Tomasello et al. 1993). The process can be seen as repeated inventiveness that leads to incrementally better adaptation; that is, more efficient, secure, . . . survival and reproduction" (Schofield et al., 2018, p. 113). Efficient cultural transmission of successful learned behavior is **enormously powerful in boosting biological fitness**, and it accounts for those features of human life, such as science, technology, governments, and so on, that distinguish us most from all other species on the planet (Koenigshofer, 2011, 2016).

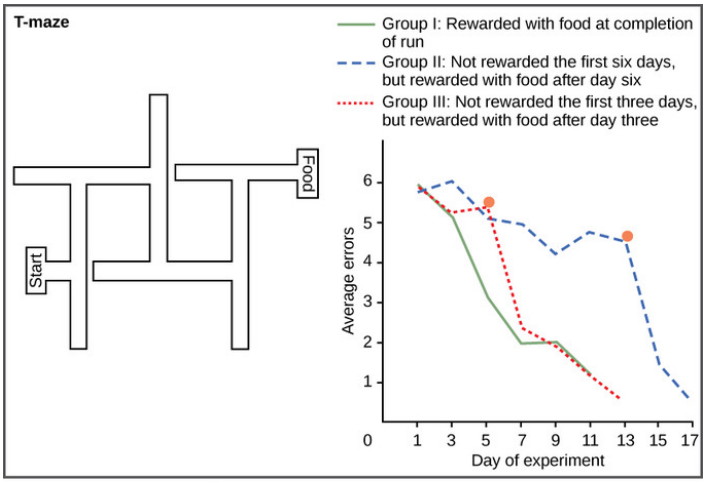
10.2.15 **Cognitive Learning**

Classical and operant conditioning are only two of the ways for humans and other intelligent animals to learn. Some primates, including humans, are able to learn by imitating the behavior of others and by taking instructions. The development of complex language by humans has made cognitive learning, a change in knowledge as a result of experience or mental manipulation of existing knowledge, the most prominent method of human learning. In fact, that is how you are learning right now by reading this information, you are experiencing a change in your knowledge. Humans, and probably some non-human animals, can make mental images of objects or organisms, imagining changes to them or behaviors by them as they anticipate the consequences. Cognitive learning is so powerful that it can be used to understand conditioning (discussed in the previous modules). In the reverse scenario, conditioning cannot help someone learn about cognition.

Classic work on cognitive learning was done by Wolfgang Köhler with chimpanzees. He demonstrated that these animals were capable of abstract thought by showing that they could learn how to solve a puzzle. When a banana was hung in their cage too high for them to reach, along with several boxes placed randomly on the floor, some of the chimps were able to stack the boxes one on top of the other, climb on top of them, and get the banana. This implies that they could visualize the result of stacking the boxes even before they had performed the action. This type of learning is much more powerful and versatile than conditioning.

Cognitive learning is not limited to primates, although they are the most efficient in using it. Maze-running experiments done with rats in the 1920s were the first to show cognitive skills in a simple mammal, the rat. The motivation for the animals to work their way through the maze was the presence of a piece of food at its end. In these studies, the animals in Group I were run in one trial per day and had food available to them each day on completion of the run. Group II rats were not fed in the maze for the first six days and then subsequent runs were done with food for several days after. Group III rats had food available on the third day and every day thereafter. The results were that the control rats, Group I, learned quickly, figuring out how to run the maze in seven

days. Group III did not learn much during the three days without food, but rapidly caught up to the control group when given the food reward. Group II learned very slowly for the six days with no reward to motivate them. They did not begin to catch up to the control group until the day food was given; it then took two days longer to learn the maze. Results suggested that although there was no reward for the rats in groups II and III during several days at the beginning of the experiment, the rats were still learning. This is evidenced particularly by the performance of group III. Although not given any food reward in the maze for the first three days of the experiment, nevertheless, once food reward was added on day 4, the maze learning performance of rats in this group rapidly caught up with the control group (Group I) that had received food reward in the maze every day from the start of the experiment. This showed that even in the absence of food reward, the rats were learning about the maze. This was important because at that time many psychologists believed that learning can only occur in the presence of reinforcement. This experiment showed that learning information about a maze, gaining knowledge about it, can occur even in the absence of reinforcement. Some have referred to this as "latent learning." The learning that has taken place remains hidden or latent in behavior until a motivating factor such as food reward stimulates action that reveals the learning that had previously not been apparent in observable behavior. In this case, the latent learning of the maze in groups II and III became apparent once food reward was made available for the rats in these groups on day 6 and day 3 respectively. **Cognitive learning involves acquisition of knowledge**, in this case, about a maze; in this example, it took place in rats without food reward, as evidenced by their performance in the maze once food reward was presented as a motivator. This procedure revealed the learning that had taken place in the first few days of the experiment.



Redrawn after H. C. Blodgett, The effect of the introduction of reward upon the maze performance of rats. Univ. Calif. Publ. Psychol., 1929, 4, No. 8, pages 117 and 120.

Figure **10.2.12**: **Cognitive learning**: Group I (the green solid line) found food at the end of each trial; group II (the blue dashed line) did not find food for the first 6 days; and group III (the red dotted line) did not find food during runs on the first three days. Notice that rats given food earlier learned faster and eventually caught up to the control group. The orange dots on the group II and III lines show the days when food rewards were added to the mazes. (Image from: Book: General Biology (Boundless), Learned Animal Behavior 45.7C Unless otherwise noted, LibreTexts content is licensed by CC BY-NC-SA 3.0. Legal. https://bio.libretexts.org/Bookshelv...d_Sociobiology

Clearly this type of learning is different from conditioning. Although one might be tempted to believe that the rats simply learned how to find their way through a conditioned series of right and left turns, Edward C. Tolman proved a decade later that the rats were making **a representation** of the maze **in their minds**, which he called a “cognitive map.” This was an early demonstration of the power of cognitive learning and how these abilities were not limited just to humans. Research discussed more fully in Chapter 14 indicates that the **visual, motor, and parietal cortical areas as well as the hippocampus** are involved in ability to form **cognitive maps** and engage in cognitive learning.

10.2.16 Key Points

- Cognitive learning involves change in knowledge as a result of experience and the mental manipulation of information; it is a great deal more powerful and flexible than either operant or classical conditioning.
- The development of complex language by humans has made cognitive learning the most prominent method of human learning.
- Cognitive learning is not limited to primates; rats have demonstrated the ability to build cognitive maps, as well, which are mental representations used to acquire, code, store, recall, and decode information about the environment.

10.2.17 Key Terms

- **cognitive map**: a mental representation which serves an organism to acquire, code, store, recall, and decode information about the relative locations and attributes of phenomena in their everyday environment
- **cognitive learning**: the process by which one acquires knowledge or skill in cognitive processes, which include reasoning, abstract thinking, and problem solving

10.2.18 Outside Resources

Article: Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist, 43*, 151–160.

Book: Bouton, M. E. (2007). *Learning and behavior: A contemporary synthesis*. Sunderland, MA: Sinauer Associates.

Book: Bouton, M. E. (2009). Learning theory. In B. J. Sadock, V. A. Sadock, & P. Ruiz (Eds.), *Kaplan & Sadock's comprehensive textbook of psychiatry* (9th ed., Vol. 1, pp. 647–658). New York, NY: Lippincott Williams & Wilkins.

Book: Domjan, M. (2010). *The principles of learning and behavior* (6th ed.). Belmont, CA: Wadsworth.

Video: Albert Bandura discusses the Bobo Doll Experiment.





10.2.19 Vocabulary (click on terms for definitions)

Blocking

In classical conditioning, the finding that no conditioning occurs to a stimulus if it is combined with a previously conditioned stimulus during conditioning trials. Suggests that information, surprise value, or prediction error is important in conditioning.

Categorize

To sort or arrange different items into classes or categories.

Classical conditioning

The procedure in which an initially neutral stimulus (the conditioned stimulus, or CS) is paired with an unconditioned stimulus (or US). The result is that the conditioned stimulus begins to elicit a conditioned response (CR). Classical conditioning is nowadays considered important as both a behavioral phenomenon and as a method to study simple associative learning. Same as Pavlovian conditioning. Classical conditioning involves learning the predictive relationship between two stimulus events.

Conditioned compensatory response

In classical conditioning, a conditioned response that opposes, rather than is the same as, the unconditioned response. It functions to reduce the strength of the unconditioned response. Often seen in conditioning when drugs are used as unconditioned stimuli.

Conditioned or conditional response (CR)

The response that is elicited by the conditioned stimulus after classical conditioning has taken place. Pavlov used the term conditional response, the response conditional upon prior learning.

Conditioned or conditional stimulus (CS)

An initially neutral stimulus (like a bell, light, or tone) that elicits a conditioned/conditional response after it has been associated with an unconditioned/unconditional stimulus. The CS comes to act as a signal predicting the coming occurrence of the US (the unconditioned stimulus; the stimulus unconditional upon prior learning and which naturally leads to the unconditioned response, a response unconditional upon prior learning; often a reflex response such as salivation, as in Pavlov's original experiments with dogs).

Context

Stimuli that are in the background whenever learning occurs. For instance, the Skinner box or room in which learning takes place is the classic example of a context. However, “context” can also be provided by internal stimuli, such as the sensory effects of drugs (e.g., being under the influence of alcohol has stimulus properties that provide a context) and mood states (e.g., being happy or sad). It can also be provided by a specific period in time—the passage of time is sometimes said to change the “temporal context.”

Discriminative stimulus

In operant conditioning, a stimulus that signals whether the response will be reinforced. It is said to “set the occasion” for the operant response.

Extinction

Decrease in the strength of a learned behavior that occurs when the conditioned stimulus is presented without the unconditioned stimulus (in classical conditioning) or when the behavior is no longer reinforced (in instrumental conditioning). The term describes both the procedure (the US or reinforcer is no longer presented) as well as the result of the procedure (the learned response declines). Behaviors that have been reduced in strength through extinction are said to be “extinguished.”

Fear conditioning

A type of classical or Pavlovian conditioning in which the conditioned stimulus (CS) is associated with an aversive unconditioned stimulus (US), such as a foot shock. As a consequence of learning, the CS comes to evoke fear. The phenomenon is thought to be involved in the development of anxiety disorders in humans.

Goal-directed behavior

Instrumental behavior that is influenced by the animal’s knowledge of the association between the behavior and its consequence and the current value of the consequence. Sensitive to the reinforcer devaluation effect.

Habit

Instrumental behavior that occurs automatically in the presence of a stimulus and is no longer influenced by the animal’s knowledge of the value of the reinforcer. Insensitive to the reinforcer devaluation effect.

Instrumental conditioning

Process in which animals learn about the relationship between their behaviors and their consequences. Also known as operant conditioning.

Law of effect

The idea that instrumental or operant responses are influenced by their effects. Responses that are followed by a pleasant state of affairs will be strengthened and those that are followed by discomfort will be weakened. Nowadays, the term refers to the

idea that operant or instrumental behaviors are lawfully controlled by their consequences.

Observational learning

Learning by observing the behavior of others.

Operant

A behavior that is controlled by its consequences. The simplest example is the rat's lever-pressing, which is controlled by the presentation of the reinforcer.

Operant conditioning

See instrumental conditioning.

Pavlovian conditioning

See classical conditioning.

Prediction error or prediction value

When the outcome of a conditioning trial is different from that which is predicted by the conditioned stimuli that are present on the trial (i.e., when the US is surprising; in other words, an event must have predictive value with regard to the occurrence of the US or no conditioning to the event will take place). Prediction error is necessary to create Pavlovian conditioning (and associative learning generally). As learning occurs over repeated conditioning trials, the conditioned stimulus increasingly predicts the unconditioned stimulus, and prediction error declines. Conditioning works to correct or reduce prediction error; i.e. conditioning works to provide reliable prediction, one event reliably predicts a second event. In other words an event must have prediction value, its presence is reliably predictive of a future event that is not already reliably predicted by another stimulus event that is present. That is, stimuli which don't add to the reliability of prediction of another event/stimulus are ignored because they contain no predictive information. This shows that what animals learn in conditioning is the predictive relations between events. If a stimulus does not add to the predictability of the target event, then that stimulus is ignored.

Preparedness

The idea that an organism's evolutionary history can make it easy to learn a particular association. Because of preparedness, you are more likely to associate the taste of tequila, and not the circumstances surrounding drinking it, with getting sick. Similarly, humans are more likely to associate images of spiders and snakes than flowers and mushrooms with aversive outcomes like shocks.

Punisher

A stimulus that decreases the strength of an operant behavior when it is made a consequence of the behavior.

Quantitative law of effect

A mathematical rule that states that the effectiveness of a reinforcer at strengthening an operant response depends on the amount of reinforcement earned for all alternative behaviors. A reinforcer is less effective if there is a lot of reinforcement in the environment for other behaviors.

Reinforcer

Any consequence of a behavior that strengthens the behavior or increases the likelihood that it will be performed it again.

Reinforcer devaluation effect

The finding that an animal will stop performing an instrumental response that once led to a reinforcer if the reinforcer is separately made aversive or undesirable.

Renewal effect

Recovery of an extinguished response that occurs when the context is changed after extinction. Especially strong when the change of context involves return to the context in which conditioning originally occurred. Can occur after extinction in either classical or instrumental conditioning.

Social Learning Theory

The theory that people can learn new responses and behaviors by observing the behavior of others.

Social models

Authorities that are the targets for observation and who model behaviors.

Spontaneous recovery

Recovery of an extinguished response that occurs with the passage of time after extinction. Can occur after extinction in either classical or instrumental conditioning.

Stimulus control

When an operant behavior is controlled by a stimulus that precedes it.

Taste aversion learning

The phenomenon in which a taste is paired with sickness, and this causes the organism to reject—and dislike—that taste in the future.

Unconditioned response (UR)

In classical conditioning, an innate response that is elicited by a stimulus before (or in the absence of) conditioning.

Unconditioned stimulus (US)

In classical conditioning, the stimulus that elicits the response before conditioning occurs.

Vicarious reinforcement

Learning that occurs by observing the reinforcement or punishment of another person.

10.2.20 References

Chiappe, D., and MacDonald, K. (2005). The evolution of domain-general mechanisms in intelligence and learning. *Journal of General Psychology*, *132* (1), 5-40.

Dawson, E. H., Avarguès-Weber, A., Chittka, L., & Leadbeater, E. (2013). Learning by observation emerges from simple associations in an insect model. *Current Biology*, *23*(8), 727-730.

Dyer, F.C., and Dickinson, J.A. (1994). Development of sun compensation by honeybees: How partially experienced bees estimate the sun's course. *Proceedings of the National Academy of Sciences, USA, 91*, 4471-4474.

Gallistel, C. R., Brown, A. L., Carey, S., Gelman, R., & Keil, F. (1991). Lessons from animal learning for the study of cognitive development. *The epigenesis of mind: Essays on biology and cognition*, 3-36. Eds. Susan Carey and Rochel Gelman. Psychology Press, 2014.

Gallistel, CR. (1992). Classical conditioning as an adaptive specialization: A computational model. In D.L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 35-67). San Diego: Academic Press.

Gallistel, C. R. (2000). The replacement of general-purpose learning models with adaptively specialized learning modules. *The Cognitive Neurosciences*, *2*, 1179-1191.

Gallistel, C. R. (2003). The principle of adaptive specialization as it applies to learning and memory. In *Principles of learning and memory* (pp. 259-280). Birkhäuser, Basel.

Garcia, J., and Koelling R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science*, *4*, 123-124.

Kandel, E. (1976). *Cellular Basis of Behavior*. San Francisco. W.H. Freeman and Company.

Koenigshofer, K.A. (2011). *Mind Design: The Adaptive Organization of Human Nature, Minds, and Behavior*. Pearson Education. Boston.

Koenigshofer, K.A. (2016). *Mind Design: The Adaptive Organization of Human Nature, Minds, and Behavior*. Revised Edition. Amazon e-book.

Koenigshofer, K. A. (2017). General Intelligence: Adaptation to Evolutionarily Familiar Abstract Relational Invariants, Not to Environmental or Evolutionary Novelty. *The Journal of Mind and Behavior*, 119-153.

Linnman, C., Rougemont-Bücking, A., Beucke, J. C., Zeffiro, T. A., & Milad, M. R. (2011). Unconditioned responses and functional fear networks in human classical conditioning. *Behavioural brain research*, *221*(1), 237-245.

Rescorla, R. A. (1966). Predictability and number of pairings in Pavlovian fear conditioning. Psychon. Sci., 4, 383-384.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of comparative and physiological psychology*, *66* (1), 1.

Schmajuk, N. A., & DiCarlo, J. J. (1991). A neural network approach to hippocampal function in classical conditioning. *Behavioral neuroscience*, *105* (1), 82.

Schofield, D. P., McGrew, W. C., Takahashi, A., & Hirata, S. (2018). Cumulative culture in nonhumans: overlooked findings from Japanese monkeys?. *Primates*, *59* (2), 113-122.

Seligman, M. (1971). Phobias and preparedness. *Behavior Therapy*, 2, 307–321.

Tolman, E. C., and Brunswik, E. (1935). The organism and the causal texture of the environment. *Psychological review*, *42* (1), 43.

Towne, W. F. (2008). Honeybees can learn the relationship between the solar ephemeris and a newly-experienced landscape. *Journal of Experimental Biology*, *211*(23), 3737-3743.

10.2.21 Attributions

"Overview," "Habituation and Adaptation," "Conditioning and Biological Adaptation," "Conditioning Involves Learning Predictive Relations," "Operant Conditioning," "Specialized Forms of Learning," "Taste Aversion Learning," "Adaptively Specialized Learning for Navigation and Song Acquisition," and "Observation Learning and Cultural Transmission Improve Biological Fitness in Non-human Animals" were written by Kenneth A. Koenigshofer, PhD. and are licensed under CC BY 4.0.

Some text and images adapted by Kenneth A. Koenigshofer, PhD, from LibreTexts, Book: General Biology (Boundless), Learned Animal Behavior 45.7A, and from Mark E. Bouton (2021), University of Vermont, Conditioning and Learning. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology.* Champaign, IL: DEF publishers. Retrieved from http://noba; Conditioning and Learning by Mark E. Bouton is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Permissions beyond the scope of this license may be available in our Licensing Agreement.

Cognitive Learning adapted by Kenneth A. Koenigshofer, Ph.D., from LibreTexts, **Book: General Biology (Boundless)**, Learned Animal Behavior 45.7C Unless otherwise noted, LibreTexts content is licensed by CC BY-NC-SA 3.0. Legal.

LICENSES AND ATTRIBUTIONS

CC LICENSED CONTENT, SHARED PREVIOUSLY

- Curation and Revision. **Provided by**: Boundless.com. **License**: ***CC BY-SA: Attribution-ShareAlike***

CC LICENSED CONTENT, SPECIFIC ATTRIBUTION

- OpenStax College, Biology. October 17, 2013. **Provided by**: OpenStax CNX. **Located at**: **http://cnx.org/content/m44879/latest...ol11448/latest**. **License**: ***CC BY: Attribution***
- innate. **Provided by**: Wiktionary. **Located at**: **en.wiktionary.org/wiki/innate**. **License**: ***CC BY-SA: Attribution-ShareAlike***
- habituation. **Provided by**: Wiktionary. **Located at**: **en.wiktionary.org/wiki/habituation**. **License**: ***CC BY-SA: Attribution-ShareAlike***
- imprinting. **Provided by**: Wiktionary. **Located at**: **en.wiktionary.org/wiki/imprinting**. **License**: ***CC BY-SA: Attribution-ShareAlike***
- OpenStax College, Behavioral Biology: Proximate and Ultimate Causes of Behavior. October 17, 2013. **Provided by**: OpenStax CNX. **Located at**: **http://cnx.org/content/m44879/latest...e_45_07_06.jpg**. **License**: ***CC BY: Attribution***
- OpenStax College, Biology. October 17, 2013. **Provided by**: OpenStax CNX. **Located at**: **http://cnx.org/content/m44879/latest...ol11448/latest**. **License**: ***CC BY: Attribution***
- operant conditioning. **Provided by**: Wiktionary. **Located at**: **en.wiktionary.org/wiki/operant_conditioning**. **License**: ***CC BY-SA: Attribution-ShareAlike***
- classical conditioning. **Provided by**: Wiktionary. **Located at**: **en.wiktionary.org/wiki/classical_conditioning**. **License**: ***CC BY-SA: Attribution-ShareAlike***

- OpenStax College, Behavioral Biology: Proximate and Ultimate Causes of Behavior. October 17, 2013. **Provided by**: OpenStax CNX. **Located at**: http://cnx.org/content/m44879/latest...e_45_07_06.jpg. **License**: ***CC BY: Attribution***
- OpenStax College, Behavioral Biology: Proximate and Ultimate Causes of Behavior. October 17, 2013. **Provided by**: OpenStax CNX. **Located at**: http://cnx.org/content/m44879/latest...e_45_07_07.jpg. **License**: ***CC BY: Attribution***
- OpenStax College, Behavioral Biology: Proximate and Ultimate Causes of Behavior. October 17, 2013. **Provided by**: OpenStax CNX. **Located at**: http://cnx.org/content/m44879/latest...e_45_07_08.jpg. **License**: ***CC BY: Attribution***
- OpenStax College, Biology. October 17, 2013. **Provided by**: OpenStax CNX. **Located at**: http://cnx.org/content/m44879/latest...ol11448/latest. **License**: ***CC BY: Attribution***
- OpenStax College, Biology. October 23, 2013. **Provided by**: OpenStax CNX. **Located at**: http://cnx.org/content/m44879/latest...ol11448/latest. **License**: ***CC BY: Attribution***
- cognitive map. **Provided by**: Wikipedia. **Located at**: **en.Wikipedia.org/wiki/cognitive%20map**. **License**: *CC BY-SA: Attribution-ShareAlike*
- Boundless. **Provided by**: Boundless Learning. **Located at**: **www.boundless.com//biology/de...itive-learning**. **License**: *CC BY-SA: Attribution-ShareAlike*

This page titled 10.2: Types of Learning and Biological Adaptation is shared under a mixed license and was authored, remixed, and/or curated by Kenenth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

10.3: Introduction to Memory and the Brain

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

10.3.1 Learning Objectives

1. Explain the differences between learning and memory
2. Describe the different types of memory and the brain structures involved in each
3. Explain the three stage model of memory
4. Explain the differences between implicit and explicit memory
5. Describe types of declarative memory and their features
6. Discuss the case of H.M. and what it tells us about the role of the medial temporal lobes in declarative memory
7. Describe the concept of the "engram"?

10.3.2 Overview

As you probably remember from your course in introductory psychology, there are a number of different types and stages of memory, involving different brain areas and processes. In this section, we briefly outline the different types of memory and some of what is known about the localization of different types of memory in the brain. We discuss the intriguing case of H.M. whose profound loss of ability to form new long-term memory of facts and events led to much of what we know about declarative memory and the involvement of the medial temporal lobes in memory formation. Formation of memories about how to do things, procedural or motor memory involves different brain areas, in particular, the cerebellum.

10.3.3 Memory and the Brain

As discussed earlier in this chapter, learning can be defined as the acquisition of information during the lifetime of the individual organism. This information is stored in memory systems of the brain of the animal. By contrast, genetic information, which is acquired over the evolutionary history of the species, is stored in DNA. Remember that behavioral adaptations can be organized by information from both sources. The relative contribution of each source of information to the organization of any particular behavior varies with the type of behavior and the species. Typically, learning is more important to organization of behavior in the more complex animal species with larger, more complex brains, yet evidence of memory, as discussed earlier in this chapter, is even found in honey bees and other invertebrates.

Memory refers to the processes of storage and retrieval of learned information.

10.3.4 Types of Memory

Just as there are different types of learning, there are also different types of memory.

One major distinction is between **verbal memory**, memory for words and ideas stated in words (dependent on language areas in **frontal and temporal cortex**) and memories in the form of **visual images** (dependent upon **occipital, temporal, and parietal lobe cortex**; see Figure 10.3.1). For example, when I remember a hotel that I stayed at in Hong Kong many years ago, I get a visual image of the hotel and the streets of Hong Kong nearby. Memories can be in the form of other sensory modalities as well. When I remember talking to a friend on the phone, I can almost "hear" my friend's voice in my mind (memory in the form of an auditory "image"). Verbal memories themselves can often involve auditory "images" in your mind, the sounds of the words you are thinking about. Think about what a lemon tastes like and you will almost "taste" it in your mind (memory in the form of a taste "image").

Another important distinction between different forms of memory is memory for **facts and events**. This is **explicit or declarative memory**, which is of two types, **episodic** (memory of episodes in your life) and **semantic** memory (memory for facts and knowledge). A second type, **implicit memory**, is also of two types, **procedural memory** (memory for how to do things, such as skills like riding a bike, **sometimes called motor memory**), and memory for classically conditioned responses and priming effects.

10.3.5 Types of Memory and the Brain

Different areas of the brain are involved in these different types of memory. The **cerebellum**, located in the hindbrain, is involved in formation and storage of motor programs for learned skilled movements (**procedural or motor memory**) and is also the source of learning and memory of simple **conditioned reflexes** such classical conditioning of the eye blink reflex (Kim & Thompson, 1997; Christian & Thompson, 2003). By contrast, the **hippocampus** has been implicated in the **"consolidation"** or fixing of **new long term episodic or declarative memories** into permanent storage (see Bird & Burgess, 2008), as well as the **encoding of the temporal sequencing** of events (Ranganath & Hsieh, 2016), **spatial memory** (Bird & Burgess, 2008; Voss, et al., 2017), and **retrieval of memories by reactivating cortical representations** of the retrieved memories (Tanaka, et al., 2014).

10.3.6 The Three Stage Model of Memory and the Brain

One simple and influential model of memory proposes three stages of memory (see Baddeley, 1982):

Sensory memory------Short-term or working memory (STM)-----Long-term memory (LTM)

Sensory memory is very short duration but very high capacity memory. To illustrate, just close your eyes at this moment and you will "see" a **mental image** of what it was you were just looking at. That is visual sensory memory and likely involves the visual cortex in **occipital lobe**. Notice it is nearly a replica of what you were just looking at; the mental image is a very detailed image, an information dense, high information capacity form of storage. But the image quickly fades--that is, this stage of memory is of very brief duration, not more than a few seconds.

Short-term memory is a bit longer in duration, up to a few minutes. But the capacity of STM is not nearly as large as sensory memory. It is usually said that STM has the capacity to store 7 plus or minus 2 separate items of information at a time (Miller, 1956). This is a limited storage capacity compared to the other stages of memory. Short term memory is also called **working memory.** This is because when you use information from your memory to do some task you are using it while it is in this stage of memory, the short term memory store, thus the term working memory is often used to refer to this stage.

According to the three-stage model, information then moves from STM (working memory) into the third stage of memory, **Long term memory** (memories here may last a lifetime; this memory stage has very large capacity and very long duration). This movement of information from STM to LTM and its **"consolidation"** into a long term trace appears to critically involve the limbic system, more specifically, the **hippocampus** (Squire, et al., 2015), located in the **medial temporal lobes** (see Figure 10.3.1). When you are studying for an exam, you are trying to get information encoded into long-term memory. Consolidation of the new semantic memory by your hippocampus, and the surrounding medial temporal lobe, is essential to this process.

Information can also move from LTM back into STM or working memory. For example, if I ask you to name the capital of Japan, you can probably say Tokyo but this information was not in your conscious awareness before I asked you the question. The information had to be pulled from your long term memory into your conscious awareness. This movement of information from LTM (your long duration, large capacity memory) to your conscious awareness (into your "working memory") is **memory retrieval**. How does your brain do it? How can it locate and pull into your conscious working memory a piece of information from among the countless pieces of information stored in your long term memory over your lifetime? Apparently, there are something like **associative networks** formed in long term memory that help facilitate an efficient search when you are trying to locate a particular piece of information for retrieval into conscious working memory.

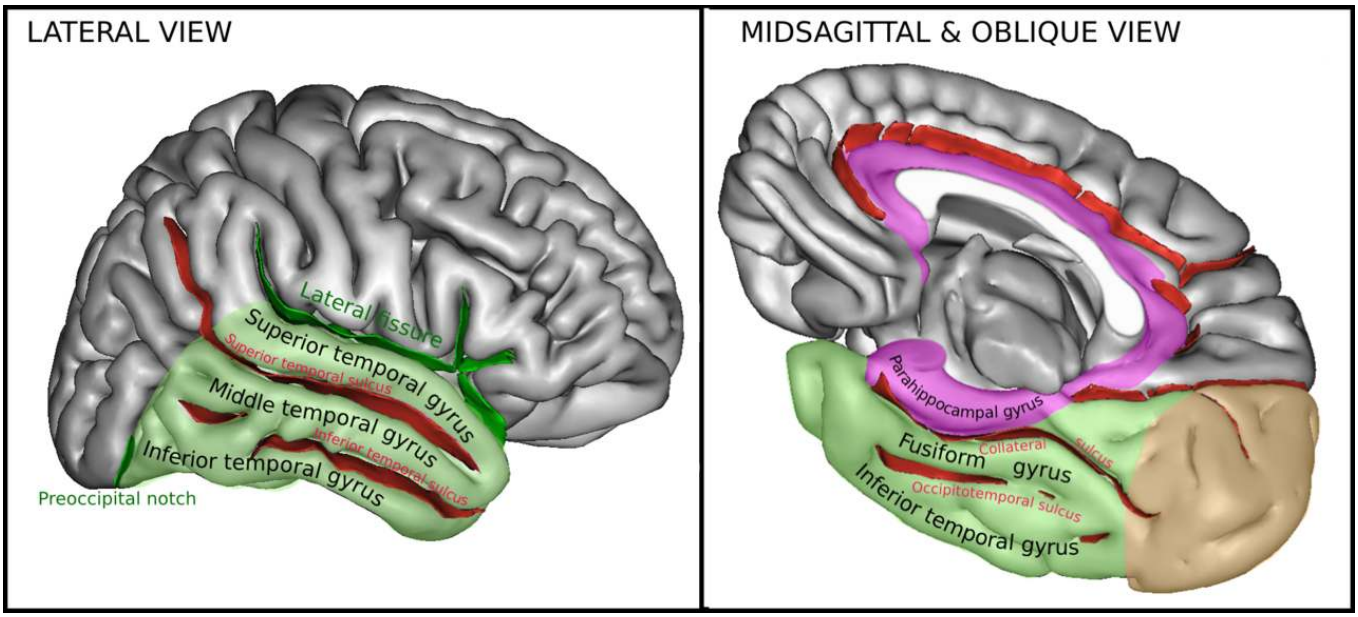
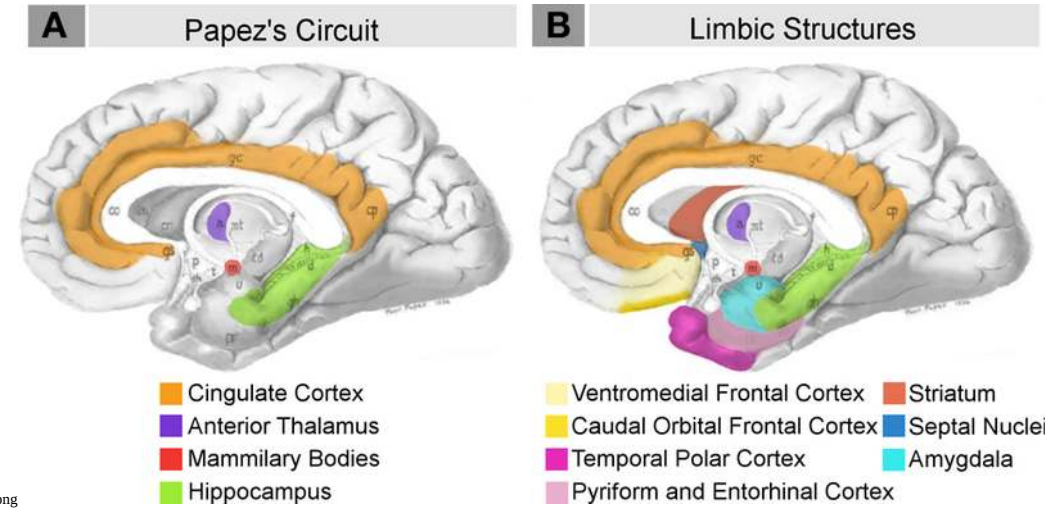


Figure **10.3.1**: Right human cerebral hemisphere highlighting temporal lobe (green) showing locations of major cortical gyri and fissures of temporal lobe. The hippocampus (not shown) is buried within medial temporal cortex important in declarative memory. (Image from Wikimedia Commons; Medial temporal cortex; TempCapts.png; by Sebastian023; https://commons.wikimedia.org/wiki/F...:TempCapts.png; licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license. Caption by Kenneth A. Koenigshofer, PhD.).

10.3.7 The Medial Temporal Lobes and Declarative Memory

Research has led to a consensus that the **medial temporal lobes** have an essential function in **declarative memory**. As Squire et al. (2004, p. 279) summarize: "The medial temporal lobe includes a system of anatomically related structures that are essential for declarative memory (conscious memory for facts and events). The system consists of the **hippocampal region** (CA fields, dentate gyrus, and subicular complex) and the adjacent **perirhinal, entorhinal, and parahippocampal cortices**."



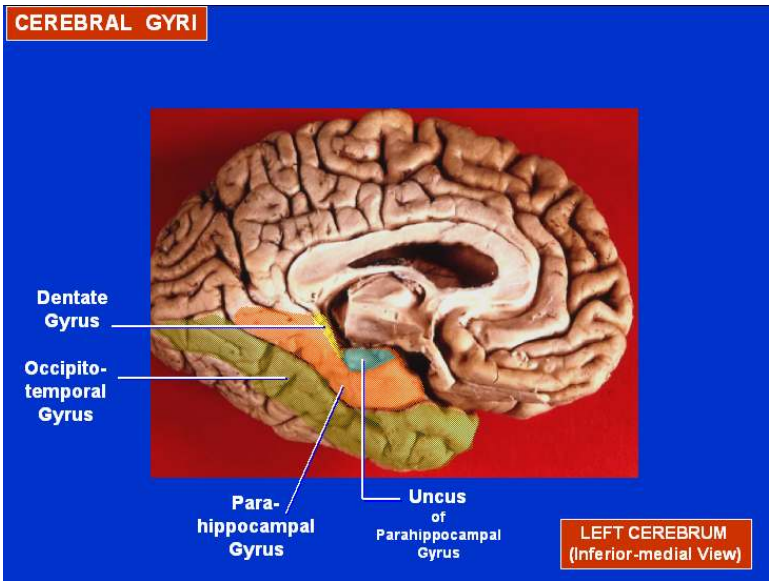
File:Parahippocampal gyrus - inferiror view.png

Figure **10.3.2**: (Left) Parahippocampal cortex. (Middle and Right). Papez Circuit and Limbic system structures. Some are within the medial temporal lobe and are involved in memory. Many limbic structures are also involved in emotion. (Images from Wikimedia Commons; (left) Parahippocampal gyrus; https://commons.wikimedia.org/wiki/F...riror_view.png; by Polygon data generated by Database Center for Life Science (DBCLS)[2]; licensed under Creative Commons Attribution-Share Alike 2.1 jp; (middle and right) File:Neural systems proposed to process emotion.png; https://commons.wikimedia.org/wiki/F...ss_emotion.png;

by Barger N, Hanson KL, Teffer K, Schenker-Ahmed NM and Semendeferi K; licensed under the Creative Commons Attribution 3.0 Unported license. Caption by Kenneth A. Koenigshofer, PhD).

A fascinating research enterprise contributing greatly to our understanding of memory is the case of H.M. **H.M.** had **bilateral removal** of much of his **medial temporal lobes** including both of his hippocampi (remember that the hippocampus is located deep beneath the temporal lobe cortex), surgery performed in order to control his epileptic seizures (which were originating from the temporal lobe). Up until his death in December of 2008 of natural causes at age 82, H.M. still believed that Eisenhower was President of the United States. He knew almost nothing about anything that had happened after his brain surgery. According to a researcher working with H.M. with whom I spoke, before H.M. died whenever he was asked to describe his physical appearance, even years after his surgery, H.M. continued to describe himself as a young man with a thick head of wavy black hair, even though, at his age, he had almost no hair at all. After the surgical removal of much of his medial temporal lobes, H.M. could remember most things that happened before his surgery, but he **could not remember any new facts or events** for more than a few minutes. His **old long-term episodic and semantic memory was fine**. His **short-term memory** was also **normal**. What he **could not** do was to consolidate or **fix any new information about facts or events into** his **long-term memory**. Yet, his **implicit memory and his ability to consolidate new procedural memories into permanent storage was normal.** He learned to play computer games with great skill, but never was able to remember having learned them, and he was often puzzled about why he was so good when he couldn't even remember having seen computers or computer games before.

I sometimes speculate about what it must have been like for H.M. when he got up in the morning and looked in the mirror, expecting to see himself as a young man and finding an old man instead. Or I wonder when awoke in the morning and looked over at his wife lying next to him. He can only remember her as she looked as a much younger woman. What must he have experienced when he saw the older woman asleep next to him? Did he know who she was? He must have been puzzled about what could have happened to her overnight (younger and prettier, and now much older and perhaps not so pretty). Soon, however, he would forget what he had just seen minutes before. The life of H.M. must have been truly extraordinary. Studies of H.M. for many decades before his death in 2008 revealed much that we know about memory and the brain today, in particular the role of the **medial temporal lobe** structures in declarative (conscious) memory formation.



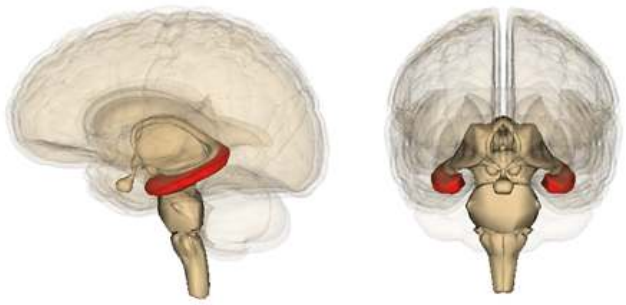


Figure **10.3.3**: (Left) Medial view, left human cerebral hemisphere highlighting Dentate (yellow) and Parahippocampal gyri (orange) within medial temporal lobe important in declarative memory. (Right) Hippocampus (red). (Images from Wikimedia Commons; (Left) Dentate gyrus; File:Cerebral Gyri - Medial Surface2.png; https://commons.wikimedia.org/wiki/F...l_Surface2.png; by John A Beal, PhD. Dep't. of Cellular Biology & Anatomy, Louisiana State University Health Sciences Center Shreveport; licensed under the Creative Commons Attribution 2.5 Generic license. (Right) Hippocampus; File:Hippocampus image.png; https://commons.wikimedia.org/wiki/F...mpus_image.png; by Life Science Databases(LSDB); licensed under the Creative Commons Attribution-Share Alike 2.1 Japan license. Caption by Kenneth A. Koenigshofer, PhD.).

10.3.8 The Engram

Many years earlier, Karl Lashley, a physiological psychologist, now well known in the history of psychology, investigated the physical basis of memory in the brain by teaching rats a maze and then destroying various parts and amounts of their brain tissue to see which areas were critical to memory. He was searching for the physical representation of memory in the brain. He referred to the physical changes in the brain, which occur when we learn and remember something, the **"engram," or "memory trace."**

His results were inconclusive, but he did find that the more cortex he destroyed the greater the disruption of memory. He called this the **"principle of mass action."** Today memory research continues and new insights have been discovered about the nature of the "engram," which appears to involve the **modification of synapses** (Howland & Wang, 2008), the **growth of new synapses**, and even the growth of **new neurons** (Deng et al., 2010; Ranganath & Hsieh, 2016). We discuss more about the physical changes in the brain that appear to be involved in learning and memory in later sections of this chapter.

10.3.9 Summary

Different types of memory have been designated by researchers as explicit and implicit, declarative, semantic and episodic, procedural or motor memory, and some forms of memory involve three stages, sensory, short-term, and long-term. Different types of memory and different stages involve different brain areas and processes. The case of H.M. led researchers to recognize the essential role of the hippocampus and other regions of the medial temporal lobes in the formation of long-term declarative memories, while the cerebellum stores motor programs of learned skilled movements and memory of at least some reflexes such as the eye-blink reflex. Mechanisms that seem to be the basis for learning and memory at the more microscopic level appear to depend on synaptic changes.

10.3.10 References

Baddeley, A. D. (1982). Implications of neuropsychological evidence for theories of normal memory. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *298* (1089), 59-72.

Bird, C. M., & Burgess, N. (2008). The hippocampus and memory: insights from spatial processing. *Nature Reviews Neuroscience*, *9* (3), 182-194.

Christian, K. M., & Thompson, R. F. (2003). Neural substrates of eyeblink conditioning: acquisition and retention. *Learning & memory*, *10* (6), 427-455.

Deng, W., Aimone, J. B., & Gage, F. H. (2010). New neurons and new memories: how does adult hippocampal neurogenesis affect learning and memory?. *Nature reviews neuroscience*, *11* (5), 339-350.

Howland, J. G., & Wang, Y. T. (2008). Synaptic plasticity in learning and memory: stress effects in the hippocampus. *Progress in brain research*, *169*, 145-158.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63* (2), 81.

Kim, J. J., & Thompson, R. E. (1997). Cerebellar circuits and synaptic mechanisms involved in classical eyeblink conditioning. *Trends in neurosciences*, *20* (4), 177-181.

Ranganath, C., & Hsieh, L. T. (2016). The hippocampus: a special place for time. *Annals of the new York Academy of Sciences*, *1369* (1), 93-110.

Squire, L. R., Stark, C. E., & Clark, R. E. (2004). The medial temporal lobe. *Annu. Rev. Neurosci.*, *27*, 279-306.

Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory consolidation. *Cold Spring Harbor perspectives in biology*, *7* (8), a021766.

Tanaka, K. Z., Pevzner, A., Hamidi, A. B., Nakazawa, Y., Graham, J., & Wiltgen, B. J. (2014). Cortical representations are reinstated by the hippocampus during memory retrieval. *Neuron*, *84* (2), 347-354.

Voss, J. L., Bridge, D. J., Cohen, N. J., & Walker, J. A. (2017). A closer look at the hippocampus and memory. *Trends in cognitive sciences*, *21* (8), 577-588.

10.3.10.1 Attributions

1. Section 10.3, "Introduction to Memory and the Brain," is original material written by Kenneth A. Koenigshofer, Ph.D., and is licensed under CC BY 4.0

This page titled 10.3: Introduction to Memory and the Brain is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

10.4: Synaptic Mechanisms of Learning and Memory

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

10.4.1 Learning Objectives

1. Describe Hebb's theory of the engram
2. Define Hebb's rule
3. Define Hebb synapses and cell assemblies
4. Describe Kandel's findings about the synaptic changes mediating habituation and sensitization in the sea slug, Aplysia
5. Explain the dual-trace theory of memory
6. Describe the role of the hippocampus in learning and memory
7. Describe the role of the cerebellum and amygdala in learning and memory
8. Discuss Long-term Potentiation (LTP) and Long-term depression and learning and memory
9. Describe changes in dendritic spines associated with learning and memory
10. Discuss how LTP and anatomical changes in synaptic spines might be related to Hebb's dual-trace theory

10.4.2 Overview

Donald Hebb's (1949) dual-trace theory of the physical basis of memory focused on changes at the synapse as the basis for learning and memory. This influential theory led to the discovery of changes in transmitter release associated with learning in the invertebrate nervous system of the sea slug. Other researchers found sustained increases in synaptic conductance caused by high frequency stimulation of the pre-synaptic neuron (Long-term Potentiation, LTP) as well as sustained decrease in synaptic strength known as Long-term Depression (LTD). Researchers have also discovered anatomical changes in dendritic spines associated with LTP and learning and memory, suggesting anatomical changes at synapses as a possible mechanism for learning and long-term memory at the synaptic level. Although early studies in search of the "engram", the physical basis for learning and memory, concluded that these functions were widely distributed throughout the brain, later research found that the hippocampus is critical for formation of long-term explicit memories and that the cerebellum is involved in implicit memory, whle the amygdala plays an essential role in emotional memories.

10.4.3 **The** Search for Learning and Memory in the Synapse

by Kenneth A. Koenigshofer, Ph.D.

Just as information is stored on digital media such as DVDs, hard drives, and flash drives, the information in our long-term memory must be physically stored in the brain. According to current theory, the ability to maintain information in long-term memory involves a gradual strengthening of the synaptic connections among neurons. When pathways in these neural networks are frequently and repeatedly activated, the synapses become more efficient permitting enhanced communication among neurons in the network, and these changes create memory (Saylor Foundation, 2012). As we will see, this view of the physical basis of memory was heavily influenced by the ideas of a Harvard biological psychologist named Donald Hebb.

Over a century ago a Russian physiologist named Ivan Pavlov proposed a theory of how learning occurred in the brain. As you recall from from module 10.2, Pavlov discovered classical conditioning when he observed that repeated pairings of a bell and food eventually led dogs to salivate at the sound of the bell, even when food was not present. To explain this phenomenon, Pavlov hypothesized that the conditioned stimulus (CS), a bell in this case, generated a locus of neural activity in sensory cortex (auditory cortex in this experiment) which radiated outward over the cortical surface. This was followed by a similar locus of cortical neural activity generated by the unconditioned stimulus (US) (meat powder in Pavlov's classic experiments) which also set up waves of radiating neural activity. Pavlov proposed that these two expanding fields of neural activity, originating from different areas of cortex, would intersect one another. According to Pavlov, this intersection of the cortical fields of neural activity, generated by the CS and US, formed the neurological basis for the association between CS and US.

This early theory of the physical basis of learning and memory was put to the test by experimental psychologist, Karl Lashley, of Harvard and then later, the Yerkes Laboratory of Primate Biology. Lashley made crisscross cuts throughout the cerebral cortex of rats and then attempted to condition them. He did this to disrupt any radiating neural activity in the cortex that might be present during conditioning, as Pavlov had proposed. Lashley found that the rats still could be conditioned and that they retained the conditioned response later in spite of the crisscross cuts over their entire cerebral cortex. This disproved Pavlov's theory that the physical basis of learning an association by classical conditioning was the intersection of waves of neural activity radiating outward from cortical loci.

Lashley, like Pavlov, was interested in finding the "engram." This term refers to the physical memory trace, the neural representation of memory in the brain. Toward this end, Lashley performed additional experiments in which he trained rats to navigate a maze and then destroyed different parts of their brains. He found that no matter where the brain damage was located, rats still retained some memory of the maze (Lashley, 1929, 1943, 1950). Lashley interpreted these results as evidence that it was the amount of cortical tissue removed, not its location, that determined the degree of impairment in learning and memory. He also hypothesized that memories were widely distributed throughout the brain and that therefore there was no particular area of the brain that was especially critical for memory formation.

Lashley proposed two principles derived from his research on the physical basis of learning and memory: 1) mass action (the amount of cortex destroyed determines degree of impairment in learning and memory) and 2) equipotentiality (any part of cortex within a functional area can take over the functions of any other part of that same functional area).

Lashley's conclusions had two primary impacts. First, his analysis "tended to discredit, or at least deemphasize, the role of the interior [subcortical] parts of the brain in learning and memory and, on the other, tended to localize the mechanisms of learning and memory within the confines of the cerebral cortex" (Thompson, 1974). Second, Lashley's principles of mass action and equipotentiality led to an **anti-localization bias** in the study of learning and memory, which discouraged attempts to find specific brain structures involved in these functions. However, consistent with current views, and **contradicting Lashley's findings**, Kaada, Rasmussen, and Kveim (1961) reported that **lesions of the hippocampo-fornix system impaired maze performance.** This was an early study that confirms what we know today, that a subcortical structure, the **hippocampus**, is critical to many forms of learning and memory.

10.4.3.1 Hebb's Rule, Cell Assemblies, and the Engram

After receiving his doctorate from Harvard in 1936, one of Lashley's graduate students, Donald Hebb, published *The Organization of Behavior: A Neuropsychological Theory*. In this influential book, Hebb proposed that the **"engram"** consisted of **changes at synapses** during learning and memory formation. A quote from the book (Hebb, 1949, p. 62) explains **Hebb's central principle** of synaptic change in learning and memory:

> “When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A’s efficiency, as one of the cells firing B, is increased."

Here Hebb is proposing an **increase in the strength of synaptic connections between neurons that consistently fire together** and that this involves growth or metabolic events in one or both of the cells.

More simply stated: "Neurons that fire together, wire together." This is known as "**Hebb's rule**" (retrieved from https://can-acn.org/donald-olding-hebb/ August 8, 2021).

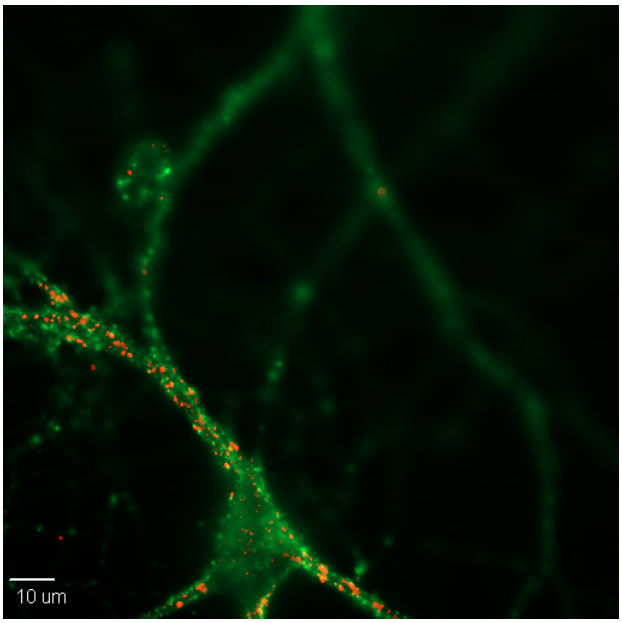


Figure 10.4.1: A 20X image of a cultured mouse cortical neuron in cell culture. Synapses are labeled for pre- (green) and post-synaptic (red) proteins, synaptophysin and PSD-95, respectively. (Image and caption from Wikimedia, Synapse; https://commons.wikimedia.org/wiki/F...n_Synapses.jpg; author, Dchordpdx; licensed under the Creative Commons Attribution 4.0 International license).

Neuroscientists use the term, **Hebbian synapses**, to refer to synapses that follow this principle. Hebb also proposed the idea of "cell assemblies." Hebb hypothesized that when cells repeatedly firing together, wired together, eventually they would form larger structures, "**cell assemblies**," which would form neural representations of whole, complex perceptions, ideas, memories, and other cognitive structures, such as schemas, and highly abstract categories and concepts. All of this could be built from changes in the efficiency of synaptic transmission at specific synapses, following Hebb's principle that "neurons that fire together, wire together." One can imagine that *entire cell assemblies* that fire together might also wire together, creating much larger and much more complex neural representations in the brain of complex, abstract ideas including such things as scientific theories and mathematical formulations.

10.4.3.1 Non-Association Learning in Aplysia

These ideas have inspired the hunt by neuroscientists for changes at synapses and in pre- or post-synaptic neurons, during and after learning--changes which might be the physical basis in the nervous system for learning and memory (Lashley's "engram," which, by the way, has nothing to do with the non-scientific use of the term by Scientology). Much of the pioneering scientific work on the search for the "engram" has involved looking for these physical changes associated with learning in the simple nervous system of a marine species, Aplysia californica, a type of sea slug. A major reason for the selection of this species for study, aside from the simplicity of its nervous system (just 20,000 neurons), is the large size of its neurons which makes the work and the observations easier.

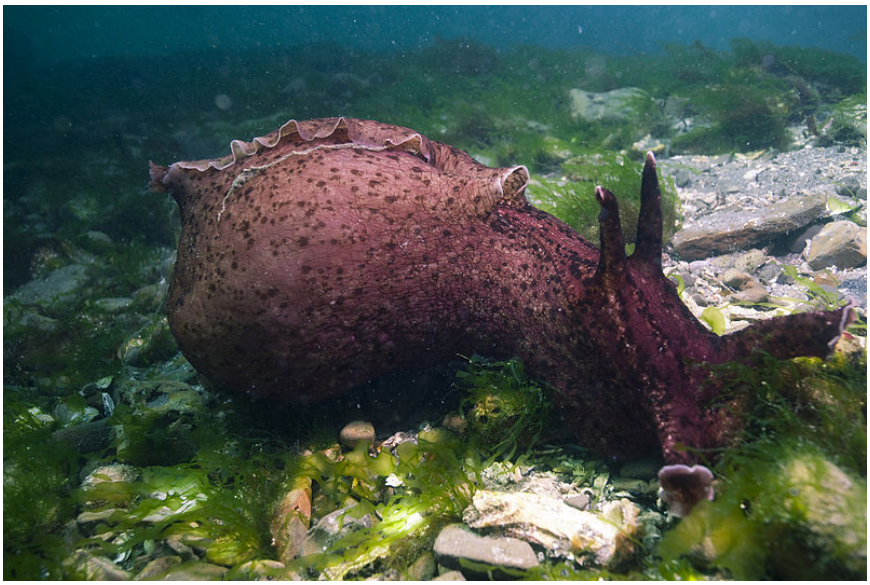


Figure **10.4.2**: Aplysia californica. (Image from Wikimedia, Aplysia californica; https://commons.wikimedia.org/wiki/F...,_Monterey.jpg; Chad King / NOAA MBNMS; this image is in the ***public domain*** because it contains materials that originally came from the U.S. *National Oceanic and Atmospheric Administration*, taken or made as part of an employee's official duties).

Eric Kandel (1976) and his colleagues did pioneering work which showed synaptic changes that mediate habituation (see sections above) of the Aplysia's siphon and gill defensive withdrawal reflex. A **decrease in conductivity** was found **at synapses** involved in the siphon and gill withdrawal reflex **in Aplysia after habituation** to a repeated presentation of a harmless novel stimulus. Kandel and co-workers showed that the **pre-synaptic sensory neuron released less neurotransmitter onto the post-synaptic motor neuron** as a result of repeated presentation of the novel stimulus leading to habituation of the reflex. The opposite was found for increased responsiveness. Sensitization refers to enhancement of responsiveness to a familiar stimulus. A single small electric shock to the tail of Aplysia heightens its gill withdrawal response for minutes to hours. Kandel and colleagues found that **sensitization** was mediated by an **increase in** release of **neurotransmitter from the pre-synaptic sensory neuron onto** the **post-synaptic motor neuron** serving the gill muscle. Thus, sensitization is the opposite of habituation both behaviorally and at the level of transmitter release (Kolb and Whishaw, 2001). Later studies showed synaptic changes involving **modified transmitter release** during **classical conditioning** as well (Kandel and Schwartz, 1982).

10.4.3.1 Hebb's Dual-Trace Theory of Memory

Hebb also proposed a **dual-trace theory of memory**--a **short-term dynamic electrical process for short-term memory** (Hebb referred to "reverberating circuits" holding this brief electrical activity) **and,** following sometime **later, an enduring structural change at synapses** as the physical basis (the engram) for long-term memory (Hebb, 1949). Decades of research followed using various agents, including electroconvulsive shock (ECS), shortly after learning in rats to disrupt the short-term memory trace, which according to Hebb, consisted of circulating or "reverberating" electrical activity, thus preventing long-term storage. However, although ECS in rats within about 30 seconds of learning seemed to disrupt consolidation of the short-term memory trace into a permanent long-term trace as evidenced by apparent amnesia for the learning task, additional research suggested alternative explanations for these results. One alternative was that the ECS did not prevent consolidation of the long-term memory trace, but instead disrupted retrieval of the memory. Evidence for this failure of retrieval hypothesis suggests that the original failure of consolidation interpretation of the ECS research was likely incorrect (Miller and Martin, 2014).

10.4.3.1 The Special Roles of the Hippocampus, Cerebellum, and Amygdala

As discussed above, research since Lashley has revealed that his speculation that all areas of brain were equally involved in learning and memory was incorrect. Now we know that one of the most important brain regions in **explicit memory** is the **hippocampus**, which serves as a preprocessor and elaborator of information (Squire, 1992). The hippocampus helps us encode information about **spatial relationships**, the **context** in which events were experienced, and the associations among memories

(Eichenbaum, 1999). The hippocampus also serves in part as a switching point that **holds the memory for a short time and then directs** the information **to other parts of the brain**, such as the **cortex**, to actually do the rehearsing, elaboration, and **long-term storage** (Jonides, Lacey, & Nee, 2005; Saylor Foundation, 2015). We also now know that different parts of the brain are involved in different kinds of memory.

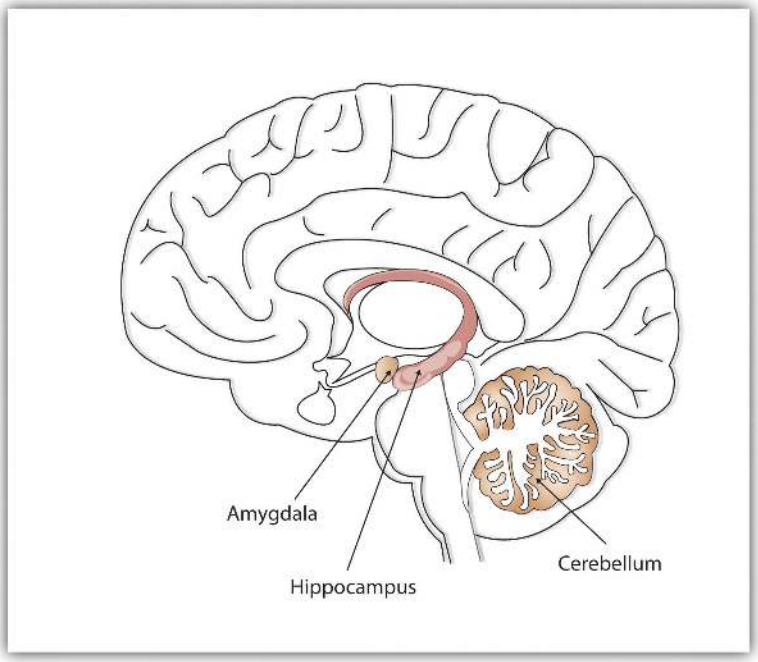


Figure **10.4.3**: Different brain structures help us remember different types of information. The hippocampus is particularly important in explicit memories, the cerebellum in implicit memories, and the amygdala in emotional memories (Image and caption from the Saylor Foundation, 2015; Remembering and Judging; https://learn.umgc.edu/d2l/le/conten.../25917628/View; licensed under a Creative Commons Attribution 3.0 Unported License.).

While the hippocampus is handling explicit memory, the **cerebellum** and the **amygdala** are critically involved in **implicit and emotional memories, respectively**. Research shows that the cerebellum is more active when we are learning associations and in priming tasks (based on implicit memory), and animals and humans with damage to the cerebellum are impaired in classical conditioning (Krupa, Thompson, & Thompson, 1993; Woodruff-Pak, Goldenberg, Downey-Lamb, Boyko, & Lemieux, 2000). The storage of many of our most important emotional memories, and particularly those related to fear, is initiated and controlled by the amygdala (Sigurdsson, Doyère, Cain, & LeDoux, 2007).

10.4.4 Changes at the Synapse Correlated with Learning and Memory

Hebb's theory that long-term memory was stored by physical changes at the synapse has been so influential that research on the neural basis of learning and memory has focused primarily on synaptic events (see Chapter 5 on synapses and synaptic transmission). Focus on the synapse was reinforced by the discovery in 1973 in rabbit hippocampus that a **long-lasting increase in synaptic conductivity (synaptic strength)** could be produced **by high frequency stimulation of** the **pre-synaptic neuron** (Bliss and Lomo, 1973). This finding that synaptic strength can be increased for an extended period of time following high frequency pre-synaptic stimulation is now referred to as **Long-Term Potentiation (LTP)**. LTP has been found in many species and in many parts of the brain. However, it has been studied most in the hippocampus of the rat. Since this early research, additional forms of synaptic change have also been discovered.

10.4.5 **Synaptic Plasticity**

Synaptic plasticity is the strengthening or weakening of synapses over time in response to increases or decreases in their activity. Plastic change also results from the alteration of the number of receptors located at a synapse. Synaptic plasticity is the basis of learning and memory, enabling a flexible, functioning nervous system. Synaptic plasticity can be either short-term (synaptic

enhancement or synaptic depression) or long-term. Two processes in particular, **long-term potentiation (LTP) and long-term depression (LTD)**, are important forms of synaptic plasticity that occur in **synapses in the hippocampus**.

10.4.6 Synaptic Plasticity: short-term enhancement, long-term potentiation and long-term depression

Key Points

- Short-term synaptic enhancement occurs when the amount of available neurotransmitter is increased, while short-term synaptic depression occurs when the amount of vesicles with neurotransmitters is decreased.
- Synapses are strengthened in long-term potentiation (LTP) when AMPA receptors (which bind to negatively-charged glutamate) are increased, allowing more calcium ions to enter the cell, causing a higher excitatory response.
- Long-term depression (LTD) occurs when the AMPA receptors are decreased, which decreases the amount of calcium ions entering the cell, weakening the synaptic response to the release of neurotransmitters.
- The strengthening and weakening of synapses over time controls learning and memory in the brain.

Key Terms

- **long-term potentiation**: a long-lasting (hours in vitro, weeks to months in vivo) increase, typically in amplitude, of the response of a postsynaptic neuron to a pattern of high frequency stimuli from a presynaptic neuron
- **long-term depression**: a long-term weakening of a synaptic connection
- **plasticity**: the property of neuron's conductivity or synaptic strength that allows it to be strengthened or weakened
- **NMDA receptor**: N-methyl-D-aspartate (NMDA) post-synaptic receptor and ion channel that is activated when Glutamate transmitter binds to it; an ionotropic (see Chapter 5) glutamate receptor
- **AMPA receptor**: Alpha-Amino-3-Hydroxy-5-Methyl-4-Isoxazole Propionic Acid (AMPA) post-synaptic receptor and ion channel that is activated when Glutamate transmitter binds to it; an ionotropic (see Chapter 5) glutamate receptor; both NMDA and AMPA receptors are important in learning and memory

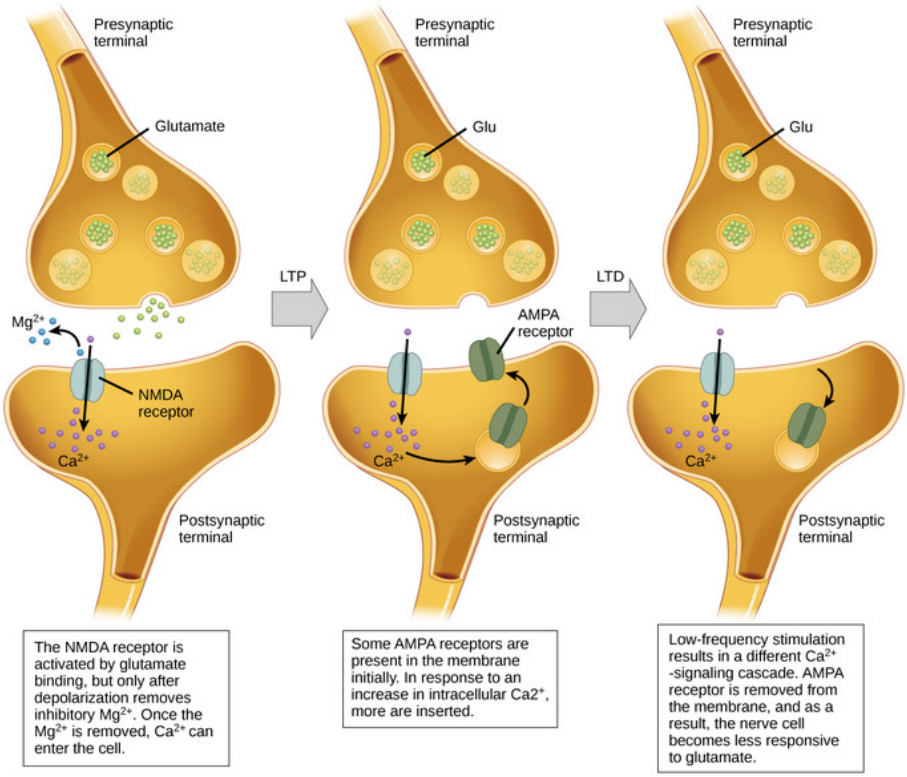


Figure **10.4.4**: **Long-term potentiation and depression**: Calcium entry through postsynaptic NMDA receptors can initiate two different forms of synaptic plasticity: long-term potentiation (LTP) and long-term depression (LTD). LTP arises when a single synapse is repeatedly stimulated. This stimulation causes a calcium- and CaMKII-dependent cellular cascade, which results in the

insertion of more AMPA receptors into the postsynaptic membrane. The next time glutamate is released from the presynaptic cell, it will bind to both NMDA and the newly-inserted AMPA receptors, thus depolarizing the membrane more efficiently. LTD occurs when few glutamate molecules bind to NMDA receptors at a synapse (due to a low firing rate of the presynaptic neuron). The calcium that does flow through NMDA receptors initiates a different calcineurin and protein phosphatase 1-dependent cascade, which results in the endocytosis of AMPA receptors. This makes the postsynaptic neuron less responsive to glutamate released from the presynaptic neuron. (Image and caption from Lumen Boundless Biology, How Neurons Communicate; https://courses.lumenlearning.com/bo...s-communicate/; curation and Revision. Provided by: Boundless.com. License: ***CC BY-SA: Attribution-ShareAlike***.)

10.4.6.1 Long-term Potentiation (LTP)

Long-term potentiation (LTP) is a persistent strengthening of a synaptic connection, which can last for minutes or hours or even weeks. LTP is based on the Hebbian principle: “cells that fire together wire together.” There are various mechanisms, none of which are fully understood, behind the synaptic strengthening seen with LTP.

One known mechanism involves a type of postsynaptic glutamate receptor: NMDA (N-Methyl-D-aspartate) receptors. These receptors are normally blocked by magnesium ions. However, when the postsynaptic neuron is depolarized by multiple presynaptic inputs in quick succession (either from one neuron or multiple neurons), the magnesium ions are forced out and Ca^{2+} ions pass into the postsynaptic cell. Next, Ca^{2+} ions entering the cell initiate a signaling cascade that causes a different type of glutamate receptor, AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) receptors, to be inserted into the postsynaptic membrane. Activated AMPA receptors allow positive ions to enter the cell.

Therefore, the next time glutamate is released from the presynaptic membrane, it will have a larger excitatory effect (EPSP) on the postsynaptic cell because the binding of glutamate to these AMPA receptors will allow more positive ions into the cell. The insertion of additional AMPA receptors strengthens the synapse so that the postsynaptic neuron is more likely to fire in response to presynaptic neurotransmitter release.

LTP has many similarities with the synaptic changes Hebb proposed as basis for long-term memory, including two key features. First, LTP is long lasting, up to a year with repeated trials. Second, many forms of LTP require simultaneous activation in pre-synaptic and post-synaptic neurons at the synapse where LTP takes place ("neurons that fire together, wire together"); this is because NMDA (N-methyl-d-aspartate) receptors for glutamate (the most common excitatory transmitter in the brain), which are prominent at synapses where LTP occurs, have the same requirement for simultaneous activity in both pre- and post-synaptic neurons in order for these receptors to become activated. Research showing that many features of LTP are similar to features of long-term memory provides strong circumstantial evidence that LTP is related to the mechanisms of learning and memory. For example, LTP can be stimulated by low intensity stimulation similar to that produced in single neurons; LTP is most prominent in structures associated with learning; LTP is produced in the hippocampus by learning; drugs that enhance or impair learning also enhance or impair LTP; and LTP occurs in the nervous system of simple invertebrates (see discussion of Aplysia above) at the specific synapses involved in the learning (Pinel & Barnes, 2021).

10.4.6.2 Long-term Depression (LTD)

Another related phenomenon is **long-term depression (LTD)**, associated with decreases in synaptic conductivity. LTD might be part of the processes involved in creating and modifying patterns of excitation and inhibition in large populations of neurons for the coding of learned movements, sensory experience, and perhaps the mental representations of complex cognitive structures such as perceptions, whole memories, concepts, and even abstract ideas (Churchland, 2013). Coding of mental events in terms of patterns of excitation and inhibition in large populations of neurons is discussed later in this chapter when we examine neural network modeling of learning and memory. A related possibility is the pruning away of some synapses to form permanent changes in neural circuitry.Long-term depression (LTD) is essentially the reverse of LTP: it is a long-term weakening of a synaptic connection. One mechanism known to cause LTD also involves AMPA receptors. In this situation, calcium that enters through NMDA receptors initiates a different signaling cascade, which results in the removal of AMPA receptors from the postsynaptic membrane. With the decrease in AMPA receptors in the membrane, the postsynaptic neuron is less responsive to the glutamate released from the presynaptic neuron. While it may seem counterintuitive, LTD may be just as important for learning and memory as LTP. The

weakening and pruning of unused synapses trims unimportant connections, leaving only the salient connections strengthened by long-term potentiation.

10.4.6.3 **Short-term Synaptic Enhancement and Depression**

Short-term synaptic plasticity acts on a timescale of tens of milliseconds to a few minutes. Short-term synaptic enhancement results from more synaptic terminals releasing transmitters in response to presynaptic action potentials. Synapses will strengthen for a short time because of either an increase in size of the readily- releasable pool of packaged transmitter or an increase in the amount of packaged transmitter released in response to each action potential. Depletion of these readily-releasable vesicles causes synaptic fatigue. Short-term synaptic depression can also arise from post-synaptic processes and from feedback activation of presynaptic receptors.

10.4.7 Changes at the Synapse Correlated with Learning and Memory: Anatomical Changes in Dendritic Spines

Modifications in synaptic strengths in excitatory synapses in the hippocampus appear to play a critical role in storage and recall of information in mammals. In humans at least, the role of the hippocampus is especially critical in explicit episodic (autobiographical) memory. As noted above, it is also crucially involved in spatial memory for many mammal species, including humans. Changes in synaptic strength throughout much of the brain may be important in learning and memory.

10.4.7.1 Dendritic Spines

Dendritic spines are tiny protruding structures located on the shaft of dendrites and are associated with synaptic connections. These spines "are present in large numbers on the surface of dendrites. For example, a single pyramidal neuron in the hippocampal CA1 region possesses as many as 30,000 dendritic spines. A majority of excitatory synapses are formed on the surface of these dendritic spines" (Irie & Yamaguchi, 2009, p. 1141). The primary sites of excitatory synaptic interaction in the mammalian central nervous system appear to be at dendritic spines.

These spiny protrusions have a head and neck of varied morphology (see Figures 10.4.5 and 10.4.6). Changes in spines play a significant role during brain development. "In a developing brain, spines exhibit a high degree of structural and functional plasticity, reflecting the formation and elimination of synapses during the maturation of neuronal circuits. The morphology of spines in developing neurons is affected by synaptic activity, hence contributing to the experience-dependent refinement of neuronal circuits, learning, and memory. Thus, understanding spine dynamics and its regulation is of central importance to studies of synaptic plasticity in the brain" (Bertling, et al., 2012, p. 391). During brain development synapses are formed, modified and sometimes eliminated as a function of input to them.

As discussed above, LTP has been shown to lead to functional changes at synapses where it is induced. These changes due to LTP are increased synaptic conductivity and enhanced responsiveness of the post-synaptic neuron. It appears that anatomical changes in spines accompany these changes in synaptic strength which occur during LTP and learning and memory.

Remember Hebb's proposal that long-term memory storage involves some persistent anatomical change at synapses. If Hebb was right, then we should expect to see anatomical changes at synapses when learning and memory occur. And if LTP is involved in the formation of these hypothesized anatomical changes, then we should expect to see anatomical modifications at the synapse associated with the LTP-induced increases in synaptic conductivity and responsiveness of the post-synaptic neuron. Consistent with Hebb's hypothesis about the physical basis of learning and memory, this is just what neuroscientists have observed. Changes in synaptic strength, associated with LTP, are accompanied by LTP-induced alterations of the shape and size of dendritic spines (Chidambaram, et al., 2019; Harris, et al., 2003), anatomical changes at the synapse, just as Hebb predicted.

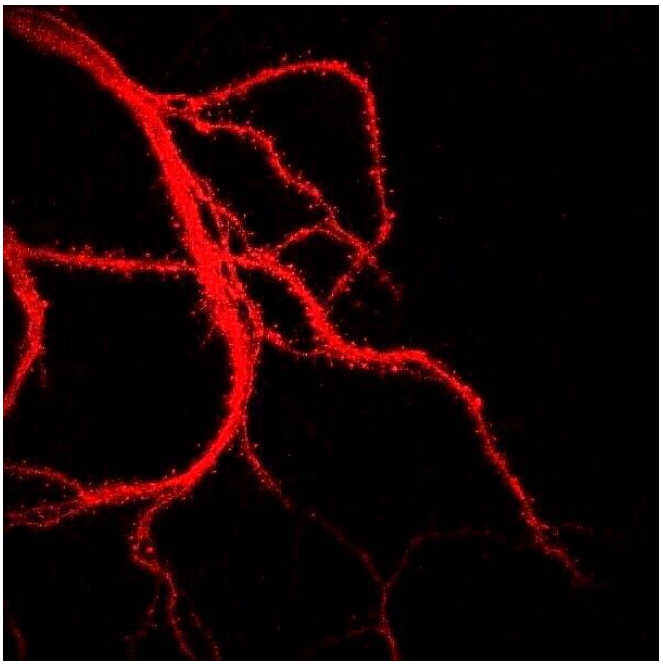


Figure 10.4.5: Branching dendrites of a neuron showing dendritic spines (tiny bristle-like projections lining each dendritic branch). The photograph was obtained with a laser scanning microscope. Dendritic spines can rapidly change in size and shape and numbers, and are important in learning and memory. (Caption by Kenneth A. Koenigshofer, PhD. Image from Wikimedia Commons; File:Нейрональные отростки с шипиками.jpg; https://commons.wikimedia.org/wiki/F...0%BC%D0%B8.jpg; by Sergb95; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

Evidence strongly suggests that new protein synthesis is necessary for these effects on the shape and size of dendritic spines (Chidambaram, et al., 2019). As noted above, both growth and metabolic changes at synapses were proposed by Hebb as possible basis for the long-term memory. These anatomical changes in dendritic spines induced by LTP can be long-lasting. For example, "In mature networks, synaptic connections at dendritic spines can be quite stable, as newly emergent spines generated after motor learning have been shown to persist for months. . . While the distribution of spine size across the dendritic arbor of a single neuron can be quite variable, spine size generally correlates with excitatory synapse strength . . . it is generally accepted that spine head diameter and synapse strength co-vary during the expression of long term potentiation (LTP) . . ." (Henry, et al., 2017). It appears that Hebb's hypothesis about the physical basis of learning and memory in the brain was quite predictive of what later research would show--one mark of a good scientific theory.

Consistent with Hebb's hypothesis, dendritic spines appear to function as storage sites for synaptic strength and help transmit electrical potentials to the neuron's cell body. Dendrites of a single neuron can contain hundreds, even thousands of spines. Dendritic spines usually receive excitatory input from axons, although sometimes both inhibitory and excitatory connections are made onto the same dendritic spine (Kasthuri, et al., 2015). In addition to spines providing a potential anatomical substrate for memory storage and synaptic transmission, they may also increase the number of possible contacts between neurons (Alvarez & Sabatini, 2007). Spines are found on the dendrites of most principal neurons in the brain, including the pyramidal neurons of the neocortex (in prefrontal cortex for cognition; corticospinal tract for movement), the medium spiny neurons (GABAergic inhibitory) of the striatum (striate nucleus of basal ganglia; motor and reward systems; receives Glutamatergic and Dopaminergic inputs), and the Purkinje cells (GABAergic inhibitory neurons) of the cerebellum which receive relatively weaker excitatory (Glutamatergic) synapses to spines on the Purkinje cell dendrite. Hippocampal and cortical pyramidal neurons may receive tens of thousands of mostly excitatory inputs from other neurons onto their spines, whereas the number of spines on Purkinje neuron dendrites in the cerebellum is even greater, up to ten times greater.

Three of the four most notable classes of spine shape are shown in the figure below: "thin", "stubby", and "mushroom." The fourth category, "branched", is not shown. Studies using electron microscopy have revealed a continuum of shapes between these categories (Ofer, et al., 2021). The variable spine shape and volume is correlated with the strength and maturity of each spine-synapse. "Dendritic spines are also highly motile undergoing changes in size and shape over a timescale of seconds to minutes.

Because small, thin dendritic spines are most likely to undergo these structural changes, whereas large, so-called ‘mushroom’ spines tend to maintain their form, it has been suggested that mushroom spines are more stable memory spines whereas the more plastic thin spines are learning spines. . . . [L]arge spines are sites of strong synapses and, accordingly, the growth of the spine head likely correlates with a strengthening of synaptic transmission" (Leuner & Shors, 2010, p. 349).

Both shorter-term transient and longer-term sustained changes in structural plasticity of dendritic spines have been observed by researchers--within the first two minutes following stimulation of the pre-synaptic neuron, an initial 300% expansion of the post-synaptic dendritic spine occurs, followed by a reduction of the spine's volume so that it is elevated to about 70-80% larger than the original pre-stimulation volume, a sustained change in structural plasticity lasting about 30 minutes or more. LTP is initiated with the transient stage of dendritic spine growth. Both stages of spine growth are hypothesized to be involved in learning and memory. Large spines are more stable than smaller ones and may be resistant to modification by additional synaptic activity (Kasai, et al., 2003). These structural changes in dendritic spines induced by pre-synaptic stimulation associated with LTP are believed by neurosciences to be important in the encoding, storage, and retrieval of memories (Murakoshi, et al., 2011). These changes in spine size and shape involve changes in the microstructure of the spine and abnormalities at this level may be involved in memory disorders. "The actin cytoskeleton is the structural element underlying changes in dendritic spine morphology and synapse strength. The proper morphology of spines and proper regulation of the actin cytoskeleton have been shown to be important in memory and learning; defects in regulation lead to various memory disorders. Thus, understanding actin cytoskeleton regulation in dendritic spines is of central importance to studies of synaptic and neuronal function" (Koskinen, et al., 2012, p. 47).

The morphogenesis of dendritic spines (the process by which they attain their shape) is critical to the induction of long-term potentiation (LTP) (Kim & Lisman, 1999; Krucker, et al., 2000). Interestingly, Hayashi-Takagi et al. (2015) found that memory could be disrupted if the potentiated spines, within a neuron ensemble involved in motor learning, were specifically shrunk. This suggests that spine growth is essential to at least some forms of memory.

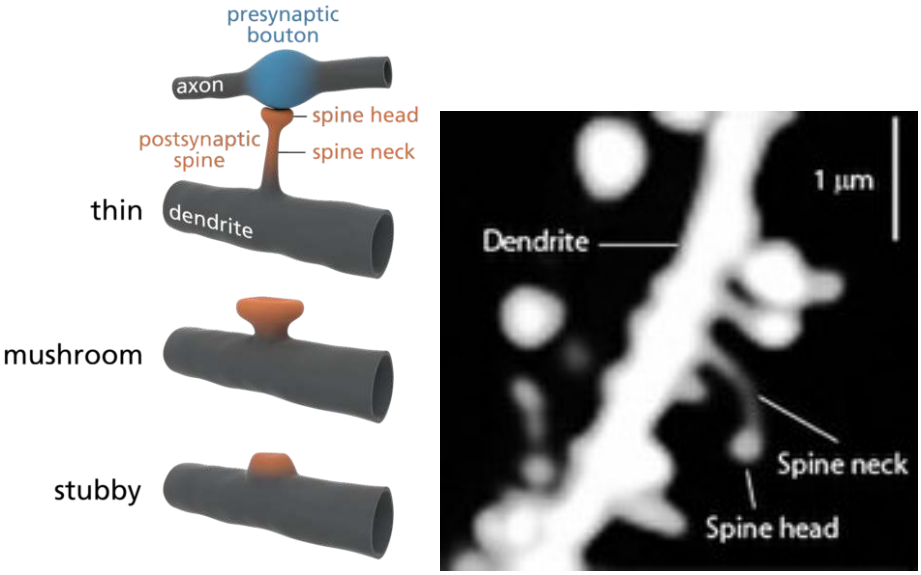


Figure **10.4.6**: (Left) Varying morphology of types of dendritic spines. Changes in their shape and size are associated with LTP and learning and memory. (Right) Spines on the dendrite of a medium spiny striatal neuron. The image was obtained by expressing Enhanced Green Fluorescent Protein (EGFP) in the neurons and imaging them using a laser scanning two photon microscope. (Caption for image on left by Kenneth A. Koenigshofer, PhD; Image on left from Wikimedia Commons, Dendritic spines; https://commons.wikimedia.org/wiki/F...e_types_3D.png; original work of Thomas Splettstoesser (www.scistyle.com); licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license. Image on right and its caption from Wikimedia Commons, Dendritic spines; https://commons.wikimedia.org/w/inde...=Go&type=image; Released into Public Domain by the image author).

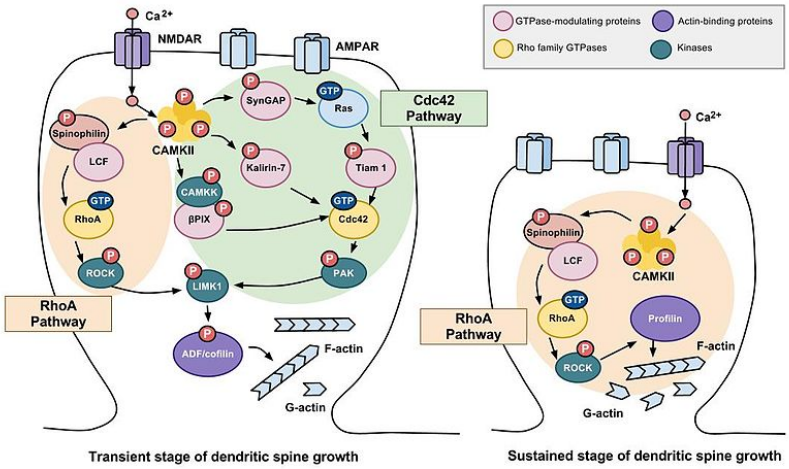


Figure 10.4.7: Transient vs. Sustained Dendritic Spine Growth following High-Frequency Stimulation. Details inside each spine show the chemical cascades involved in the two stages of spine growth after pre-synaptic stimulation. It is not necessary for the student to know the details of these cascades. However, note that these cascades begin with calcium influx through NMDA receptors (top of spine) and that the final step is action on actin (bottom right of spine) leading to changes in shape and size of the spine. (Image from Wikimedia Commons; Dendritic spines; https://commons.wikimedia.org/wiki/F...timulation.jpg; author of image and image title, Itzy; licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license; caption by Kenneth A. Koenigshofer, PhD.).

Yang, et al. (2009) studied changes in numbers of synaptic spines during learning and novel sensory experience in mice. They state that their "results suggest that new experience leads to the pruning of existing synapses and could cause significant functional changes in cortical circuits. Indeed, we found that 1 week after motor training, motor performance strongly correlated with the degree of spine elimination . . . Thus, motor learning and novel sensory experience involve not only new spine formation but also permanent removal of connections established early in life" (Yang, et al. 2009, p. 921). These results are easier to understand if we recall that during prenatal development of the brain there are excess numbers of synaptic connections and that the elimination of nonfunctional synapses is essential to normal brain development and functioning. Learning may involve neural pruning as well as the formation and enhanced conductivity of synapses via changes in dendritic spines. As Yang, et al., note, "In addition to promoting synapse formation, experience plays an important role in eliminating excessive and imprecise synaptic connections formed early during development" (Yang, et al. 2009, p. 921).

Chidambaram, et al. (2019, p.161) summarize the relationship of changes in synaptic conductivity to changes in dendritic spines: "During synaptic plasticity the number and shapes of dendritic spines undergo radical reorganizations. Long-term potentiation (LTP) induction promotes spine head enlargement and the formation and stabilization of new spines. Long-term depression (LTD) results in their shrinkage and retraction." These authors also note some interesting observations relating atypical numbers of synaptic spines to several neurologically based disorders. "Reports indicate increased spine density in the pyramidal neurons of autism and Fragile X syndrome patients and reduced density in the temporal gyrus loci of schizophrenic patients. Post-mortem reports of Alzheimer's brains showed reduced spine number in the hippocampus and cortex" and atypical spines may play a role in neurodegenerative diseases (Chidambaram, et al., 2019, p. 161).

If changes in spines are involved in the formation and retention of long-term memories, then morphological changes in spines must be very durable. According to Yang and colleagues (2009, p. 920), "Stably maintained dendritic spines are associated with lifelong memories. . . . a small fraction of new spines induced by novel experience, together with most spines formed early during development and surviving experience-dependent elimination, are preserved throughout the entire life of an animal. These studies indicate that learning and daily sensory experience leave minute but permanent marks on cortical connections and suggest that lifelong memories are stored in largely stably connected synaptic networks."

Hebb's theory has proven to be impressively predictive of discoveries made many years after his book was published in 1949. Findings like those described in this module continue to confirm many aspects of Hebb's theory.

10.4.8 Cellular and Receptor Level Mechanisms Revisited

In the sections above, we discussed **Long-Term Potentiation (LTP)** and its possible involvement in memory processes. Here, it is appropriate to consider LTP and associated processes in greater detail. As noted previously, Long-term potentiation (LTP) is a process in which synapses are strengthened. It has been the subject of extensive research since the mid-1970s because of its likely role in several types of memory.

A related phenomenon also discussed above is Long-Term Depression **(LTD**), a phenomenon that is the opposite of long-term potentiation (LTP). In LTD, communication across the synapse is silenced. LTD plays an important role in the cerebellum, in implicit procedural memory (motor memory, implicit memory of how to do something like ride a bicycle or hit a baseball), where the neural networks involved in erroneous movements are inhibited by the silencing of their synaptic connections. LTD is what allows us to correct our motor procedures when learning how to perform a task. Thus, long-term depression (LTD) involves a weakening of a synapse as a means of improving performance of learned skilled movements (recall that weakening of synaptic strength is associated with shrinkage of dendritic spines). Imagine that you are trying to learn a gymnastics routine that involves a number of flips and other complicated movements in sequence. When you initially start to learn the routine, you make many incorrect moves that you must suppress if you are going to learn to do the routine correctly with the error free precision that the judges will look for. LTD is the cellular level mechanism that allows you to suppress the incorrect muscle movements that lead to errors in the routine.

As discussed previously, there is another possible role for LTD in coding of perceptions, concepts, ideas, and other complex cognitive structures. One of the most difficult problems in neuroscience is how the activities of large populations of neurons encode and create complex elements of psychological experience such as perceptions, thoughts, ideas, understanding, knowledge, and other complex mental representations. Modeling of brain processes in artifical neural networks (see the supplement later in this chapter on artificial networks) suggests that inhibition is just as important as excitation in neural coding of complex mental representations. Patterns of excitation and inhibition in large populations of neurons may underlie the neural coding of complex cognition and perception. If so, then LTD may be equally as important as LTP in the neural coding of complex representations such as perceptions, thoughts, concepts, ideas and even emotions (see *Matter and Consciousness* by Paul Churchland for a more complete discussion of these ideas). Just as dots and dashes are equally important in Morse Code, excitation and inhibition may be equally important elements in neural coding.

Although a synapse is weakened in LTD , by contrast, in LTP, after intense stimulation of the pre-synaptic neuron, a synapse is strengthened and the amplitude of the post-synaptic neuron's response increases. The stimulus applied to the pre-synaptic neuron to produce LTP is generally of short duration (less than 1 second) but high frequency (over 100 Hz). The excitatory potential (EPSP; see Chapter 5) measured in a post-synaptic neuron, after high frequency pre-synaptic stimulation, is increased for a long period. For example, when the axons that make connections to the pyramidal neurons of the hippocampus are exposed to a high-frequency stimulus, the amplitude of the excitatory potential measured in these pyramidal neurons is increased for up to several weeks. What synaptic events are involved in the production of LTP? Let's review, now in greater detail.

Glutamate, the neurotransmitter released into these hippocampal synapses (glutamate is excitatory and mediates fast synaptic neurotransmission in the brain), binds to several different sub-types of receptors on the post-synaptic hippocampal neuron. Two of these glutamate receptor sub-types, the receptors for AMPA and NMDA, are especially important for LTP (Traynelis, et al., 2010).

The AMPA receptor (a glutamate receptor named after the derivative that activates it, AMPA short for Alpha-Amino-3-Hydroxy-5-Methyl-4-Isoxazole Propionic Acid) is paired with an ion channel so that when glutamate binds to this receptor, this channel lets sodium ions enter the post-synaptic neuron. This influx of sodium causes the post-synaptic dendrite to become locally depolarized (an EPSP, a positive shift in voltage).

The **NMDA receptor**, which also uses glutamate as its transmitter, is also paired with an ion channel (NMDA is N-methyl-D-aspartate, a derivative of glutamate). This channel admits calcium ions into the post-synaptic cell when it is activated. However, when the cell is at resting potential, the calcium channel is blocked by magnesium ions (Mg^{2+}), so that even if glutamate binds to the receptor, calcium cannot enter the neuron. For these magnesium ions to withdraw from the channel, the dendrite's membrane potential must be depolarized. And that is exactly what happens during the high-frequency stimulation that causes LTP: the post-synaptic neuron becomes depolarized following the sustained activation of its AMPA receptors! The magnesium then withdraws from the NMDA receptors and allows large numbers of calcium ions to enter the cell.

This increased concentration of calcium in the dendrite sets off several biochemical reactions that make this synapse more efficient for an extended period (Bliss & Collingridge, 1993; Bliss, et al., 2018; Citri & Malenka, 2008). These calcium ions are extremely important intracellular messengers that activate many enzymes by altering their conformation. One of these enzymes is calmoduline, which becomes active when four calcium ions bind to it. It then becomes Ca2+/calmodulin, the main second messenger for LTP. Ca2+/calmodulin then in turn activates other enzymes that play key roles in this process, such as **adenylate cyclase** and Ca2+/calmodulin-dependent protein kinase II **(CaM kinase II)**. These enzymes in turn modify the spatial conformation of other molecules, usually by adding a phosphate ion to them. This common catalytic process is called **phosphorylation**.

The activated adenylate cyclase manufactures cyclic adenosine mono-phosphate **(cAMP)**, which in turn catalyzes the activity of another protein, kinase A (**PKA**). In short, there is a typical cascade of biochemical reactions which can have many different effects.

For example, PKA phosphorylates the AMPA receptors, allowing them to remain open longer after glutamate binds to them. As a result, the post-synaptic neuron becomes further depolarized, thus contributing to LTP.

Other experiments have shown that **CREB** protein is another target of PKA, protein kinase A. CREB plays a major role in gene transcription, and its activation leads to the creation of new AMPA receptors that can increase synaptic efficiency still further.

According to Park, et al. (2021), Long-term potentiation (LTP) at hippocampal CA1 synapses (see Figure 10.4.8) can be expressed by an increase either in the number (N) of AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptors or in their single channel conductance (γ)--that is, increased synaptic strength. In their experiments, they established how these distinct synaptic processes contribute to the expression of LTP in hippocampal slices obtained from young adult rodents. LTP induced by compressed theta burst stimulation (TBS), with a 10 second inter-episode interval, involves purely an increase in the number of AMPA receptors (LTP_N). In contrast, either a spaced TBS, with a 10 min inter-episode interval, or a single TBS, delivered when PKA is activated, results in LTP that is associated with a transient increase in single channel conductance (γ) or increased synaptic strength (LTP_γ), caused by the insertion of calcium-permeable (CP)-AMPA receptors. Activation of CaMKII is necessary and sufficient for LTP_N (increase in numbers of AMPA receptors, while PKA is additionally required for LTP_γ (increase in single channel conductance). Thus, two mechanistically distinct forms of LTP co-exist at these synapses.

The other enzyme activated by Ca2+/calmodulin, CaM kinase II, has a property that is decisive for the persistence of LTP: it can phosphorylate itself! Its enzymatic activity continues long after the calcium has been evacuated from the cell and the Ca2+/calmodulin has been deactivated.

CaM kinase II can then in turn phosphorylate the AMPA receptors and probably other proteins such as MAP kinases, which are involved in the building of dendrites, or the NMDA receptors themselves, whose calcium conductance would be increased by this phosphorylation.

LTP involves at least two phases: **establishment** (or **induction**), which lasts about an hour, and **maintenance** (or **expression**), which may persist for several days.	The first phase, **establishment**, can be experimentally induced by a single, high-frequency stimulation. It involves the activity of various enzymes (kinases) that persist after the calcium is eliminated, but no protein synthesis occurs.	To trigger the **maintenance** phase, however, a series of high-frequency stimuli must be applied. Unlike the establishment phase of LTP, the maintenance phase requires the synthesis of new proteins–for example, the ones that form the receptors and the ones that contribute to the growth of new synapses (another phenomenon that occurs during the maintenance phase of LTP).

Table 10.4.1: Establishment and Maintenance Phases of LTP

To let the calcium enter the cell, the **NMDA receptor must be activated by glutamate and subjected to depolarization simultaneously.** The necessity for **these two simultaneous conditions gives this receptor associative properties**. This lets it **detect the coincidence of two events** and makes it the **key element in long-term potentiation.**

Significantly, if this receptor is blocked with a drug, or if the gene involved in its construction is disabled, LTP cannot occur.

The spines on post-synaptic dendrites form separate compartments to isolate biochemical reactions that occur at some synapses but not at others. This anatomical specialization probably helps to ensure a certain specificity in neural connections.

The most interesting characteristic of **LTP** is that it c**an cause the long-term strengthening of the synapses between two neurons that are activated simultaneously.** In other words, exactly the kind of **association mechanism** that **Hebb** (1949) had imagined years earlier.

10.4.8.1 Experimental Evidence for Components of LTP

Many methods have been used to determine the role of a particular ion, or a second messenger, or an enzyme in a synaptic process.

For example, the role of calcium in long-term potentiation (LTP) has been confirmed in a number of ways. One experiment involved injecting the post-synaptic neuron with chelating agents such as EGTA and BAPTA, two molecules that bind to calcium and render it inactive. As a result, it becomes impossible to induce LTP. The reverse procedure has also been used. Researchers have injected special molecules into the post-synaptic neuron and then illuminated them with UV light, thus causing them to release enough calcium to induce LTP in this neuron.

Another approach is to produce mutations that make a protein non-functional or that block its action with another molecule. Blocking CaM kinase II in this way prevents LTP from becoming established, which also demonstrates the central role of this protein.

Similarly, inhibiting PKA or CREB prevents LTP from reaching its second phase and being sustained.

In certain cases, researchers have even identified the one amino acid, among the hundreds of amino acids that make up a protein, whose phosphorylation is essential for LTP (in case you want to know: Ser831 for the GluR1 sub-unit of the AMPA receptor and Thr286 for the autophosphorylation site of CaM kinase II).

Researchers have also shown that a mouse whose gene for the GluR1 sub-unit of the AMPA receptor had been knocked out could not have any LTP, thus confirming the role of CaM kinase II once again.

In mice for which the site Thr286 in CaM kinase II was deactivated, their basic synaptic transmission was maintained, but LTP could no longer be induced in them, thus proving the need for CaM kinase II. The reverse procedure also led to the same conclusion: adding activated CaM kinase II to the pyramidal neurons of the hippocampus causes a potentiation phenomenon similar to LTP.

10.4.8.2 Additional Mechanisms of LTP

“Silent synapses” are another mechanism that was discovered in the mid-1990s and that may contribute to long-term potentiation (LTP). These synapses are physically present, but under normal conditions do not contribute to synaptic transmission.

Some of these silent synapses have been found in the hippocampus. They appear to have receptors for NMDA but not for AMPA. It is thought that these synapses may be activated during LTP and thus help to strengthen the synaptic response. The discovery that after LTP, these synapses do display an electrical current associated with AMPA channels suggests that some newly synthesized AMPA receptors may be inserted into the post-synaptic membrane.

In addition to all of the post-synaptic mechanisms involved in the **establishment of LTP**, it has long been postulated that some pre-synaptic modifications occur during the ensuing **maintenance phase**. But certain modifications, such as an increase in the amount of glutamate released by the pre-synaptic neuron, would imply the presence of a retrograde messenger that goes back to this neuron and modifies it. Because nitric oxide (**NO**) is a gas in its natural state, and can thus diffuse through cell membranes, it would be an ideal candidate for this role. But its involvement is still the subject of much debate and controversy.

10.4.8.3 Video Reviews of LTP and Memory

For a review, take a look at these brief videos of the events in LTP:

https://www.youtube.com/watch?v=KyQUBukwwO8

https://www.youtube.com/watch?v=vso9jgfpI_c

https://www.youtube.com/watch?v=-mHgPfXHzJE

10.4.9 LTP, Neurochemical Cascades, and Stages of Memory

According to Rozensweig (2007), different parts of the neurochemical cascade, associated with learning and memory, can be related to different stages in memory processing. Bennett, et al. (1964), made an early discovery that enriched experience in rats causes increased rates of protein synthesis and increased amounts of protein in the cortex. Mizumori, et al. (1985), using the protein-synthesis inhibitor, anisomycin, found that protein must be synthesized in the cortex soon after training if LTM is to be formed; however, short-term memory (STM) did not require protein synthesis, findings consistent with the two kinds of memory traces that Hebb (1949) had proposed: transient, labile memory traces on the one hand and stable structural traces on the other. Using chicks, several investigators traced a cascade of neurochemical events from initial sensory stimulation to synthesis of protein and structural changes in the brain (Rose, 1992). Rozensweig (2007) summarizes some of these events as follows:

"The cascade is initiated when sensory stimulation activates receptor organs that stimulate afferent neurons by using various synaptic transmitter agents such as acetylcholine (ACh) and glutamate. Inhibitors of ACh synaptic activity such as scopolamine and pirenzepine can prevent STM as can inhibitors of glutamate receptors including both the NMDA and AMPA receptors. Alteration of regulation of ion channels in the neuronal membrane can inhibit STM formation, as seen in effects of lanthanum chloride on calcium channels and of ouabain on sodium and potassium channels. Inhibition of second messengers is also amnestic, for example, inhibition of adenylate cyclase by forskolin or of diacylglycerol by bradykinin. These second messengers can activate protein kinases — enzymes that catalyze additions of phosphate molecules to proteins. We found that two kinds of protein kinases are important in formation, respectively, of ITM (an intermediate stage in memory formation noted by Rozenzweig, 2007) or LTM. Agents that inhibit calcium/calmodulin protein kinases (CaM kinases) prevent formation of ITM, whereas agents that do not inhibit CaM kinases, but do inhibit protein kinase A (PKA) or protein kinase C (PKC) prevent formation of LTM (Rosenzweig, et al., 1992; Serrano P.A., et al., 1994)."

Rose (1995) suggested that in chicks a kind of LTM that lasts a few hours (Rozensweig's ITM) involves a first wave of glycoprotein synthesis, whereas “true long-term memory” (LTM) requires a second wave of glycoprotein synthesis, occurring about 6 hours after training.

Rozensweig (2007) reviews evidence that the neurochemical cascades in *memory* in the chick are similar to the cascades in formation of LTP in different species. He states:

"The neurochemical cascade involved in formation of memory in the chick was soon shown to be similar to the cascade involved in long-term potentiation in the mammalian brain (Colley & Routtenberg, 1993) and in the nervous systems of invertebrates (Krasne & Glanzman, 1995). DeZazzo and Tully (1995) compared STM, ITM, and LTM in fruit flies, chicks, and rats. Tully and coworkers have shown that the three stages of memory in the fruit fly depend on three different genes (Tully et al., 1996)."

More recent research on LTP also confirms a three-stage process in LTP. In a review of the literature on LTP, Bliss, et al. (2018) note:

"The labels LTP1 and LTP2 equate to the forms of LTP that are, respectively, independent of and dependent on *de novo* protein synthesis. These are frequently referred to as early-phase LTP and late-phase LTP (E-LTP and L-LTP, respectively) implying that protein synthesis is not required initially but is required at later stages, with the switch-over [between E-LTP and L-LTP] occurring during a period of a few hours."

Although LTP has many properties that make it a good candidate for the mechanism of learning and memory in the brain, critical evidence in behaving animals is still needed. Nevertheless, neuroscientists are optimistic, and for good reason. According to Bliss, et al. (2018), "Today, LTP can be studied at every level from the purely molecular to the cognitive. Although definitive proof that the mechanisms of LTP subserve learning and memory in the behaving animal is still lacking, few neuroscientists doubt that such proof will eventually be forthcoming . . . there is now very strong evidence that an LTP-like mechanism mediates at least some aspects of memory." Much of this optimism comes from numerous studies showing that physiological, genetic, or pharmacological manipulations of LTP (either facilitating it or inhibiting it) have similar effects on learning and memory (Bliss, et al., 2018; Rozenzweig, 2007).

10.4.9.1 Anatomical Subregions of the Hippocampus and Memory

As mentioned above, hippocampal and cortical pyramidal neurons may receive tens of thousands of mostly excitatory inputs from other neurons onto their dendritic spines. As shown below in Figure **10.4.8**, the hippocampus consists of a number of subregions. These regions are involved in different functions related to learning and memory. "The hippocampus proper is defined by the dentate gyrus and Cornu Ammonis (CA). While the dentate gyrus contains the fascia dentata and the hilus, the CA is anatomically and functionally differentiated into distinct subfields named CA1, CA2, CA3, and CA4. . . . [T]he CA3 subfield . . . with inputs from the dentate gyrus and entorhinal cortex . . . is implicated in encoding spatial representations and episodic memories. . . The mossy fiber pathway . . . translates . . . cortical signals to a . . . hippocampal code, essential for memory formation" (Cherubini & Miles, 2015, p.19). According to Yang et al. (2014), "CA3 pyramidal neurons form extensive recurrent connections with each other. Such connections are able to learn to associate components of an input pattern with each other." Perhaps this capacity would be required to link the components of a memory into a unified whole.

A prominent model of the organization of the hippocampus highlights what neuroscientists have dubbed the "trisynaptic circuits," circuits with three synapses (dentate--CA3--CA1), which according to the "lamellar hypothesis," are stacked upon one another along the body of the hippocampus. According to this model, "the hippocampus is organized as a stack of parallel, trisynaptic circuits" (Yang et al., 2014, p. 12919).

However, additional research suggests a network oriented perpendicular to the trisynaptic circuits. Yang et al. (2014, 12919-20) found a "well-organized, longitudinally projecting synaptic network among CA1 pyramidal neurons . . . [and] that synapses of this network are capable of supporting synaptic plasticity, including long-term potentiation, and a short-term memory mechanism called 'dendritic hold and read.' . . . [Furthermore,] LTP can strengthen interlamellar CA1-to-CA1 connections as well as the well-established CA3-to-CA1 connections in the transverse plane" (Yang et al., 2014, p. 12921). Might these two networks in planes perpendicular to one another in the hippocampus suggest a kind of grid-like arrangement which might code location of objects, including oneself, in three-dimensional space? As mentioned above, CA3 is implicated in the representation of space and spatial relationships. In addition, neuroscientists have found "place cells" in the hippocampus which fire when an animal is in a specific place in a maze, for example, and "place cells" tuned to fire to specific locations also seem to be present in the human hippocampus. According to Yang et al. (2014, p. 12923), "These properties suggest that this system may be an integral component of the larger 3D information processing network of the hippocampus." Clearly, processing of three-dimensional space and one's spatial relations to other objects is critical to navigating the world successfully.

However, Yang and colleagues consider another possibility--that the CA1 network may "transform a time sequence into a spatial sequence" so that the "resulting 'time-to-space' transform is effectively an efficient sequence memory mechanism" such as would be necessary when "information typically arrives as a time sequence over a finite interval and its meaning can only be revealed if each sequence is viewed as a whole and in proper order. . . [Furthermore,] behavioral studies have suggested that area CA1 possesses the capacity for sequence memory" (Yang et al., 2014, p. 12923). To understand what this means, think about events in the world--they take place over time and their meaning, including often their adaptive significance, may only become apparent when they are perceived as a whole sequence. If you introspect for a moment and consider what a memory is like when you recall it, you may get a sense of this. For example, I have a very distinct memory of the first time I saw the planet Saturn and its rings through a telescope I bought when I was 14 (which I still have) using money I had saved from delivering newspapers door-to-door on my paper route (yes, I am that old). After what is now 61 years ago, I still remember setting up my telescope on a cold night in the driveway of my parents' house, finding Saturn almost by accident, viewing the rings around the planet, and then running inside through the kitchen door to tell my parents what I was seeing. Could it be that the CA1 subregion of the hippocampus is an important component in the ability to remember events in their proper sequence? Sequence memory appears to be an essential property of the brain in order to represent and store vital information about the temporal order of events we experience in the world.

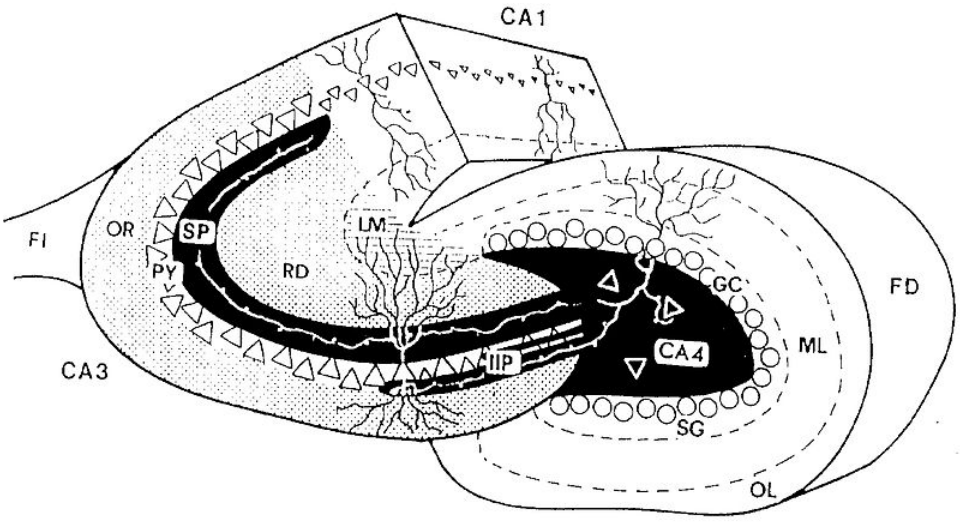


Figure 10.4.8: Diagram showing anatomy of the hippocampus. Pyramidal neurons in area CA1 have as many as 30,000 dendritic spines per neuron on their dendrites. Spines are associated with synapses which are primarily excitatory. The hippocampal subregion CA3-CA4 is indicated in black, stippled, and hatched areas. Black areas: suprapyramidal (SP), intra- and infrapyramidal (IIP) and hilar (CA4) mossy fiber terminal fields originating from the dentate gyrus. Stippled area: strata oriens (OR) and radiatum (RD). Hatched area: stratum lacunosum-moleculare (LM). CA1, subregion of the hippocampus without mossy fibers; FI, fimbria hippocampi; FD, fascia dentata; OL and ML, outer and middle molecular layers of the fascia dentata; SG, supragranular layer; GC, granular cells (Image and caption from Wikimedia Commons; File:Diagram of a Timm-stained cross-section of the hippocampus.JPEG; https://commons.wikimedia.org/wiki/F...ppocampus.JPEG; by Sluyter, Frans; Laure Jamot, Jean-Yves Bertholet, Wim Crusio (2005-04-22). "Prenatal exposure to alcohol does not affect radial maze learning and hippocampal mossy fiber sizes in three inbred strains of mouse". *Behavioral and Brain Functions* **1** (1): 5. DOI:10.1186/1744-9081-1-5. ISSN 1744-9081. Retrieved on 2007-12-21; licensed under the Creative Commons Attribution 2.0 Generic license).

10.4.9.2 References

Bennett E.L., et al. (1964). Chemical and anatomical plasticity of the brain. Science, 146:610. [PubMed]

Bertling, E., Ludwig, A., Koskinen, M., & Hotulainen, P. (2012). Methods for three-dimensional analysis of dendritic spine dynamics. In *Methods in enzymology* (Vol. 506, pp. 391-406). Academic Press.

Bliss T.V. & Collingridge G.L. (1993). A synaptic model of memory: long-term potentiation in the hippocampus. *Nature*. 361 (6407): 31–39

Bliss, T. V., Collingridge, G. L., Morris, R. G., & Reymann, K. G. (2018). Long-term potentiation in the hippocampus: discovery, mechanisms and function. *Neuroforum*, *24* (3), A103-A120.

Cherubini, E., & Miles, R. M. (2015). The CA3 region of the hippocampus: how is it? What is it for? How does it do it?. *Frontiers in cellular neuroscience*, *9*, 19.

Citri A., & Malenka R.C. (2008). Synaptic plasticity: multiple forms, functions, and mechanisms. *Neuropsychopharmacology*. 33 (1): 18–41.

Colley P.A., Routtenberg A. (1993). Long-term potentiation as synaptic dialogue. *Brain Res Brain Res Rev.* 18:115. [PubMed]

DeZazzo J., Tully T. (1995). Dissection of memory formation: from behavioral pharmacology to molecular genetics. *Trends Neurosci*. 18:212. [PubMed]

Hebb, D.O. (1949). *The Organization of Behavior*. New York. Wiley.

Irie, F., & Yamaguchi, Y. (2009). Eph receptor signaling and spine morphology. Encyclopedia of Neuroscience, p. 1141-1145.

Koskinen, M., Bertling, E., & Hotulainen, P. (2012). Methods to measure actin treadmilling rate in dendritic spines. In *Methods in enzymology* (Vol. 505, pp. 47-58). Academic Press.

Krasne F.B., Glanzman D.L. (1995). What we can learn from invertebrate learning. *Annu Rev Psychol*. 46:585.

Leuner, B., & Shors, T. J. (2010). Synapse formation and memory. In *Encyclopedia of Behavioral Neuroscience* (pp. 349-355). Elsevier Inc.

Mizumori S.J., Rosenzweig M.R., Bennett E.L. (1985). Long-term working memory in the rat: effects of hippocampally applied anisomycin. Behav Neurosci.;99:220. [PubMed]

Nicoll, R. A., & Kauer, J. A. Malenka. RC 1988. *The current excitement in long-term potentiation. Neuron*, *1*, 97-103.

Park, P., Georgiou, J., Sanderson, T.M. et al. (2021). PKA drives an increase in AMPA receptor unitary conductance during LTP in the hippocampus. *Nat Commun* **12,** 413. This article is licensed under a Creative Commons Attribution 4.0.

Rose S.P.R. (1992). *Neuropsychology of Memory*. Squire L.R., et al., editors. Guilford; New York: p. 547.

Rose S.P. (1995). Glycoproteins and memory formation. *Behav Brain Res.* 66:73. [PubMed]

Rosenzweig M.R., et al. (1992). *Neuropsychology of Memory*. Squire LR, et al., editors. Guilford; New York: p. 533.

Rosenzweig, M. R. (2007). Modification of brain circuits through experience. In *Neural Plasticity and Memory: From Genes to Brain Imaging*. CRC Press/Taylor & Francis, Boca Raton (FL); 2007. PMID: 21204433.

Serrano P.A., et al. (1994). Differential effects of protein kinase inhibitors and activators on memory formation in the 2-day-old chick. *Behav Neural Biol. 61*:60. [PubMed]

Traynelis S.F., Wollmuth L.P., McBain C.J., Menniti F.S., Vance K.M., Ogden K.K., Hansen K.B., Yuan H., Myers S.J., Dingledine, R. (2010). Glutamate receptor ion channels: structure, regulation, and function. *Pharmacol. Rev*. *62* (3): 405–496

Tully T, et al. A return to genetic dissection of memory in Drosophila. Cold Spring Harb Symp Quant Biol. 1996;61:207. [PubMed]

Yang, S., Yang, S., Moreira, T., Hoffman, G., Carlson, G. C., Bender, K. J., ... & Tang, C. M. (2014). Interlamellar CA1 network in the hippocampus. *Proceedings of the National Academy of Sciences*, *111*(35), 12919-12924.

10.4.10 Additional References

Alvarez, V. A., & Sabatini, B. L. (2007). Anatomical and physiological plasticity of dendritic spines. *Annu. Rev. Neurosci.*, *30*, 79-97.

Bliss, T. V., & Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *The Journal of physiology*, *232*(2), 331-356.

Chidambaram, S. B., Rathipriya, A. G., Bolla, S. R., Bhat, A., Ray, B., Mahalakshmi, A. M., ... & Sakharkar, M. K. (2019). Dendritic spines: revisiting the physiological role. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *92*, 161-193.

Eichenbaum, H. (1999). Conscious awareness, memory, and the hippocampus. *Nature Neuroscience*, *2* (9), 775–776.

Hayashi-Takagi, A., Yagishita, S., Nakamura, M., Shirai, F., Wu, Y.I., Loshbaugh, A.L., Kuhlman, B., Hahn, KM., Kasai, H. (2015). Labelling and optical erasure of synaptic memory traces in the motor cortex. *Nature 525*, 333–338.

Harris, K. M., Fiala, J. C., & Ostroff, L. (2003). Structural changes at dendritic spine synapses during long-term potentiation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358* (1432), 745-748.

Hebb, D.O. (1949). *The Organization of Behavior*. New York. Wiley.

Henry, F. E., Hockeimer, W., Chen, A., Mysore, S. P., & Sutton, M. A. (2017). Mechanistic target of rapamycin is necessary for changes in dendritic spine morphology associated with long-term potentiation. *Molecular Brain*, *10* (1), 1-17.

Jonides, J., Lacey, S. C., & Nee, D. E. (2005). Processes of working memory in mind and brain. *Current Directions in Psychological Science*, *14* (1), 2–5.

Kaada, B. R., Rasmussen, E. W., & Kveim, O. (1961). Effects of hippocampal lesions on maze learning and retention in rats. *Experimental Neurology*, *3*(4), 333-355.

Kandel, E. (1976). *Cellular Basis of Behavior*. San Francisco. W.H. Freeman and Company.

Kandel, E. R., & Schwartz, J. H. (1982). Molecular biology of learning: Modulation of transmitter release. *Science*, 218 (4571), 433–443

Kasai, Haruo; Matsuzaki, Masanori; Noguchi, Jun; Yasumatsu, Nobuaki; Nakahara, Hiroyuki (1 July 2003). "Structure–stability–function relationships of dendritic spines". *Trends in Neurosciences*. **26** (7): 360–368.

Kasthuri, N., Hayworth, K. J., Berger, D. R., Schalek, R. L., Conchello, J. A., Knowles-Barley, S., ... & Lichtman, J. W. (2015). Saturated reconstruction of a volume of neocortex. *Cell*, *162* (3), 648-661.

Kim, C. H., & Lisman, J. E. (1999). A role of actin filament in synaptic transmission and long-term potentiation. *Journal of Neuroscience*, *19* (11), 4314-4324.

Kolb, B. & Whishaw, I.Q. (2001). *In Introduction to Brain and Behavior*. New York. Worth Publishers.

Krucker, T., Siggins, G. R., & Halpain, S. (2000). Dynamic actin filaments are required for stable long-term potentiation (LTP) in area CA1 of the hippocampus. *Proceedings of the National Academy of Sciences*, *97*(12), 6856-6861.

Krupa, D. J., Thompson, J. K., & Thompson, R. F. (1993). Localization of a memory trace in the mammalian brain. *Science*, *260*(5110), 989–991.

Lashley, K. S. (1929). The effects of cerebral lesions subsequent to the formation of the maze habit: Localization of the habit. In *Brain mechanisms and intelligence: A quantitative study of injuries to the brain* (pp. 86–108). Chicago, IL: University of Chicago Press.

Miller, R. R., & Marlin, N. A. (2014). Amnesia following electroconvulsive shock. *Functional disorders of memory*, 143-178. New York. Psychology Press.

Murakoshi, H., Wang, H., & Yasuda, R. (2011). Local, persistent activation of Rho GTPases during plasticity of single dendritic spines. *Nature*, *472* (7341), 100-104.

Ofer, N., Berger, D. R., Kasthuri, N., Lichtman, J. W., & Yuste, R. (2021). Ultrastructural analysis of dendritic spine necks reveals a continuum of spine morphologies. *Developmental Neurobiology*.

Pinel, P.J. & Barnes, S. (2021). *Biopsychology* (11th Edition). Boston. Pearson Education.

Sigurdsson, T., Doyère, V., Cain, C. K., & LeDoux, J. E. (2007). Long-term potentiation in the amygdala: A cellular mechanism of fear learning and memory. *Neuropharmacology*, *52* (1), 215–227.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*(2), 195–231.

Thompson, R. (1974). Localization of the “maze memory system” in the white rat. *Physiological Psychology*, *2* (1), 1-17.

Woodruff-Pak, D. S., Goldenberg, G., Downey-Lamb, M. M., Boyko, O. B., & Lemieux, S. K. (2000). Cerebellar volume in humans related to magnitude of classical conditioning. *Neuroreport: For Rapid Communication of Neuroscience Research*, *11*(3), 609–615.

Yang, G., Pan, F., & Gan, W. B. (2009). Stably maintained dendritic spines are associated with lifelong memories. *Nature*, *462* (7275), 920-924.

10.4.11 Attributions

"Learning Objectives," "Overview," "The Search for Learning and Memory in the Synapse," "Changes at the Synapse Correlated with Learning and Memory," "Changes at the Synapse Correlated with Learning and Memory: Anatomical Changes in Dendritic Spines," "LTP, Neurochemical Cascades, and Stages of Memory," original material written by Kenneth A. Koenigshofer, Ph.D., is licensed under CC BY 4.0.

"Synaptic Plasticity," "Synaptic Plasticity: short-term enhancement, long-term potentiation and long-term depression," "Short-term Synaptic Enhancement and Depression," "Long-term Potentiation (LTP)," and "Long-term Depression (LTD)," adapted by Kenneth

A. Koenigshofer, Ph.D., from Lumen Boundless Biology, How Neurons Communicate; https://courses.lumenlearning.com/bo...s-communicate/; curation and Revision. Provided by: Boundless.com. License: ***CC BY-SA: Attribution-ShareAlike.***

"Cellular and Receptor Level Mechanisms Revisited" adapted from The Brain from Top to Bottom; license: Copyleft, https://thebrain.mcgill.ca/flash/pop.../pop_copy.html; modified by Kenneth A. Koenigshofer, PhD., licensed under CC BY 4.0.

This page titled 10.4: Synaptic Mechanisms of Learning and Memory is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

10.5: Brain Mechanisms of Memory Disorders

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

10.5.0.1 Learning Objectives

1. Explain the roles of the hippocampus and medial temporal lobes in memory.
2. Describe the case of H.M. and how it contributed to our understanding of the brain mechanisms in memory.
3. Describe global amnesic syndrome.
4. Discuss retrograde and anterograde amnesias including temporally graded amnesia.
5. Describe amnesia of the frontal lobes.
6. Discuss specific amnesias due to specific and circumscribed cortical damage.

10.5.0.1 Overview

We now examine the roles of the hippocampus and medial temporal lobes in memory and memory loss, focusing on amnesic syndromes associated with Korsakoff's Syndrome and Alzheimer's Disease. We also consider frontal lobe amnesia and highly specific forms of amnesia for highly specific types of information due to localized cortical damage. Damage to the amygdala disrupts emotional memories including memories of traumatic events. Diseases or other causes of cellular loss in the hippocampi can produce amnesia.

10.5.0.1 **Amnesia, Hippocampus, and Medial Temporal Lobes**

The Russian psychologist A. R. Luria (1968) has described the abilities of a man known as “S,” who seems to have unlimited memory. S remembers strings of hundreds of random letters for years at a time, and seems in fact to never forget anything. But what would this be like?

Shereshevsky, or “S,” the mnemonist studied by Luria was a man who almost never forgot. His memory appeared to be virtually limitless. He could memorize a table of 50 numbers in under 3 minutes and recall the numbers in rows, columns, or diagonals with ease. He could recall lists of words and passages that he had memorized over a decade before. Yet Shereshevsky found it difficult to function in his everyday life because he was constantly distracted by a flood of details and associations that sprung to mind. His case history suggests that remembering everything is not always a good thing. You may occasionally have trouble remembering where you parked your car, but imagine if every time you had to find your car, every single former parking space came to mind. The task would become impossibly difficult to sort through all of those irrelevant memories. Thus, forgetting is adaptive in that it makes us more efficient. The price of that efficiency is those moments when our memories seem to fail us (Schacter, 1999).

Clearly, remembering everything would be maladaptive, but what would it be like to remember nothing? We will now consider a profound form of forgetting called amnesia that is distinct from more ordinary forms of forgetting. Most of us have had exposure to the concept of amnesia through popular movies and television. Typically, in these fictionalized portrayals of amnesia, a character suffers some type of blow to the head and suddenly has no idea who they are and can no longer recognize their family or remember any events from their past. After some period of time (or another blow to the head), their memories come flooding back to them. Unfortunately, this portrayal of amnesia is not very accurate. What does amnesia typically look like?

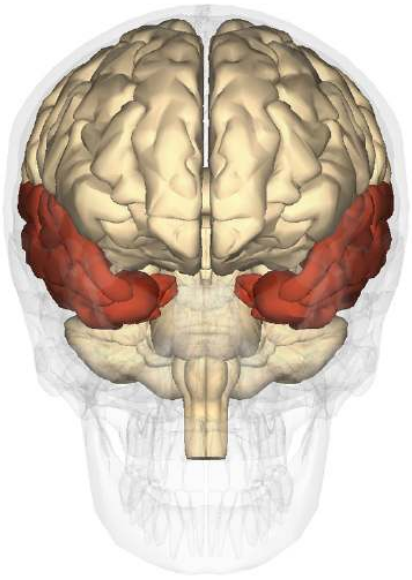


Figure **10.5.1**: Patients with damage to the temporal lobes may experience anterograde amnesia and/or retrograde amnesia. [Image from Wikimedia Commons; File:Temporal lobe - anterior view.png; https://commons.wikimedia.org/wiki/F...erior_view.png; by Anatomography; licensed under the Creative Commons Attribution-Share Alike 2.1 Japan license; CC BY-SA 2.1 JP, https://goo.gl/BDF2Z4]

As previously discussed, the most widely studied amnesic patient was known by his initials H. M. (Scoville & Milner, 1957). H.M.'s disorder is an example of **global amnesic syndrome** characterized by severe anterograde amnesia and more moderate retrograde amnesia (see below). This syndrome results from bilateral lesions of the medial portion of the temporal lobe, and more specifically, of the hippocampus and its neighboring structures (the parahippocampal, entorhinal, and perirhinal cortices). These lesions can be due to surgical ablation, as in the case of H.M., or to other causes such as tumors, ischemic episodes, head traumas, and various forms of encephalitis.

As a teenager, H. M. suffered from severe epilepsy, and in 1953, he underwent surgery to have both of his medial temporal lobes removed to relieve his epileptic seizures. The **medial temporal lobes** encompass the hippocampus and surrounding cortical tissue. Although the surgery was successful in reducing H. M.'s seizures and his general intelligence was preserved, the surgery left H. M. with a profound and permanent memory deficit. From the time of his surgery until his death in 2008, H. M. was unable to learn new information, a memory impairment called **anterograde amnesia**. H. M. could not remember any event that occurred since his surgery, including highly significant ones, such as the death of his father. He could not remember a conversation he had a few minutes prior or recognize the face of someone who had visited him that same day. He could keep information in his short-term, or working, memory, but when his attention turned to something else, that information was lost for good. It is important to note that H. M.'s memory impairment was restricted to **declarative memory**, or conscious (explicit) memory for facts and events. H. M. could learn new motor skills and showed improvement on motor tasks even in the absence of any memory for having performed the task before (Corkin, 2002).

In addition to anterograde amnesia, H. M. also suffered from **temporally graded retrograde amnesia**. **Retrograde amnesia** refers to an inability to retrieve old memories that occurred before the onset of amnesia. Extensive retrograde amnesia in the absence of anterograde amnesia is very rare (Kopelman, 2000). More commonly, retrograde amnesia co-occurs with anterograde amnesia and shows a temporal gradient, in which memories closest in time to the onset of amnesia are lost, but more remote memories are retained (Hodges, 1994). In the case of H. M., he could remember events from his childhood, but he could not remember events that occurred a few years before the surgery.

Amnesiac patients with damage to the **hippocampus and surrounding medial temporal lobes** typically manifest a similar clinical profile as H. M. The degree of anterograde amnesia and retrograde amnesia depend on the extent of the **medial temporal lobe damage**, with greater damage associated with a more extensive impairment (Reed & Squire, 1998). Anterograde amnesia provides evidence for the role of the **hippocampus** in the formation of long-lasting declarative memories, as damage to the hippocampus

results in an inability to create or consolidate this type of new long-term memory. Similarly, temporally graded retrograde amnesia can be seen as providing evidence for the importance of **memory consolidation** (Squire & Alvarez, 1995). A memory depends on the hippocampus until it is consolidated (or "fixed") and transferred into a more durable form that is **stored in the cortex**. According to this theory, an amnesiac patient like H. M. could remember events from his remote past because those memories were fully consolidated and no longer depended on the hippocampus, but instead had been transferred to other brain areas, primarily to particular areas of the cerebral cortex.

The classic amnesiac syndrome we have considered here is sometimes referred to as **organic amnesia**, and it is distinct from functional, or dissociative, amnesia. Functional amnesia involves a loss of memory that cannot be attributed to brain injury or any obvious brain disease and is typically classified as a mental disorder rather than a neurological disorder (Kihlstrom, 2005). The clinical profile of dissociative amnesia is very different from that of patients who suffer from amnesia due to brain damage or deterioration. Individuals who experience **dissociative amnesia** often have a history of trauma. Their amnesia is retrograde, encompassing autobiographical memories from a portion of their past. In an extreme version of this disorder, people enter a dissociative fugue state, in which they lose most or all of their autobiographical memories and their sense of personal identity. They may be found wandering in a new location, unaware of who they are and how they got there. Dissociative amnesia is controversial among research psychologists, as both the causes and existence of it have been called into question. The memory loss associated with dissociative amnesia is much less likely to be permanent than it is in organic amnesia.

10.5.0.1 Korsakoff's Amnesia

Another well-known form of amnesia is **Korsakoff's syndrome**, encountered for the first time in chronic alcoholics. Korsakoff's syndrome is similar to global amnesic syndrome, except that people with Korsakoff's are more prone to confabulation to cover up gaps in their memories of their own past. Korsakoff's syndrome is also known as **diencephalic amnesia** (or Wernicke-Korsakoff syndrome), because the vitamin B1 deficiency that results from alcoholism causes bilateral damage to the **mammillary bodies of the hypothalamus**. Similar symptoms are also produced by damage to the **dorsomedial thalamic nuclei, the mammillothalamic tract, and the upper portion of the brainstem**. Once again, other etiologies, such as strokes and tumors, can affect the same structures and produce the same results.



Figure 10.5.2: Brain affected by Wernicke-Korsakoff syndrome. Note the pigmentation of the grey matter around the third ventricle. (Image Source: University of Texas (Houston); https://thebrain.mcgill.ca/flash/a/a...07_cr_oub.html; The Brain from Top to Bottom; Copyleft licence).

10.5.0.1 Alzheimer's Disease

In Alzheimer's disease, beta-amyloid, an insoluble toxic substance, forms clumps known as "senile plaques" around the neurons. These plaques release free radicals that strip atoms from organic molecules that are vital to the neurons, including molecules in their cell membranes. Holes thus develop in these membranes, allowing large amounts of harmful substances to enter and kill the neurons. The memory circuits that depend on them, especially those in the hippocampus, are thus permanently damaged.

The National Institute on Aging reports on research indicating that the brains of Alzheimer's patients also contain neurofibrillary tangles which consist of a protein called tau. This protein normally plays a role in healthy microtubules inside neurons that assist with the transport of vital substances within the neuron cell body to the axon and dendrites. In Alzheimer's they form clumps that disrupt the neuron's transport system interfering with synaptic transmission. Complex interactions between beta-amyloid and

tangles and other factors may then result in Alzheimer's. "It appears that abnormal tau accumulates in specific brain regions involved in memory. Beta-amyloid clumps into plaques between neurons. As the level of beta-amyloid reaches a tipping point, there is a rapid spread of tau throughout the brain" (https://www.nia.nih.gov/health/what-...eimers-disease; retrieved 5/11/2022).

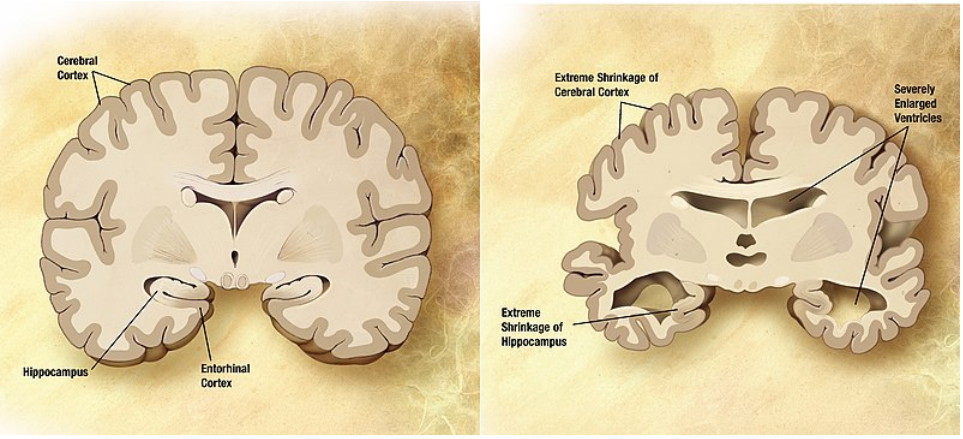


Figure **10.5.3**: Diagram showing changes of the brain caused by Alzheimer's disease. Note extreme shrinkage of cerebral cortex and hippocampus, and extensive deterioration and loss of brain tissue resulting in extremely enlarged cerebral ventricles (Image and caption from Wikimedia Commons; File:Alzheimer's disease brain comparison.jpg; https://commons.wikimedia.org/wiki/F...comparison.jpg; by SEVERESLICE_HIGH.JPG: ADEAR: "Alzheimer's Disease Education and Referral Center, a service of the National Institute on Aging," modifications by Garrondo; this work is in the **public domain**. This applies worldwide).

According to the National Institute on Aging, "at first, Alzheimer's disease typically destroys neurons and their connections in parts of the brain involved in memory, including the entorhinal cortex and hippocampus. It later affects areas in the cerebral cortex responsible for language, reasoning, and social behavior. Eventually, many other areas of the brain are damaged. Over time, a person with Alzheimer's gradually loses his or her ability to live and function independently. Ultimately, the disease is fatal" (https://www.nia.nih.gov/health/what-...eimers-disease; retrieved 5/11/2022).

10.5.0.1 Specific Cortical Damage and Memory Loss

There is also an **amnesia of the frontal lobe** due to damage at this site. People with this disorder do not suffer from global amnesia, but do show a memory deficit in tasks involving temporal planning of sequences of events. These people also have problems with the sources of newly acquired knowledge and have deficient meta-memory (they cannot make judgments about their memory's contents).

Other types of damage to the cortex can cause forms of amnesia that are **sometimes highly specific**. For example, if the part of the cortex that perceives colors is damaged, people can lose their knowledge of color. And since the memory of colours is reconstructed at this same location, this memory disappears as well.

Other localized cortical lesions can prevent people from accessing certain items in their semantic memory and thus cause all sorts of specialized aphasias (language disorders).

10.5.0.1 Other Specific Forms of Memory Loss

A specific injury to the amygdala can prevent people from recording memories of traumatic events. In normal people, such memories are formed when particularly stressful conditions make certain details of a scene practically unforgettable.

Certain encephalopathies due to anoxias, ischemias, hypoglycemias, carbon monoxide poisoning, or prolonged epileptic attacks can cause the **loss of large numbers of neurons in both hippocampi.**

The pyramidal neurons of hippocampal area CA1, as well as the cortical neurons of layers 3, 5 and 6, the Purkinje cells, and the striate neurons are especially sensitive to lack of oxygen and energy.

Since these neurons are involved in various memory systems, malfunctions in their circuits inevitably lead to memory problems.

Thus, damage to the temporal lobes of the cortex can cause severe, permanent anterograde amnesia, as well as retrograde amnesia extending back from three to ten years before the accident.

When selective neuronal losses occur in area CA1 of the hippocampus, the resulting anterograde amnesia is just as severe, but the resulting retrograde amnesia generally remains slight (extending only one to two years before the accident).

Lastly, certain **transitory global amnesias** can be triggered suddenly, causing people to completely lose their memory for a few hours. Though these transitory amnesic episodes are frightening, they are brief and do not cause any permanent damage to the brain. They seem to be due to a temporary vascular insufficiency in the brain tissue.

10.5.1 Summary

Amnesiac patients show us what a life without memory would be like. Amnesiac patients suggest a special role for the hippocampus and surrounding medial temporal lobe structures in memory. Korsakoff's Syndrome is caused by vitamin B1 deficiency due to long term use of alcohol. Alzheimer's disease is associated with damage to neurons caused by plaques and tangles. Damage to specific cortical areas can cause very specific forms of memory loss including specific aphasias.

10.5.1.0.1 Outside Resources

10.5.1.0.1 (click resource for link)

Web: Brain Case Study: Patient HM

https://bigpictureeducation.com/brain-case-study-patient-hm

Web: Self-experiment, Penny demo

http://www.indiana.edu/~p1013447/dictionary/penny.htm

Web: The Man Who Couldn't Remember

http://www.pbs.org/wgbh/nova/body/corkin-hm-memory.html

10.5.1.1 References

Corkin, S. (2002). What's new with the amnesic patient H. M.? *Nature Reviews Neuroscience*, 3, 153–160.

Hodges, J. R. (1994). Retrograde amnesia. In A. Baddeley, B. A. Wilson, & F. Watts (Eds.), *Handbook of Memory Disorders* (pp. 81–107). New York: Wiley.

Kihlstrom, J. F. (2005). Dissociative disorders. Annual *Review of Clinical Psychology*, 1, 227– 253.

Kopelman, M. (2000). Focal retrograde amnesia and the attribution of causality: An exceptionally critical review. *Cognitive Neuropsychology*, 17, 585–621.

Luria, A. R. (1968). *The mind of a mnemonist: A little book about a vast memory* (L. Solataroff, Trans.). New York: Basic Books.

Reed, J. M. & Squire, L. R. (1998). Retrograde amnesia for facts and events: Findings from four new cases. *Journal of Neuroscience*, 18, 3943–3954.

Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist*, 54, 182–203.

Scoville, W. B. & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology*, Neurosurgery, & Psychiatry, 20, 11–21.

Squire, L. R., & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: A neurobiological perspective. *Current Opinions in Neurobiology*, 5, 169–177.

10.5.1.2 Attributions

Adapted by Kenneth A. Koenigshofer, PhD., from Forgetting and Amnesia by Nicole Dudukovic and Brice Kuhl, licensed by NOBA under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Permissions beyond the

scope of this license may be available in NOBA's Licensing Agreement. Dudukovic, N. & Kuhl, B. (2021). Forgetting and amnesia. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology.* Champaign, IL: DEF publishers. Retrieved from http://noba.to/m38qbftg; **Creative Commons License**

Korsakoff's Amnesia, Alzheimer's Disease, Specific Cortical Damage and Memory Loss, and Other Specific Forms of Memory Loss, adapted by Kenneth A. Koenigshofer, Ph.D., from "The Brain From Top to Bottom," Memory and the Brain; under Copyleft license.

This page titled 10.5: Brain Mechanisms of Memory Disorders is shared under a mixed license and was authored, remixed, and/or curated by Kenenth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

10.6: Localization of Types and Stages of Memory

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

10.6.1 Learning Objectives

1. Explain the role of the dorsolateral prefrontal cortex in memory, problem-solving and reasoning.
2. Discuss working memory and its various components including the central executive, the phonological loop, episodic buffer, and visuospatial sketch pad
3. Describe the region of brain associated with the operation of the central executive
4. Describe the brain structures involved in consolidation of new memories into long-term storage
5. Explain the different types of memory, three stages of memory, and the brain areas associated with each as discussed below
6. Describe the role of place cells in the formation of spatial maps in the hippocampus
7. Explain the trisynaptic circuit of the hippocampus and its role in memory

10.6.2 Overview

Here we examine memory from the standpoint of its different types focusing on the role of the hippocampus and other medial temporal lobe structures and circuits in the consolidation of new declarative memories. The hippocampal/mammillothalamic tract, also called Papez's circuit, plays an essential role in this process. After consolidation, long-term memories are stored in the cerebral cortex and become independent of the hippocampus. We also consider the role of the prefrontal cortex in working memory, important for the retention of information during problem-solving and reasoning. In addition, we discuss the role of the hippocampus in spatial memory and in the formation of cognitive maps or mental models of space.

10.6.3 **Types and Stages of Memory Localized in Different Brain Regions**

A large body of evidence indicates that the **dorsolateral prefrontal cortex** plays an important role in certain forms of memory work (Barbey, et al., 2013), in particular those that involve alternating between two memory tasks and exploring various possibilities before making a choice (Hunt et al., 2015). It seems fairly certain that dorsolateral prefrontal cortex holds information required for reasoning processes that are in progress.

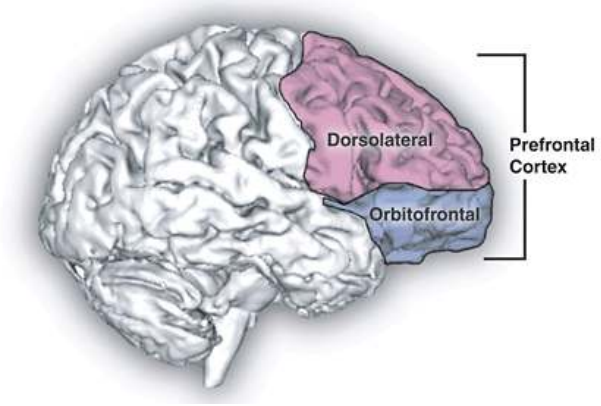


Figure **10.6.1**: Divisions of Human Prefrontal Cortex (Image from Wikimedia Commons; File:Prefrontal cortex.png; https://commons.wikimedia.org/wiki/F...tal_cortex.png; by Natalie M. Zahr, Ph.D., and Edith V. Sullivan, Ph.D.; image is a work of the National Institutes of Health and is in the public domain).

10.6.3.1 Short-term Memory

Short-term memory depends on the attention paid to the elements of sensory memory. Short-term memory lets you retain a piece of information for less than a minute and retrieve it during this time. One typical example of its use is the task of repeating a list of items that has just been read to you, in their original order. In general, you can retain 5 to 9 items (or, as it is often put, 7±2 items)

in short-term memory. It is clear that **short-term memory does not depend on** the **medial temporal cortex including the hippocampus and surrounding structures**. We know this because H.M. who had bilateral removal of his medial temporal lobe still retained his short-term explicit memory. If someone came into the room, he could remember the visit for a short time. Of course, as you have learned, if we checked his recollection a few minutes later, without ability to consolidate the memory into a long-term form, he would have no recollection of the visit that occurred just minutes before. This suggests that the **dorsolateral prefrontal cortex**, left intact in H.M., may be at least **one of the structures capable of retaining information for a brief period, particularly when engaged in reasoning and decision-making.** But its precise role remains the subject of much debate.

10.6.3.2 Working Memory

Working memory is a more recent extension of the concept of short-term memory. As techniques for studying memory have become more refined, it has become increasingly apparent that the original conception of short-term memory as a mere temporary receptacle for long-term memory is too simplistic. In fact, it is becoming increasingly clear that there is no strict line of demarcation between memories and thoughts. In order to test some hypotheses that may provide a better understanding of this complex phenomenon, the concept of working memory has therefore been advanced.

Working memory is used to perform cognitive processes on the items that are temporarily stored in it. It would therefore be heavily involved in processes that require reasoning, such as reading, or writing, or performing computations. One typical example of the use of working memory is the task of repeating a list of items that has just been read to you, but in the reverse of their original order. Another good example is the task of simultaneous language interpretation, where the interpreter must store information in one language while orally translating it into another. In the course of a day, there are many times when you need to keep some piece of information in your head for just a few seconds. Maybe it is a number that you are “carrying over” to do a subtraction, or a persuasive argument that you are going to make as soon as the other person finishes talking. These are good examples of why you usually hold information in your short-term/working memory: to accomplish something that you have planned to do. Perhaps the most extreme example of working memory is a chess master who can explore several possible solutions mentally before choosing the one that will lead to checkmate.

Working memory appears to be composed of several independent systems, which would imply that we are not aware of all the information that is stored in it at any given time. For example, when you drive a car, you are performing several complex tasks simultaneously. It is unlikely that all of the various types of information involved are being handled by a single short-term memory system.

Baddeley & Hitch (1974) proposed a model of **working memory** with **several components**: a certain number of auxiliary “slave” systems and a **central executive** that supervises and controls the flow of information to and from the "slave systems," which are short-term storage repositories for particular types or domains of information. One of these slave systems, the **phonological or articulatory loop**, specializes in processing linguistic information, while another, the **visuospatial sketchpad**, specializes in processing visual-spatial information. The **episodic buffer** is a third storage system, **dedicated to linking information across domains to form integrated units of visual, spatial, and verbal information** (e.g., the memory of a story, a movie scene, or an integrated autobiographical memory). The episodic buffer is also assumed to have **links to long-term memory**.

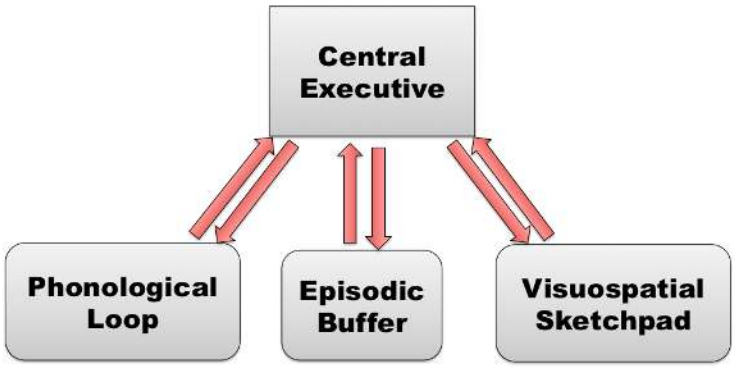


Figure 10.6.2: Baddeley's Model of Working Memory. (Image from Wikimedia Commons; Baddeley's Model of Working Memory; https://commons.wikimedia.org/wiki/F...ing_Memory.png; author, AmandaSilver15; permission is granted to copy, distribute and/or modify this document under the terms of the **GNU Free Documentation License**, Version 1.2 or any later version published by the Free Software Foundation).

The phonological or articulatory loop plays an important role in everyday life. For example, when you repeat a phone number to yourself in your head, you are activating this loop. This loop is also heavily involved in reading and writing. The presence **in working memory of another slave system that can manipulate mental images** of visual objects is suggested by tests where subjects are asked to rotate such images.

Perhaps the most important but least understood component in Baddeley's model of working memory is the **central processor or central executive**, whose job would be to select, initiate, and halt the routines performed by its slave systems.

Baddeley's model of **working memory** has proven especially fruitful for research on the brain areas involved in problem solving. This model posits a **central executive** or **central processor** that coordinates the activity of two sub-systems. Many brain-imaging studies show high activity in the frontal lobe, in particular the **prefrontal lobe,** when this central executive is working. For example, the image shown below (see Figure 10.5.3.2) was produced by **functional magnetic resonance imaging (fMRI)**. As you have learned in a previous chapter, fMRI is a technique based on the increased blood flow to the most active areas of the brain. In this image, taken while the subject was holding an image of a face in his memory, the yellow area in the **prefrontal cortex** is very active, confirming the role of these prefrontal cortical areas in working memory.

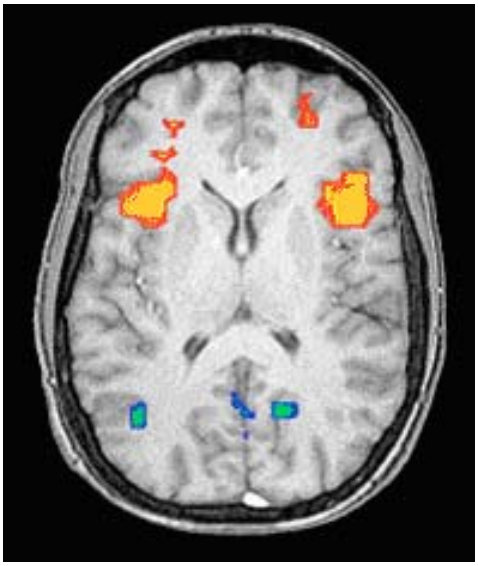


Figure 10.6.3: fMRI images of brain showing increased activity in the frontal lobes while the subject is keeping an image of a face in his memory (Image from NIMH Laboratory of Brain and Cognition. Published in *Nature*, Vol 386, April 10, 1997, p. 610, Courtney et al., 1997).

As you have learned in an earlier chapter, this region, at the very front of the brain, is highly developed in humans. It appears that the **prefrontal cortex**, by holding information in storage during **working memory**, may play a significant role in functions related to **attention, problem-solving, and intelligence** (Conway et al., 2003; Engle et al., 1999; Kane & Engle, 2002; Matzel and Kolata, 2010; Barbey et al., 2014; Waltz et al., 1999).

Recall that Baddeley's model also postulates the existence of a **phonological** (acoustic and linguistic) memory and a **visual/spatial** memory (containing mental images). Brain imaging studies have also revealed distinct neuroanatomical bases for both of these forms of short-term working memory.

The **phonological loop** activates certain areas in the **left hemisphere** that are associated with the production of language, such as **Wernicke's area in the temporal lobe** and **Broca's area in the frontal lobe**. **Visual/spatial memory** seems to be associated with a region of the **occipital cortex** generally associated with visual processing. This is consistent with the hypothesis that visual imagination in which humans can manipulate visual images in the "mind's eye" to mentally test out and anticipate future

consequences of action involves parts of the visual system "recruited" into this new function during evolution (Koenigshofer, 2017).

Meanwhile, certain sub-regions of the **prefrontal cortex** are activated only if the memorization exercise is somewhat difficult for the subject, thus confirming the coordinating role of the **central executive**.

The phenomenon of working memory is made all the more complex by the fact that it takes place over time. For example, the experimental results illustrated below show how various areas of the subjects' brains alter their activity levels as the subjects are presented with various visual stimuli. When the subjects are shown a blurred image, the activity level (represented by the blue bars in the graph) becomes highest in area 1, the visual part of the brain. When the subjects are shown an image of a face, brain activity (black bars) becomes highest in the **associative and frontal regions** (4, 5, and 6). Lastly, when the subjects are retaining an image of a **face** in their working memory, brain activity (red bars) is highest in the **frontal** regions, while the **visual areas are scarcely stimulated at all**.

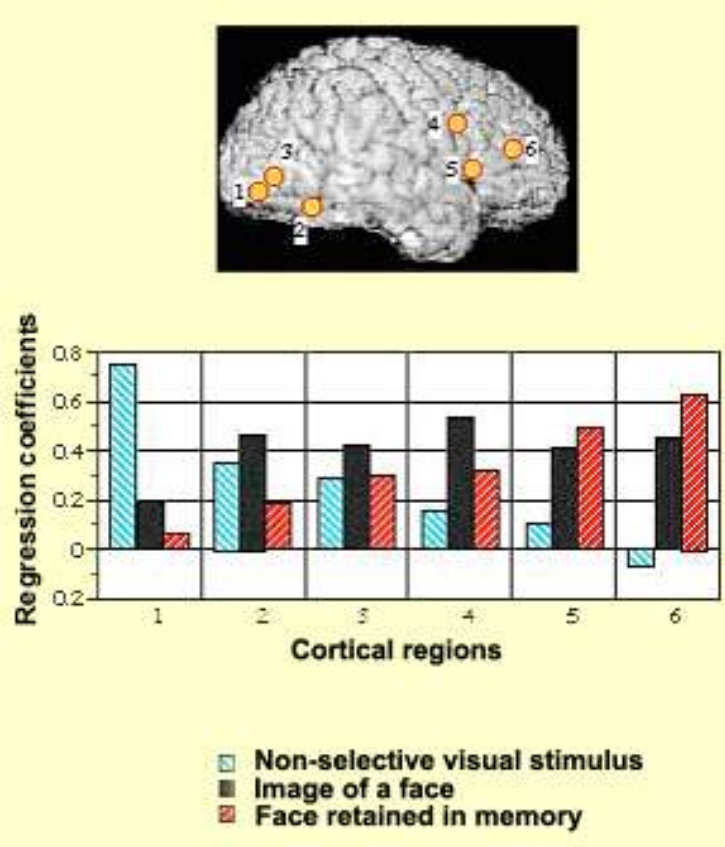


Figure **10.6.4**: Activity in cortical areas (labeled 1-6 in image of cortical surface) primarily in occipital lobe with presentation of blurred image (blue bars, tallest in area 1), in frontal and association cortical areas with presentation of a photo of a face (black bars), and in frontal areas (areas 5 and 6), with little activity in visual areas (red bars in cortical areas 1 and 2), when subject retains memory of a face in working memory (red bars). (Image source: NIMH Laboratory of Brain and Cognition. Published in *Nature*, Vol. 386, April 1997, p. 610, Courtney et al., 1997).

It has also been observed that distinct processes appear to be involved in the storage and recall of items memorized with the phonological loop and the visual/spatial sketchpad. One thing is certain: the **prefrontal cortex** plays a fundamental role in **working memory**. It enables people to keep information available that they need for their current reasoning processes. For this purpose, the prefrontal cortex must cooperate with other parts of the cortex from which it extracts information for brief periods. For this information to eventually pass into longer-term memory, the limbic system has to be brought into play.

10.6.3.3 Long-Term Memory

Recent research has provided a complex, highly intricate picture of memory functions and their loci in the brain. The **hippocampus, the temporal lobes, and the structures of the limbic system that are connected to them** are essential for the **consolidation of long-term declarative memory**. The **hippocampus receives** connections **from** the **cortex's primary sensory areas, unimodal associative areas** (those that involve only one sensory modality), and **multimodal associative areas**, as well as from the **rhinal and entorhinal cortexes.** While these anterograde connections converge at the hippocampus, other, retrograde **pathways emerge from it**, returning **to** the **primary cortexes**, **where** they record in the cortical synapses the **associations facilitated by the hippocampus**. Thus, in the mechanism of memorization, we find **feedback loops among these various brain structures.**

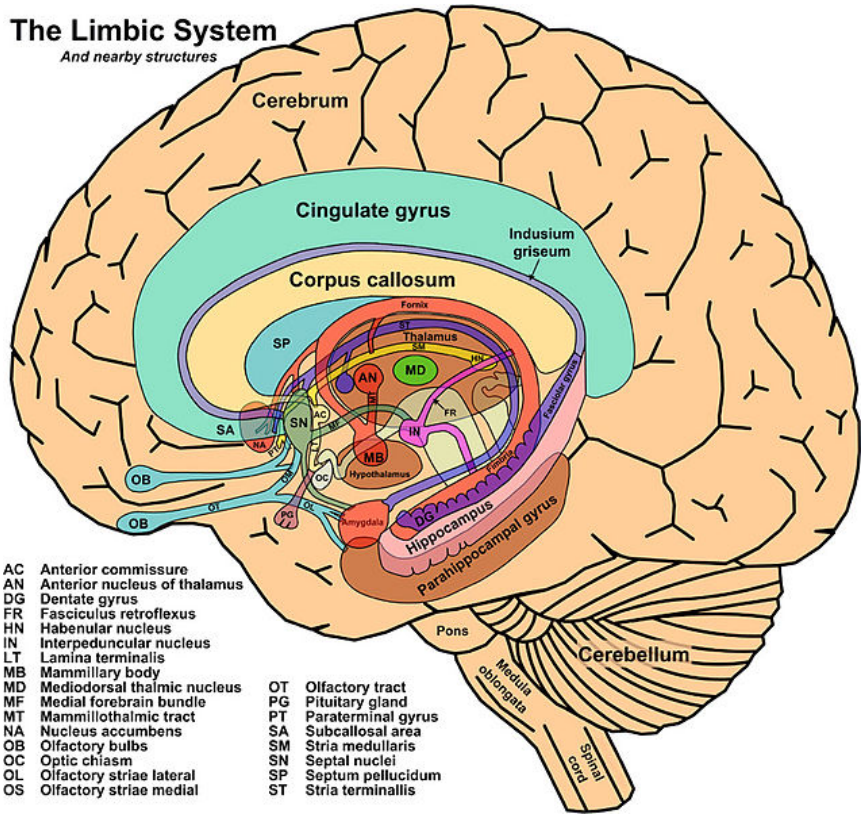


Figure **10.6.5**: Limbic system and surrounding structures (Image from Wikimedia; Structures of the Limbic System; https://commons.wikimedia.org/wiki/F...ohn_Taylor.jpg; This work has been released into the **public domain** by its author, John Taylor).

10.6.3.4 Hippocampus

The **hippocampus facilitates associations among various parts of the cortex**, for example, between a tune that you heard at a dinner party represented in auditory cortex in the temporal lobe and the faces of the other guests who were at the table, represented in the visual areas and frontal areas of cortex. However, all other things being equal, such associations would naturally fade over time, so that your mind did not become cluttered with useless memories. What might cause such associations to be strengthened and eventually etched into long-term memory very often depends on "limbic" factors, such as how interested you were in the occasion, or what emotional charge it may have had for you, or how gratifying you found its content.

Recall from prior modules of this chapter that the **hippocampus, the cortical structures surrounding it, and the neural pathways that connect them to the cortex** as a whole are all heavily involved in **declarative memory**–the memory of facts and events. Declarative (explicit) memory is composed of **semantic memory** (memory for facts and knowledge) and **episodic memory** (memory for episodes in your life).

Semantic memory is the system that you use to store your knowledge of the world. It is a knowledge base that we all have and much of which we can access quickly and effortlessly. It includes our memory of the meanings of words–the kind of memory that lets us recall not only the names of the world's great capitals, but also social customs, the functions of things, and their color and odor and other sensory properties.

Semantic memory also includes our memory of the rules and concepts that let us construct a **mental representation of the world** without any immediate perceptions. Its content is thus **abstract and relational** and is associated with the meaning of verbal symbols.

The **hippocampus** also plays a fundamental role in **episodic memory** (sometimes called **autobiographical memory**), the kind of memory that lets you remember events that you personally experienced at a specific time and place. For example, episodic memory lets you remember an especially pleasant dinner party years later. In fact, it seems to be the **hippocampus** that **enables** you to "**play the scene back**", by reactivating this particular activity pattern in the various regions of the cortex. This phenomenon would be very important during dreams, and would explain the incorporation of events from the last few days into them.

The various structures of the limbic system exert their influence on the hippocampus and the temporal lobe via **Papez's circuit**, also known as the **hippocampal/mammillothalamic tract**. This circuit is a sub-set of the numerous connections that the limbic structures have with one another. The route that information travels in this circuit is **from** the **hippocampus to** the **mammillary bodies of** the **hypothalamus,** then on to the **anterior thalamic nucleus,** the **cingulate cortex,** and the **entorhinal cortex,** before finally returning **to** the **hippocampus** (see Figure 10.5.5).

Once the temporary associations of cortical neurons generated by a particular event have made a certain **number of such "passes" through Papez's circuit**, they will have undergone a physical remodeling that **consolidates** them. Eventually, these associations will have been strengthened so much that they will stabilize and become **independent of the hippocampus**. As you have already learned, bilateral lesions of the hippocampus will prevent new long-term memories from forming, but will not erase those that were encoded and consolidated before the injury.

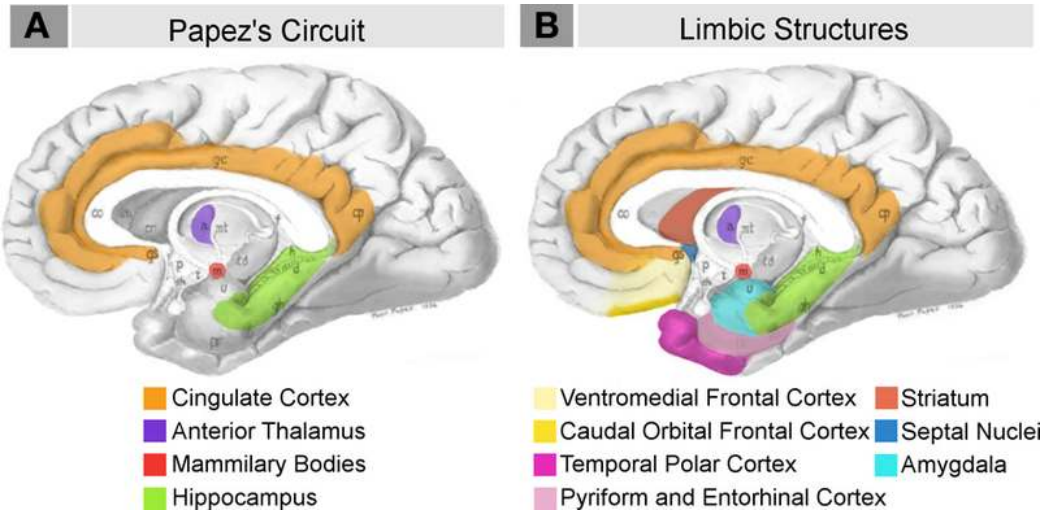


Figure 10.6.6: Papez Circuit and Limbic system structures some of which are involved in memory and emotion. (Images from Wikimedia Commons; File:Neural systems proposed to process emotion.png; https://commons.wikimedia.org/wiki/F...ss_emotion.png; by Barger N, Hanson KL, Teffer K, Schenker-Ahmed NM and Semendeferi K; licensed under the Creative Commons Attribution 3.0 Unported license. Caption by Kenneth A. Koenigshofer, PhD).

With this gradual disengagement of the **limbic system**, the **memories** will **no longer pass through Papez's circuit**, but **instead will be encoded in specific areas** of the **cortex:** the **same ones where the sensory information that created the memories was initially received** (the occipital cortex for visual memories, the temporal cortex for auditory memories, etc.).

Some very intense personal memories that bring what is sometimes called **emotional memory** into play appear to involve another structure of the limbic system besides the hippocampus. This structure is the **amygdala**, located beneath temporal lobe cortex near

the hippocampus, and not surprisingly, as discussed in an earlier chapter, the amygdala is already known to manage our reactions to fear. Emotional memories can be very strong and particularly long-lasting. The **release of norepinephrine** during highly emotional events may contribute to the rapid and effective **consolidation** of memory for emotionally-charged events.

Injury to the Limbic System Pathways and Memory

As described above, for a piece of information to be recorded in long-term memory, it must pass through **Papez's circuit**. Injuries to this circuit can result in memory impairments. For example, a **lesion in the mammillary bodies** is responsible for an **amnesic syndrome** whose most classic example is **Korsakoff's** syndrome. In addition to the confabulation, confusion, and disorientation that accompany this syndrome, patients suffer from **anterograde amnesia**: they cannot store new information in their long-term memory. The most typical cause of this syndrome is **vitamin B1 deficiency**, often seen in chronic alcoholics.

Spatial Memory and a Role in Episodic Memory

The **hippocampus** appears to play a fundamental role in **spatial memory** in many animal species, including humans. For example, in a British study, the researchers asked taxi drivers to imagine their travels through the city of London, while their brain activity was monitored by positron emission tomography (PET scan). This task, which was so familiar to these subjects, caused a specific activation of their **right hippocampus**. Unlike our memory of facts and events, **spatial memory** appears to be confined to the hippocampus. And more specifically to the right hippocampus. This structure seems to be able to create a **mental map of space**, thanks to certain cells called **place cells**.

An amazing discovery in the 1970s demonstrated that a rat's **hippocampus** is a veritable **spatial map** of the environment through which it moves. Certain pyramidal neurons in **area CA1** of the **rat hippocampus** become active only when the rat is located in a specific part of its environment. There are 1 million of these "**place cells**" in area CA1 of the rat hippocampus, so that if each one is assigned a specific point in space, together they can form a very precise **cognitive map** that tells the animal where it is at any given time. Moreover, when a rat explores a new environment, it forms a new cognitive map that can be very stable, lasting weeks or months.

According to O'Keefe and Nadel (1978), the researchers who discovered the existence of these **cognitive maps**, one of their functions might be to provide a **context** to which **episodic memories** can be **attached**, in particular, their **location in space and time**. An event recorded in memory could **thus** be made **autobiographical** (situated in personal time and space). This would explain the fundamental **role** that the **hippocampus** plays in **episodic memory** in humans.

The **hippocampus** proper is **composed of** regions with tightly packed pyramidal neurons, mainly areas **CA1, CA2, and CA3.** ("CA" stands for Cornu Ammonis, or Horn of Ammon. The reference is to the ram's horns of the Egyptian god Ammon, whose shape these three areas together roughly resemble.) This is what is called the **trisynaptic circuit** or trisynaptic loop **of** the **hippocampus. Information enters this one-way loop via** the **axons of the entorhinal cortex** (part of the medial temporal cortex), known as **perforant fibers** (or the **perforant path,** because it penetrates through the subiculum and the space that separates it from the dentate gyrus, which is also in the temporal lobe, adjacent to the hippocampus; some anatomists consider it part of the hippocampus). These axons of the **perforant path** make the loop's **first connection**, **with** the **granule cells** of the **dentate gyrus**, the **first region where all sensory modalities merge together** to form representations and **memories that bind stimuli together** (Science Direct; https://www.sciencedirect.com/topics...0and%20memory.), a critical step in learning and memory.

From granule cells of the dentate, the **mossy fibres** in turn project to make the loop's **second connection**, **with** the dendrites of the **pyramidal cells** in area **CA3** of the hippocampus.

The **axons of these CA3 cells divide** into **two branches**. **One** branch forms the commissural **fibers** that project **to** the **contralateral hippocampus via** the **corpus callosum**. The **other branch** forms the Schaffer collateral pathways that make the **third connection** in the loop, **with** the cells in area **CA1** of hippocampus. It is in these synapses that the **spatial memory,** associated with the hippocampus, seems to be encoded.

This region also displays a **high propensity for long-term potentiation (LTP)**, though this same phenomenon is **also** observed **in many other parts of the hippocampus** as well as in the **cortex**.

Lastly, the **axons of** the **cells in CA1** project **to** the neurons of the **subiculum** and of the **entorhinal cortex**. The **receiving portion of** the **hippocampal formation** thus consists of the **dentate gyrus**, while the **sending portion** consists of the **subiculum**. The

axons of the **large pyramidal neurons of** the **subiculum** then **project to** the **subcortical nuclei via** the **fimbria**, a thin tract of white matter at the inner edge of the hippocampus (see Figure 10.4.8). **Lastly**, the information **returns to** the **sensory cortical areas from which it came** before it was processed by the hippocampus.

Thus, **entorhinal cortex projects strongly to dentate gyrus and hippocampus** and then **from hippocampus back to cortex** (Van Hoesen and Pandya, 1975b). Being **situated between cortex and the hippocampus, entorhinal cortex plays a critical role in memory** (Zola-Morgan et al., 1989). Even though the neurons of the hippocampus may seem like just a transit point in the establishment of long-term memory, they actually display a great deal of plasticity. This plasticity, as you have learned in previous sections, is manifested chiefly through **long-term potentiation (LTP)**, which was first discovered in the **hippocampus** but has subsequently been demonstrated in **many regions of** the **cortex**.

10.6.4 Summary

Memory can be understood in terms of types and stages. Different brain structures and circuits are involved in these different types and stages of memory. Dorsolateral prefrontal cortex appears to hold information required for reasoning and decision making processes. The prefrontal cortex is important for retention of information during working memory. Working memory has several components which interact and which have links to long-term memory. Consolidation of new declarative memories depends upon feedback loops between the cortex, the hippocampus, and other limbic structures in the medial temporal lobes. A group of these structures involved in memory, called Papez's circuit or the hippocampal/mammillothalamic tract, includes the hippocampus, cingulate cortex, mammillary bodies, and anterior thalamus. Damage to the mammillary bodies, commonly due to B1 vitamin deficiency associated with alcoholism, causes Korsakoff's amnesia. The hippocampus in rats forms spatial cognitive maps involving the use of hippocampal pyramidal neurons that function as place cells which code particular locations in the environment that the rat moves through.

10.6.5 References

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? Trends in Cognitive Science, 4, 417-423.

Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory (Vol. 8, pp. 47--89). New York: Academic Press.

Barbey, A. K., Colom, R., Paul, E. J., & Grafman, J. (2014). Architecture of fluid intelligence and working memory revealed by lesion mapping. *Brain Structure and Function*, *219* (2), 485-494.

Barbey, A. K., Koenigs, M., & Grafman, J. (2013). Dorsolateral prefrontal contributions to human working memory. *cortex*, *49* (5), 1195-1205.

Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in cognitive sciences*, *7*(12), 547-552.

Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1997). Transient and sustained activity in a distributed neural system for human working memory. *Nature*, *386* (6625), 608-611.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. *Models of working memory: Mechanisms of active maintenance and executive control*, *4*, 102-134.

Hunt, L. T., Behrens, T. E., Hosokawa, T., Wallis, J. D., & Kennerley, S. W. (2015). Capturing the temporal evolution of choice across prefrontal cortex. *Elife*, *4*, e11945.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic bulletin & review*, *9*(4), 637-671.

Koenigshofer, K. A. (2017). General Intelligence: Adaptation to Evolutionarily Familiar Abstract Relational Invariants, Not to Environmental or Evolutionary Novelty. *The Journal of Mind and Behavior*, *38* (2), 119-153.

Matzel, L. D., & Kolata, S. (2010). Selective attention, working memory, and animal intelligence. *Neuroscience & Biobehavioral Reviews*, *34* (1), 23-30.

O'keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford university press.

Van Hoesen, G. W., & Pandya, D. N. (1975). Some connections of the entorhinal (area 28) and perirhinal (area 35) cortices of the rhesus monkey. III. Efferent connections. *Brain research*, *95*(1), 39-59.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., ... & Miller, B. L. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological science*, *10*(2), 119-125.

Zola-Morgan, S., Squire, L. R., Amaral, D. G., & Suzuki, W. A. (1989). Lesions of perirhinal and parahippocampal cortex that spare the amygdala and hippocampal formation produce severe memory impairment. *Journal of Neuroscience*, *9*(12), 4355-4370.

10.6.6 Attributions

Section 10.7 adapted by Kenneth A. Koenigshofer, PhD., from The Brain from Top to Bottom; license: Copyleft, https://thebrain.mcgill.ca/flash/pop.../pop_copy.html. licensed under CC BY 4.0.

This page titled 10.6: Localization of Types and Stages of Memory is shared under a mixed license and was authored, remixed, and/or curated by Kenenth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

10.7: Hormones, Learning, and Memory

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

10.7.1 Learning Objectives

1. Describe the structures involved in hormone release
2. Discuss the role of hormones in neuromodulation mediating the effects of motivation, reward, and emotion on learning and memory

10.7.2 **Overview**

Hormone release is under the control of the pituitary which in turn is under the control of parts of the hypothalamus at the base of the brain below the thalamus. Memory and learning can be influenced by hormones which may act as neuromodulators associated with motivation, rewards, and emotions. The effects of hormones are slower to act but produce more long-lasting effects widely distributed over large regions of the brain.

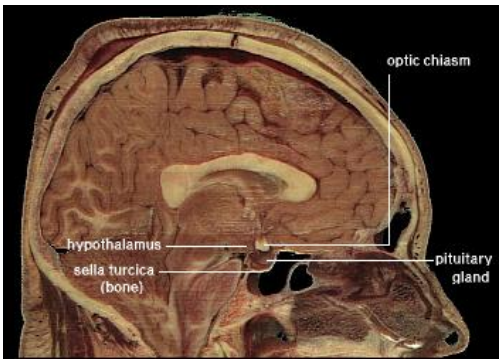


Figure **10.7.1**: Cross-section of the human head showing hypothalamus and pituitary gland. The hypothalamus control the pituitary gland, the "master gland" of the endocrine system, which releases pituitary hormones into the bloodstream which control other endocrine glands throughout the body. (Image from Wikimedia Commons; File:LocationOfHypothalamus.jpg; https://commons.wikimedia.org/wiki/F...pothalamus.jpg; by NIH; *mage is a work of the **National Institutes of Health**, part of the **United States Department of Health and Human Services**, taken or made as part of an employee's official duties. As a work of the U.S. federal government, the image is in the **public domain***).

Hormonal Influences on Memory

It is important to be aware that memory does not depend solely on a few synapses in the hippocampus or elsewhere in the brain. For example, several studies indicate that the major neuromodulation systems in the brain (such as those which use dopamine or serotonin) also greatly influence synaptic plasticity.

These neuromodulators are part of the molecular mechanisms through which factors such as motivation, rewards, and emotions can influence learning. A major source of neuromodulation is hormones. The hormonal neurons are concentrated mainly in the brainstem and the central region of the brain including the hypothalamus which regulates the pituitary gland. The neurons of the brain secrete substances that regulate the functions of the glands in the rest of the body. The main pathway through which this regulation occurs passes through the pituitary gland, a small gland that is located at the base of the brain and whose hormones influence practically all of the other glands in the body. The pituitary gland is strongly influenced by the neuromodulators produced by the hypothalamus (which are therefore classified as neurohormones). These neurons form small masses of thousands of cells, but these cells project their axons into large areas of the forebrain and the midbrain.

Just one of these neurons can therefore influence over 100,000 others through the neuromodulators that it secretes into the brain's extracellular space (rather than into a synaptic gap). Each of these groups of neurons projects its axons into large areas of the

central nervous system and thus modulates numerous behaviors, including sexual motivation and sexual behavior.

The effects of these neuromodulators take longer to become established and last longer than those of the neurotransmitters in the circuits of the brain. One reason for these differences is that many of these neuromodulator effects are mediated by “second messengers.”

Neuroscientists are getting closer to tying psychological activities in humans and animals to specific molecular processes, even though we are still far from understanding all of the influences acting on the billions of connections in our brains--synaptic connections continuously undergoing modification moment to moment throughout our lives.

10.7.3 Attributions

Adapted by Kenneth A. Koenigshofer, Ph.D., licensed under CC BY 4.0 from The Brain from Top to Bottom, https://thebrain.mcgill.ca/flash/a/a...07_cl_tra.html, license: Copyleft, https://thebrain.mcgill.ca/flash/pop.../pop_copy.html.

This page titled 10.7: Hormones, Learning, and Memory is shared under a mixed license and was authored, remixed, and/or curated by Keneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

CHAPTER OVERVIEW

11: Wakefulness and Sleep

> This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

11.1: Consciousness
11.2: What is Sleep?
11.3: Stages of Sleep
11.4: The Brain and Sleep
11.5: Theories of Sleep
11.6: Sleep Disorders

This page titled 11: Wakefulness and Sleep is shared under a mixed license and was authored, remixed, and/or curated by ASCCC OERI & Bakhtawar Bhadha (ASCCC Open Educational Resources Initiative (OERI)) .

11.1: Consciousness

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Differentiate the main theories of consciousness.
- Describe the Default Mode Network including its associated brain structures and its role in varying levels of consciousness

11.1.1 Overview

What does it mean to be conscious? Clearly when someone has fainted, they are unconscious, and when they are sitting up and solving a math problem alongside you they are conscious. But consider this, when a child I was babysitting walked into my living room one night, but refused to answer me when I asked her if everything was alright, it was my first experience with sleepwalking. Would you say she was conscious? Alternately, is there a difference between the level or kind of consciousness you experience while taking an exam, versus when you have eaten some marijuana brownies, versus when you are meditating? There are theories about whether consciousness varies on levels or types. There has also been much research looking at the brain underpinnings of these different states. So theories of consciousness as well as the Default Mode Network (brain parts involved in resting wakefulness) will be described below.

11.1.2 Theories of Consciousness

Conscious states can be thought of as global dimensions of consciousness that can modulate how we think, feel and behave (Bayne and Hohwy 2016; Bayne et al. 2016). For example, in the state of ordinary waking awareness, a wide variety of inputs can enter consciousness, and a wide variety of cognitive and behavioral capacities can be exercised. In other conscious states, however, both the range of conscious content and the range of cognitive and behavioral capacities may be curtailed. Conditions that are often associated with changes in conscious state include post-comatose disorders of consciousness (Bruno et al. 2011; Casarotto et al. 2016), sleep and drug-induced sedation (Sarasso et al. 2015), and certain pathologies of consciousness, such as epileptic absence seizures (Gloor 1986; Blumenfeld 2005; Bayne 2011). For example, Minimally Conscious State patients can track certain features of their environment (such as the presence of motion or the semantic content of simple instructions), but they lack the capacity to engage in complex forms of cognition or behavior, and they seem unable to entertain complex thoughts or ideas (Giacino et al. 2002).

The **Integrated Information Theory of Consciousness (IITC)** is one theory suggesting that consciousness can be explained in terms of levels (Tononi & Koch, 2015). However, this assumes that conscious states can be ordered along a single continuum. Indeed, using transcranial magnetic stimulation (TMS – as you learned about in chapter 2) and computerized compression of the wave patterns that result, Casali et al (2013) figured out a way to objectively measure different states of consciousness including healthy subjects who are awake, dreaming, in NREM sleep, and those who are sedated, as well as in patients who had emerged from coma. But these are still not quite levels, and might indeed be thought of as qualitatively separate states of consciousness.

On the other hand, Bayne and Carter (2018) claim that consciousness ought to be construed in multidimensional terms, and that conscious states can differ from each other along multiple dimensions. For example, one of their central claims is that although the psychedelic state is distinct from the state of ordinary waking awareness, it is neither ‘higher’ nor ‘lower’ than the state of ordinary waking awareness.

The **Global Workspace Theory of Consciousness** (Dahaene et al., 2017) holds that changes in consciousness only involve ‘vigilance’ and ‘wakefulness’ where there is global availability of content. But this theory does not consider the multidimensional nature of consciousness.

Consider the state of consciousness elicited by psychedelic drugs. The psychedelic state involves a state that certainly differs from that associated with ordinary waking awareness, but there is no reason to think that individuals in the psychedelic state are more conscious (or, for that matter, less conscious) than individuals who are not in it. Why do psychedelics increase the vividness,

complexity and possibly also the bandwidth of sensory experience? What explains the systematic effects that psychedelics have on the experience of unity across a wide range of domains (e.g. time, space and the self)? Addressing these questions have helped researchers to identify the various dimensions that structure consciousness. Additionally, as presented in the chapter 6 on drugs, since much is known about the neurochemical effects of psychedelics, using this knowledge in both animal models and human imaging might provide a window into understanding the neural basis of consciousness.

11.1.3 Default Mode Network

Raichle et al coined the term, **Default Mode Network (DMN)** in 2001 to describe how the brain is constantly active even when not doing anything actively (See Figure **11.1.1**). The regions in this network exhibit decreased activation when engaging in goal-oriented or attention-demanding tasks, and therefore facilitate a "default" functional state within the brain. (Raichle et al., 2001; Figure **11.1.1**).

The brain is constantly active even when it is not engaged in a task. In the restful yet awake state, the default mode network is primarily activated. It used to also be called the task negative network – meaning that it was deactivated when we are engaged in a particular task. The main parts in this network include the medial prefrontal cortex, posterior cingulate cortex/precuneus and angular gyrus

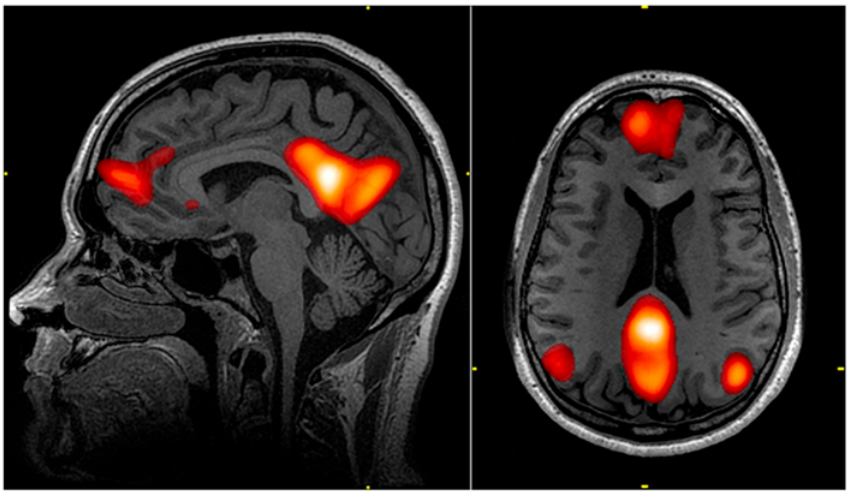


Figure **11.1.1**: FMRI of Default Mode Network from Functional MRI in the investigation of blast-related traumatic brain injury by Graner, Oakes, French and Riedy in the Public Domain.

This network is active when one is in a state of wakeful rest, for example when one is daydreaming, relaxing, or meditating.

DMN activation is modulated in different types of consciousness changes. Table **11.1.1** summarizes studies showing these connections.

Table **11.1.1**: Different studies have shown the ways in which the Default Mode Network (DMN) is affected by psychological events.

Consciousness change	Effect on DMN	Research credit
Acupuncture	DMN connectivity reduced in pain response	Huang et al, 2012
Meditation	Structural changes in areas of precuneus	Fox et al, 2014
Resting wakefulness	Increase in DMN	Picchioni et al, 2013
Sleep deprivation	Decrease in connectivity	McKenna & Eyler, 2012
Use of psychedelics	reduced blood flow to precuneus and medial prefrontal cortex	Carhart-Ellis et al, 2012
Deep brain stimulation	used to rebalance the restful brain structures	Kringelbach et al, 2011
Psychotherapy	helps stabilize the DMN in PTSD sufferers	Sripada et al, 2012
Attention training techniques	help increase connectivity in DMN	Kowalski et al, 2020
Antidepressant use	improves abnormalities in DMN in PTSD sufferers	Akiki et al, 2017
Physical activity and exercise	alters DMN	Shao et al, 2019; Muraskin et al, 2016; Voss et al, 2019

11.1.4 Summary

There has been a lot of research into what consciousness means philosophically and how that ties in with brain activity. Clearly, when one is asleep there is a difference in the awareness of stimuli, thought and function behaviorally. Biological psychologists have attempted to establish whether the differences in consciousness experienced while awake, asleep, under the influence of different drugs, while meditating, and in comatose states are qualitative or quantitative differences. Images of the brain in these different conditions have helped scientists come up with theories - and arguments - for each of these positions

11.1.5 References

Akiki, T. J., Averill, C. L., & Abdallah, C. G. (2017). A Network-Based Neurobiological Model of PTSD: Evidence From Structural and Functional Neuroimaging Studies. *Current psychiatry reports*, *19*(11), 81. https://doi.org/10.1007/s11920-017-0840-4

Bayne T. (2011). The presence of consciousness in absence seizures. Behav Neurol, 24(1:47-53. doi: 10.3233/BEN-2011-0318

Bayne, T., & Hohwy, J. (2016). Modes of consciousness. In: Sinnott-Armstrong W. (ed.), Finding Consciousness: The Neuroscience, Ethics and Law of Severe Brain Damage. New York: Oxford University Press, 57–80.

Bayne, T., Hohwy, J., & Owen, A.M. (2016). Are there levels of consciousness? Trends Cogn Sci 20, 405–413.

Blumenfeld H. (2005). Consciousness and epilepsy: why are patients with absence seizures absent? Prog Brain Res. 150, 271-286. doi: 10.1016/S0079-6123(05)50020-7.

Bruno, M. A., Vanhaudenhuyse, A., Thibaut, A., Moonen, G., & Laureys, S. (2011). From unresponsive wakefulness to minimally conscious PLUS and functional locked-in syndromes:

recent advances in our understanding of disorders of consciousness. J Neurol. 258(7),1373-1384. doi: 10.1007/s00415-011-6114-x.

Carhart-Harris, R. L., Erritzoe, D., Williams, T., Stone, J. M., Reed, L. J., Colasanti, A., Tyacke, R. J., Leech, R., Malizia, A. L., Murphy, K., Hobden, P., Evans, J., Feilding, A., Wise, R. G., & Nutt, D. J. (2012). Neural correlates of the psychedelic state as determined by fMRI studies with psilocybin. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(6), 2138–2143. https://doi.org/10.1073/pnas.1119598109

Casali, A.G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, P., Casarotto, S., Bruno, M., Laureys, S., Tononi, G., & Massimini, M. (2013) A theoretically based index of consciousness independent of sensory processing and behavior. Sci Transl Med, 14(5),198ra105. doi: 10.1126/scitranslmed.3006294.

Casarotto, S., Comanducci, A., Rosanova, M., Sarasso, S., Fecchio, M., Napolitani, M., Pigorini, A., Casali, A., Trimarchi, P. D., Boly, M., Gosseries, O., Bodart, O., Curto, F., Landi, C., Mariotti, M., Devalle, G., Laureys, S., Tononi, G., & Massimini, M. (2016). Stratification of unresponsive patients by an independently validated index of brain complexity. Ann Neurol. 80(5), 718-729. doi: 10.1002/ana.24779.

Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness, and could machines have it? *Science, 358*, 486-492. doi: 10.1126/science.aan8871.

Fox, K. C. R, Nijeboer, S., Dixon, M. L., Floman, J. L., Ellamil, M., Rumak, S. P., Sedlmeier, P., & Christoff, K. (2014). Is meditation associated with altered brain structure? A systematic review and meta-analysis of morphometric neuroimaging in meditation practitioners. Neuroscience & Biobehavioral Reviews, 43, 48-73. https://doi.org/10.1016/j.neubiorev.2014.03.016.

Giacino, J.T., Ashwal, S., Childs, N., Cranford, R., Jennett, B., Katz, D.I., Kelly, J.P., Rosenberg, J.H., Whyte, J., Zafonte, R.D., & Zasler, N.D. (2002). The minimally conscious state: definition and diagnostic criteria. Neurology,58(3):349-353. doi: 10.1212/wnl.58.3.349.

Gloor P. (1986). Consciousness as a neurological concept in epileptology: a critical review. Epilepsia. 27 Suppl 2:S14-26. doi: 10.1111/j.1528-1157.1986.tb05737.x.

Graner, J., Oakes, T. R., French, L. M. & Riedy, G. (2013). Functional MRI in the investigation of blast-related traumatic brain injury. Frontiers in Neurology, 4, 16, 1-18. https://doi.org/10.3389/fneur.2013.00016

Huang, W., Pach, D., Napadow, V., Park, K., Long, X., Neumann, J., Maeda, Y., Nierhaus, T., Liang, F., & Witt, C. M. (2012). Characterizing acupuncture stimuli using brain imaging with FMRI--a systematic review and meta-analysis of the literature. *PloS one*, *7*(4), e32960. https://doi.org/10.1371/journal.pone.0032960

Kowalski, J., Wierzba, M., Wypych, M., Marchewka, A., & Dragan, M. (2020). Effects of attention training technique on brain function in high- and low-cognitive-attentional syndrome individuals: Regional dynamics before, during, and after a single session of ATT. Behaviour Research and Therapy, 132, 103693. https://doi.org/10.1016/j.brat.2020.103693.

Kringelbach, M., Green, A., & Aziz T. (2011). Balancing the Brain: Resting State Networks and Deep Brain *Stimulation. Frontiers in Integrative Neuroscience*, *5*. doi: 10.3389/fnint.2011.00008

McKenna, B. S., & Eyler, L. T. (2012). Overlapping prefrontal systems involved in cognitive and emotional processing in euthymic bipolar disorder and following sleep deprivation: a review of functional neuroimaging studies. *Clinical psychology review*, *32*(7), 650–663. https://doi.org/10.1016/j.cpr.2012.07.003

Muraskin, J., Dodhia, S., Lieberman, G., Garcia, J. O., Verstynen, T., Vettel, J. M., Sherwin, J., & Sajda, P. (2016). Brain dynamics of post-task resting state are influenced by expertise: Insights from baseball players. *Human brain mapping*, *37*(12), 4454–4471. https://doi.org/10.1002/hbm.23321

Picchioni, D., Duyn, J. H., Horovitz, S. G. (2013). Sleep and the functional connectome. NeuroImage, 80, 387-396. https://doi.org/10.1016/j.neuroimage.2013.05.067.

Raichle, M. E., MacLeod, A. M., Snyder, A.Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. Proceedings of the National Academy of Sciences 98 (2) 676-682. doi: 10.1073/pnas.98.2.676

Sarasso, S., Boly, M., Napolitani, M., Gosseries, O., Charland-Verville, V., Casarotto, S., Rosanova, M., Casali, A. G., Brichant, J. F., Boveroux, P., Rex, S., Tononi, G., Laureys, S., & Massimini, M. (2015). Consciousness and complexity during unresponsiveness induced by Propofol, Xenon, and Ketamine. Curr Biol. 25(23), 3099-3105. doi: 10.1016/j.cub.2015.10.014.

Shao, M., Lin, H., Yin, D., Li, Y., Wang, Y., Ma, J., Yin, J., & Jin, H. (2019). Learning to play badminton altered resting-state activity and functional connectivity of the cerebellar sub-regions in adults. *PloS one*, *14*(10), e0223234. https://doi.org/10.1371/journal.pone.0223234

Sripada, R. K., King, A. P., Welsh, R. C., Garfinkel, S. N., Wang, X., Sripada, C. S., & Liberzon, I. (2012). Neural dysregulation in posttraumatic stress disorder: evidence for disrupted equilibrium between salience and default mode brain networks. *Psychosomatic medicine*, *74*(9), 904–911. https://doi.org/10.1097/PSY.0b013e318273bf33

Voss, M. W., Soto, C., Yoo, S., Sodoma, M., Vivar, C., & van Praag, H. (2019). Exercise and hippocampal memory systems. Trends in Cognitive Sciences, 23(4), 318-333. https://doi.org/10.1016/j.tics.2019.01.006

11.1.6 Attributions

Adapted by Bakhtawar Bhadha from Tim Bayne, Olivia Carter, Dimensions of consciousness and the psychedelic state, *Neuroscience of Consciousness*, Volume 2018, Issue 1, 2018, niy008, https://doi.org/10.1093/nc/niy008 Licensed CC-BY NC

Adapted by Bakhtawar Bhadha from Tononi G, Koch C. 2015 Consciousness: here, there and everywhere? Phil. Trans. R. Soc. B 370: 20140167. http://dx.doi.org/10.1098/rstb.2014.0167 Licensed CC-BY

Functional MRI in the investigation of blast-related traumatic brain injury by Graner, Oakes, French and Riedy in the Public Domain.

This page titled 11.1: Consciousness is shared under a mixed license and was authored, remixed, and/or curated by ASCCC OERI & Bakhtawar Bhadha (ASCCC Open Educational Resources Initiative (OERI)) .

11.2: What is Sleep?

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Evaluate the overall complex nature of sleep, rather than simply the absence of wakefulness.
2. Demonstrate a general understanding of the biological nature of sleep.
3. Evaluate the importance of sleep, its functions and the negative ramifications of sleep deprivation.

11.2.1 Overview

When you are asleep are you unconscious? Why do all animals appear to need sleep, some at great expense and danger? Is it true that we "need" sleep every night, or are we just wasting time (about a third of our lives if we follow the 8 hours a day recommendation)? Really all we are doing is laying down and turning off the lights, right? And our brain just shuts off for that time, right?

The nature of sleep as a biological process that is generated by the brain will be discussed in this chapter. In this section, some ideas about how these discoveries came about are contrasted against people's general belief that sleep is just something that happens when one is not doing anything else.

11.2.2 What is Sleep?

Sleep is a complex biological process. While you are sleeping, you may appear unconscious (though there is certainly a difference between being asleep and having fainted), but your brain and body functions are still active. They are doing a number of important jobs that help you stay healthy and function at your best. So when you don't get enough quality sleep, it does more than just make you feel tired. It can affect your physical and mental health, thinking, and daily functioning.

Sleep is an important part of your daily routine—you spend about one-third of your time doing it. Quality sleep – and getting enough of it at the right times -- is as essential to survival as food and water.

Everyone needs sleep, but its full biological purpose remains a mystery. Sleep affects almost every type of tissue and system in the body – from the brain, heart, and lungs to metabolism, immune function, mood, and disease resistance. Research shows that a chronic lack of sleep, or getting poor quality sleep, increases the risk of disorders including high blood pressure, cardiovascular disease, diabetes, depression, and obesity.

Sleep is important to a number of brain functions, including how nerve cells (neurons) communicate with each other. In fact, your brain and body stay remarkably active while you sleep. Recent findings suggest that sleep plays a housekeeping role that removes toxins in your brain that build up while you are awake (Xie et al, 2013).

Without sleep you can’t form or maintain the pathways in your brain that let you learn and create new memories, and it’s harder to concentrate and respond quickly.

Sleep is a complex and dynamic process that affects how you function in ways scientists are now beginning to understand.

11.2.3 Functions of sleep

11.2.3.1 Memory consolidation

Another idea about why we sleep is that sleeping reinforces learning and memory, while at the same time, helping us to forget or to clear stores of unneeded memories. During the course of a day we are inundated with experiences, some of which should be remembered while others need not be. Perhaps sleep aids in rearranging all of the experiences and thoughts from the day so that those that are important are stored and those that are not are discarded. People who get plenty of deep NREM sleep in the first half of the night and REM sleep in the second half improve their ability to perform spatial tasks. This suggests that the full night's sleep plays a role in learning—not just one kind of sleep or the other.

Other research has also shown that activity in the hippocampus during REM sleep supports the idea that dreams serve a memory consolidation function. Further the activity of the inferior parietal lobule, a part of the cortex that conveys experiences to memory, decreases during REM sleep, which probably helps to explain why we have so much trouble in remembering our dreams. There is also a lot of activity noted in the limbic regions – particularly the amygdala, probably in relation to the emotional quality of dreams that we report.

Research with rats, cats, songbirds, and humans have shown that multiple stages in sleep increase memory consolidation. Rasch and Born (2013) reviewed the multiple genetic, neurological, and chemical bases by which memory function is affected during sleep.

Studies show that a good night's sleep improves learning. Whether you're learning math, how to play the piano, how to perfect your golf swing, or how to drive a car, sleep helps enhance your learning and problem-solving skills. Sleep also helps you pay attention, make decisions, and be creative.

11.2.3.2 Energy Conservation

This theory states that we sleep to conserve energy and is based on the fact that the metabolic rate is lower during sleep. The theory predicts that total sleep time and NREM sleep time will be proportional to the amount of energy expended during wakefulness. Support for this theory is derived from several lines of evidence. For example, NREM and REM sleep states are found only in **endothermic** animals (that is, those that expend energy to maintain body temperature). Species with greater total sleep times generally have higher core body temperatures and higher metabolic rates. Consider also that NREM sleep time and total sleep time decrease in humans, with age, as do body and brain metabolism. In addition, infectious diseases tend to make us feel sleepy. This may be because molecules called **cytokines**, which regulate the function of the immune system, are powerful sleep inducers. It may be that sleep allows the body to conserve energy and other resources, which the immune system may then use to fight the infection.

11.2.3.3 Brain Development

This proposed function of sleep is related to REM sleep, which occurs for prolonged periods during fetal and infant development. This sleep state may be involved in the formation of brain synapses.

Children and teens who are sleep deficient may have problems getting along with others. They may feel angry and impulsive, have mood swings, feel sad or depressed, or lack motivation. They also may have problems paying attention, and they may get lower grades and feel stressed.

11.2.3.4 Discharge of emotions

Perhaps dreaming during REM sleep provides a safe discharge of emotions. As protection to ourselves and to a bed partner, the muscular paralysis that occurs during REM sleep does not allow us to act out what we are dreaming. Additionally, activity in brain regions that control emotions, decision making, and social interactions is reduced during sleep. Perhaps this provides relief from the stresses that occur during wakefulness and helps maintain optimal performance when awake.

11.2.4 Sleep Deprivation and Deficiency

11.2.4.1 Why Is Sleep Important?

Whichever theory is most effective, it is clear that sleep plays a vital role in good health and well-being throughout your life. Getting enough quality sleep at the right times can help protect your mental health, physical health, quality of life, and safety.

The way you feel while you're awake depends in part on what happens while you're sleeping. During sleep, your body is working to support healthy brain function and maintain your physical health. In children and teens, sleep also helps support growth and development.

The damage from sleep deficiency can occur in an instant (such as a car crash), or it can harm you over time. For example, ongoing sleep deficiency can raise your risk for some chronic health problems. It also can affect how well you think, react, work, learn, and get along with others.

11.2.4.2 Healthy Brain Function and Emotional Well-Being

As seen above, sleep helps your brain work properly. While you're sleeping, your brain is preparing for the next day. It's forming new pathways to help you learn and remember information.

Studies also show that sleep deficiency alters activity in some parts of the brain. If you're sleep deficient, you may have trouble making decisions, solving problems, controlling your emotions and behavior, and coping with change. Sleep deficiency also has been linked to depression, suicide, and risk-taking behavior.

11.2.4.3 Physical Health

Sleep plays an important role in your physical health. For example, sleep is involved in healing and repair of your heart and blood vessels. Ongoing sleep deficiency is linked to an increased risk of heart disease, kidney disease, high blood pressure, diabetes, and stroke.

- Sleep deficiency also increases the risk of obesity. For example, one study of teenagers showed that with each hour of sleep lost, the odds of becoming obese went up. Sleep deficiency increases the risk of obesity in other age groups as well.
- Sleep helps maintain a healthy balance of the hormones that make you feel hungry (ghrelin) or full (leptin). When you don't get enough sleep, your level of ghrelin goes up and your level of leptin goes down. This makes you feel hungrier than when you're well-rested.
- Sleep also affects how your body reacts to insulin, the hormone that controls your blood glucose (sugar) level. Sleep deficiency results in a higher than normal blood sugar level, which may increase your risk for diabetes.
- Sleep also supports healthy growth and development. Deep sleep triggers the body to release the hormone that promotes normal growth in children and teens. This hormone also boosts muscle mass and helps repair cells and tissues in children, teens, and adults. Sleep also plays a role in puberty and fertility.
- Your immune system relies on sleep to stay healthy. This system defends your body against foreign or harmful substances. Ongoing sleep deficiency can change the way in which your immune system responds. For example, if you're sleep deficient, you may have trouble fighting common infections.

11.2.4.4 Daytime Performance and Safety

Getting enough quality sleep at the right times helps you function well throughout the day. People who are sleep deficient are less productive at work and school. They take longer to finish tasks, have a slower reaction time, and make more mistakes. After several nights of losing sleep—even a loss of just 1–2 hours per night—your ability to function suffers as if you haven't slept at all for a day or two.

Lack of sleep also may lead to microsleep. Microsleep refers to brief moments of sleep that occur when you're normally awake. You can't control microsleep, and you might not be aware of it. For example, have you ever driven somewhere and then not remembered part of the trip? If so, you may have experienced microsleep. Even if you're not driving, microsleep can affect how you function. If you're listening to a lecture, for example, you might miss some of the information or feel like you don't understand the point. In reality, though, you may have slept through part of the lecture and not been aware of it.

11.2.4.5 Who Is at Risk for Sleep Deprivation and Deficiency?

Sleep deficiency, which includes sleep deprivation, affects people of all ages, races, and ethnicities.

Sleep deficiency can cause problems with learning, focusing, and reacting. You may have trouble making decisions, solving problems, remembering things, controlling your emotions and behavior, and coping with change. You may take longer to finish tasks, have a slower reaction time, and make more mistakes.

The signs and symptoms of sleep deficiency may differ between children and adults. Children who are sleep deficient might be overly active and have problems paying attention. They also might misbehave, and their school performance can suffer.

Sleep-deficient children may feel angry and impulsive, have mood swings, feel sad or depressed, or lack motivation.

You may not notice how sleep deficiency affects your daily routine. A common myth is that people can learn to get by on little sleep with no negative effects. However, research shows that getting enough quality sleep at the right times is vital for mental health, physical health, quality of life, and safety.

11.2.5 Summary

It is clear that sleep is a complex biological process that all humans, indeed all animals, engage in. Several brain regions are involved in sleep and wakefulness processes, and sleep serves complex functions that continue to be researched. The seemingly simple idea that we put on our pajamas, close our eyes and nod off to sleep involves far more complexity than appears on the surface.

11.2.6 References

Rasch, B. & Born, J. (2013). About sleep's role in memory. *American Psychological Society: Physiological Reviews, 93*(2). https://doi.org/10.1152/physrev.00032.2012

Xie, L., Kang, H., Xu, Q., Chen, M.J., Liao, Y. Thiyagarajan, M., O'Donnell, J., Christensen, D.J., Nicholson, C., Iliff, J.J., Takano, T., Deane, R., Nedergaard, M. (2013). Sleep drives metabolite clearance from the adult brain. *Science*, *342* (6156), 373-377. doi:10.1126/science.1241224

11.2.7 Attributions

- The Brain structures that wake you up and put you to sleep by Bruno Dubuc under a Copyleft license.
- Sleep disorders: What is sleep by Medline Plus, National Library of Medicine under public domain
- Brain basics: Understanding sleep by NIH:NINDS under public domain
- Information about Sleep by NIH in the public domain
- Sleep Deprivation and Deficiency: National Heart, Lung, and Blood Institute; National Institutes of Health; U.S. Department of Health and Human Services. Public Domain

---

This page titled 11.2: What is Sleep? is shared under a mixed license and was authored, remixed, and/or curated by ASCCC OERI & Bakhtawar Bhadha (ASCCC Open Educational Resources Initiative (OERI)) .

11.3: Stages of Sleep

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Explain the different stages of sleep.
2. Describe the nature of REM sleep, its biological underpinnings and possible functions.

11.3.1 Overview

While watching someone sleep might appear as if they are seemingly unconscious with eyes closed for an extended period of time, in fact brain wave recordings show that the brain is cycling through several different stages as an individual sleeps. These stages each have distinct characteristics of brain activity and conscious experience and are divided into two main types known as Rapid Eye Movement (REM) and Non-REM. This section will explore these characteristics and some of the peculiar and unexpected features of REM sleep, in particular.

11.3.2 The different stages of sleep

From a behavioral standpoint, sleep is defined by four criteria:

1. reduced motor activity
2. diminished responses to external stimuli
3. stereotyped posture (in humans, lying down with eyes closed)
4. relatively ready reversibility.

These criteria distinguish sleep from hibernation and unconscious states, such as coma.

Sleep is a highly organized sequence of events that follows a regular, cyclic program each night. Polysomnograms - literally "many sleep measures" - are used to measure different body functions in a sleep lab (See Figure to see where electrodes are placed for each of the measurements). Thus, the EEG (electroencephalogram) is the technique used to measure sleep most readily and often (as discussed in chapter 2). For the EEG, electrodes are placed on the head and the brain's electrical activity is amplified and recorded). EMG (electromyograph that measures muscle tension), and EOG (electrooculogram that measures eye movement) patterns also change in predictable ways several times during a single sleep period.

NREM sleep is divided into four stages according to the **amplitude** and **frequency** of brain wave activity. In general, the EEG pattern of NREM sleep is slower, often more regular, and usually of higher voltage than that of wakefulness. As sleep gets deeper, the brain waves get slower and have greater amplitude. NREM Stage 1 is very light sleep; NREM Stage 2 has special brain waves called **sleep spindles** and **K complexes**; NREM Stages 3 and 4 show increasingly more high-voltage slow waves. In NREM Stage 4, it is extremely hard to be awakened by external stimuli. The muscle activity of NREM sleep is low, but the muscles retain their ability to function. Eye movements normally do not occur during NREM sleep, except for very slow eye movements, usually at the beginning. With the exception of brain activity, the body's general physiology during these stages is fairly similar to the wake state.

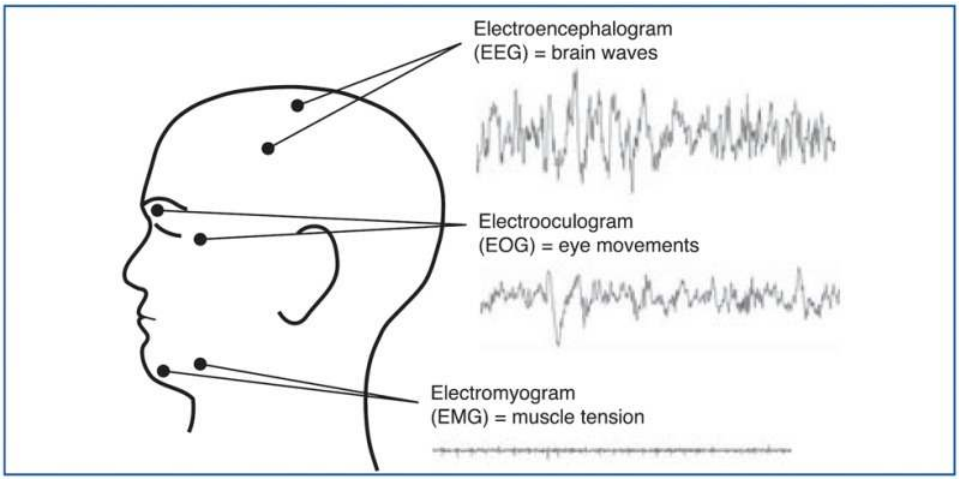


Figure **11.3.1**: Placement of electrodes to determine EEG, EOG, and EMG. (NIH in the public domain)

From the time you fall asleep to the time you reach the deepest non-REM sleep, about 1½ hours later, the amplitude of these waves increases continuously, while their frequency diminishes correspondingly. Compared with wakefulness and with REM (rapid eye movement) sleep, non-REM sleep is characterized by an electroencephalogram (EEG) in which the waves have a greater amplitude and a lower frequency.

Hypnograms were developed to summarize the voluminous chart recordings of electrical activities (EEG, EOG, and EMG) collected during a night's sleep. Hypnograms provide a simple way to display information originally collected on many feet of chart paper or stored as a large file on a computer. The figure illustrates a hypnogram. (Of course, it is important to keep in mind that individuals' hypnograms might be much less uniform than this relative "ideal" presented here.)

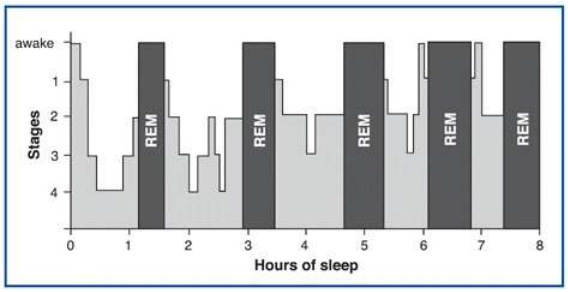


Figure **11.3.2**: A typical hypnogram from a young, healthy adult. Light-gray areas represent non–rapid eye movement (NREM) sleep (NIH in the public domain)

We can make several observations about the hypnogram in this figure. First, the periods of NREM and REM sleep alternate during the night. Second, the deepest stages of NREM sleep occur in the first part of the night. Third, the episodes of REM sleep are longer as the night progresses. This hypnogram also indicates two periods during the night when the individual awakened (at about six and seven hours into the night).

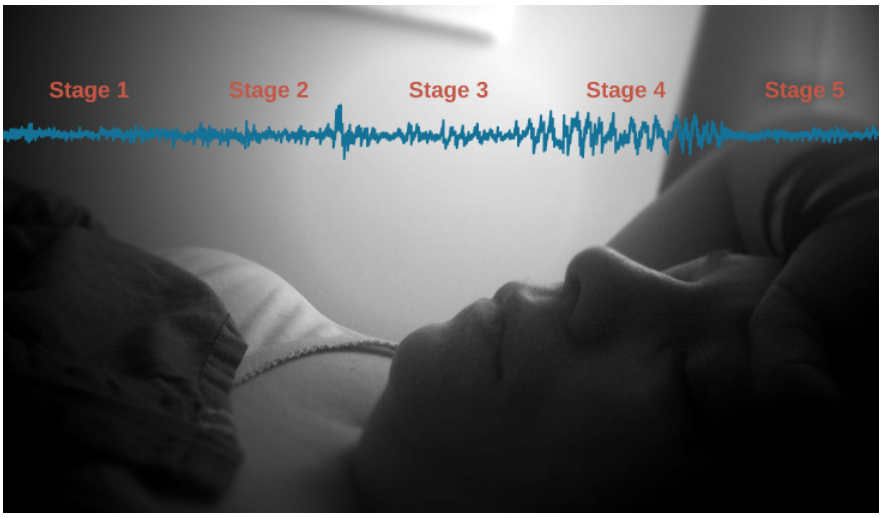
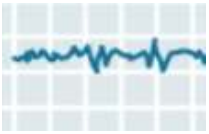
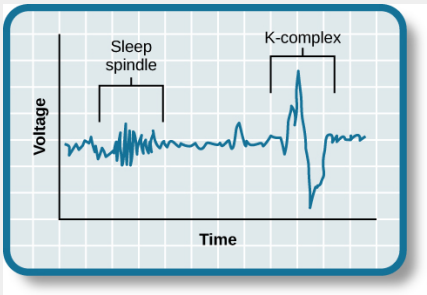
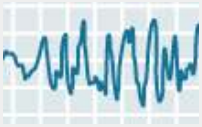


Figure **11.3.3**: Note the amplitude and frequency of the wave patterns during different stages of sleep (Stage 5 indicates REM sleep). (CC BY 4.0; Lumen Learning)

The four types of brain waves (shown in Figure **11.3.3**, and others discussed below, are important criteria that have been used to define four distinct stages of non-REM (NREM) sleep. Obviously, falling into a deeper and deeper sleep as the night progresses is actually a gradual, continuous process, but these four stages still provide a convenient means of describing the relative depth of NREM sleep.

Table 11.3.1 Stages of sleep

Stage	EEG	Description
Stage 1		Stage 1 non-REM sleep begins when you first lie down and close your eyes. After a few sudden, sharp muscle contractions in the legs, the muscles relax. Then, as you continue falling asleep, the rapid beta waves of wakefulness are replaced by the slower alpha waves of someone who is relaxed with their eyes closed. Soon, the even slower theta waves begin to emerge. Though your reactions to stimuli from the outside world diminish, Stage 1 is still the phase of sleep from which it is easiest to wake someone up. In experiments where people are awakened from Stage 1 sleep and asked about their state of consciousness, they usually report that they had just fallen asleep or had been in the process of doing so. They also often report having had stray thoughts and short dreams. Each period of Stage 1 sleep generally lasts 3 to 12 minutes,
Stage 2		Stage 2 non-REM sleep is a stage of light sleep in which the frequency of the EEG trace decreases further while its amplitude increases. The theta waves characteristic of Stage 2 sleep are interrupted by occasional series of high-frequency waves known as sleep spindles. These bursts of activity have a frequency of 8 to 14 Hz and an amplitude of 50 to 150 µV. Sleep spindles generally last 1 to 2 seconds. They are generated by interactions between thalamic and cortical neurons. During Stage 2 sleep, the EEG trace may also show a fast, high-amplitude wave form called a K-complex. The K-complex seems to be associated with brief awakenings, often in response to external stimuli. People in Stage 2 sleep are unlikely to react to a light or a noise, unless it is extremely bright or loud. It is still possible to awaken them, even if they then report that they were really sleeping during the 10 to 20 minutes that this stage lasts during the earliest of the night's sleep cycles. But because people go through Stage 2 sleep several times during the cycles in a night, this is the stage in which adults spend the greatest proportion of their sleep–nearly

Stage	EEG	Description
		50% of the total time that they sleep each night.
Stage 3		Stage 3 non-REM sleep marks the passage from moderately to truly deep sleep. Delta waves appear and soon account for nearly half of the waves in the EEG trace. Sleep spindles and K-complexes still occur, but less often than in Stage 2. The greater activity observed in the electro-oculogram (EOG) trace during stages 3 and 4 reflects the greater amplitude of EEG activity in the prefrontal areas, rather than movements of the eyes. Stage 3 lasts about 10 minutes during the first sleep cycle of the night but accounts for only about 7% of a total night's sleep. During Stage 3, the muscles still have some tonus, and sleepers show very little response to external stimuli unless they are very strong or have a special personal meaning (for example, when someone calls your name, or when a baby cries within earshot of its mother).
Stage 4		Stage 4 non-REM sleep is the deepest, the one in which we sleep the most soundly. The EEG trace is dominated by delta waves, and overall neuronal activity is at its lowest. The brain's temperature is also at its lowest, and breathing, heart rate, and blood pressure are all reduced under the influence of the parasympathetic nervous system. In adults, Stage 4 lasts about 35 to 40 minutes during the first sleep cycle of the night; it accounts for 15 to 20% of total sleep time in young adults. The muscles still have their tonus, and some movements of the arms, legs, and trunk are possible. This is the stage of sleep that accomplishes most of the body's repair work and from which it is most difficult to wake someone up. This is also the stage of sleep in which children may have episodes of somnambulism (sleepwalking) and night terrors.

11.3.3 REM/NREM Sleep Chemicals and Brainstem Structures

In addition to maintaining wakefulness, several of the nuclei of the two pathways from the brainstem and the pons to the cortex use acetylcholine and glutamate as neurotransmitters and are partly responsible for the cortical activation that occurs during REM sleep.

The midbrain reticular formation projects massively into the thalamic nuclei, which in turn influence the entire cortex. The role of this formation is to desynchronize the cortex in the broad sense, thus facilitating not only wakefulness but REM sleep as well. Formerly known as the ascending activating reticular system, it is now regarded simply as part of the wakefulness network.

11.3.3.1 Brain structures involved in REM

Rapid eye movement (REM) sleep is a distinct, homeostatically controlled brain state characterized by an activated electroencephalogram (EEG) in combination with paralysis of skeletal muscles and is associated with vivid dreaming.

In 1953, Aserinsky and Kleitman first reported the existence of rapid eye movement (REM) sleep in humans as a periodically recurring brain state marked by a low amplitude electroencephalogram (EEG) and rapid eye movements (Aserinsky and Kleitman, 1953). Soon afterward, Kleitman and Dement found that the EEG during REM sleep resembles that during alert waking and showed that REM sleep coincides with periods of vivid dreaming (Dement and Kleitman, 1957). Two years later, Jouvet discovered in cats that the activated EEG during REM sleep is associated with a complete paralysis of skeletal muscles, reflected in a flat electromyogram (EMG), and therefore coined the term paradoxical sleep (Jouvet and Michel, 1959). Besides these defining properties in EEG and EMG, REM sleep is characterized by further striking neurophysiological and behavioral features, including high-amplitude theta oscillations in the hippocampus, muscle twitches, autonomic and respiratory activation, an elevated arousal threshold and bursts of large waves in the local field potential (LFP), called PGO-waves (P= pons, G= (lateral) geniculate nucleus, O= occipital cortex), that originate in the pons and propagate to the lateral geniculate nucleus, occipital cortex, and other brain areas (Datta, 1997; Karashima et al., 2010).

The fact that the prefrontal cortex is relatively silent in REM falls in line with the fact that our dream content is often bizarre, illogical and socially inappropriate. Also the anterior cingulate gyrus which plays a role in attention and motivation is quite involved in REM sleep. And it appears that certain nuclei in the pons help trigger REM.

The amygdala is one of the parts of the brain that is most active during REM sleep, but this state is actually generated deep in the brainstem. However, the set of cortical and limbic structures involved in REM sleep do not just passively submit to orders issued by the brainstem. On the contrary, the particular kind of dream images associated with REM sleep are the result of a dynamic interaction between certain key structures in the brainstem and the rest of the brain (See Figure **11.3.4**).

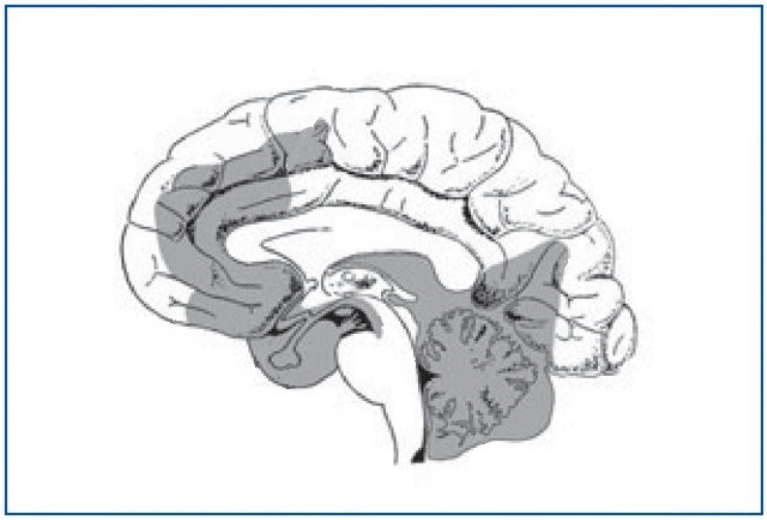


Figure **11.3.4**: Areas of brain active during REM sleep and dreaming. (Adapted from NIH in the public domain)

11.3.3.2 REM and dreams

It was shown by 1950s that the cortex is as active in REM as when someone is awake – which is one reason why REM is also called paradoxical sleep.

In humans, sleep is punctuated by REM (rapid eye movement) sleep about every 90 min (Ohayon et al., 2004). This is when most dreaming occurs (Hobson, 2009). Although some forms of dreaming can occur during non-REM sleep, such dreams are quite different from REM dreams; non-REM dreams usually are related to plans or thoughts, and they lack the visual vividness and

hallucinatory and delusory components of REM dreams (Roffwarg et al., 1962; Nielsen, 2000). Rapid eye movement during sleep is thought to be associated with visual experience of dreaming (Andrillon et al., 2015).

It was also found using MRIs that while the primary visual cortex was relatively silent in REM (clearly their eyes are closed and therefore they aren't seeing anything), the secondary visual areas of the cortex, that interpret and analyze visual information, are relatively active in REM state. This is consistent with the fact that when we are awakened from REM sleep we often report highly elaborate visual dream scenes.

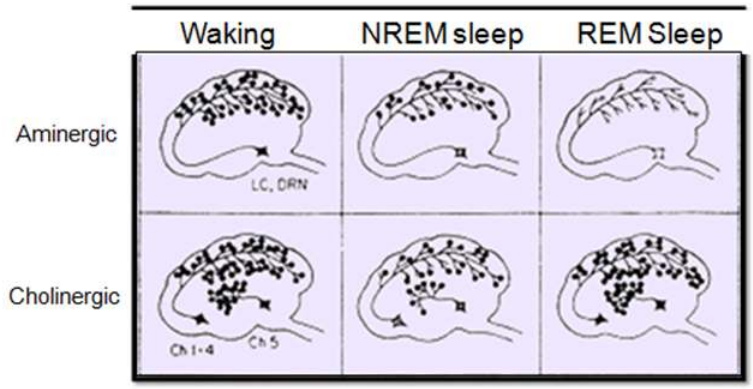


Figure **11.3.5**: The numbers of black dots are proportional to the activity of the aminergic and cholinergic systems during the various stages of wakefulness and sleep. REM sleep (paradoxical sleep) characteristics by Dr. Mohammad Torabi Nami and Dr. Khsoro Sadeghniaat. Licensed CC BY 3.0 via Webmed Central

11.3.3.3 Neurochemical bases of REM

REM is a combination of increase in cholinergic (acetylcholine) activity (in pontine tegmentum/pons and midbrain) and decrease in aminergic (monoamines - serotonin and norepinephrine) activity (in dorsal raphe nucleus and locus coeruleus).

That a sudden elevation in the activity of the cholinergic neurons of the pons is necessary for the onset of REM sleep may seem somewhat strange, because the brain's acetylcholine systems are known for their role in wakefulness. How then can this system associated with wakefulness be activated during one of the phases of sleep? The answer probably lies in the simultaneous reduction in the activity of two other nuclei that produce other wakefulness neurotransmitters, a reduction that is just as necessary for REM sleep as the increase in cholinergic activity. The two nuclei in question are the dorsal raphe nucleus, a group of serotonergic neurons, and the locus coeruleus, a group of noradrenergic neurons, and both of them are located in the rostral portion of the pons.

REM sleep cannot begin unless all activity in these two main aminergic systems of the brainstem ceases (See Figure **11.3.5** for some of the comparisons of neurochemical activation) . The serotonergic and noradrenergic neurons in these systems are referred to as REM-off neurons because of their inactivity during REM sleep, and they can be said to act as a sort of permissive system for REM sleep, in the sense that REM sleep becomes possible only when their activity has ceased. The shutdown of these neurons might, for example, help to suppress consciousness during REM sleep. Also, the periods of REM sleep end when these aminergic neurons become active again.

The rapid eye movements - that REM gets named for - come from originating signals in the pontine reticular formation sending signals to the midbrain and which coordinate the duration and direction of eye movements.

11.3.3.4 Musculature in REM

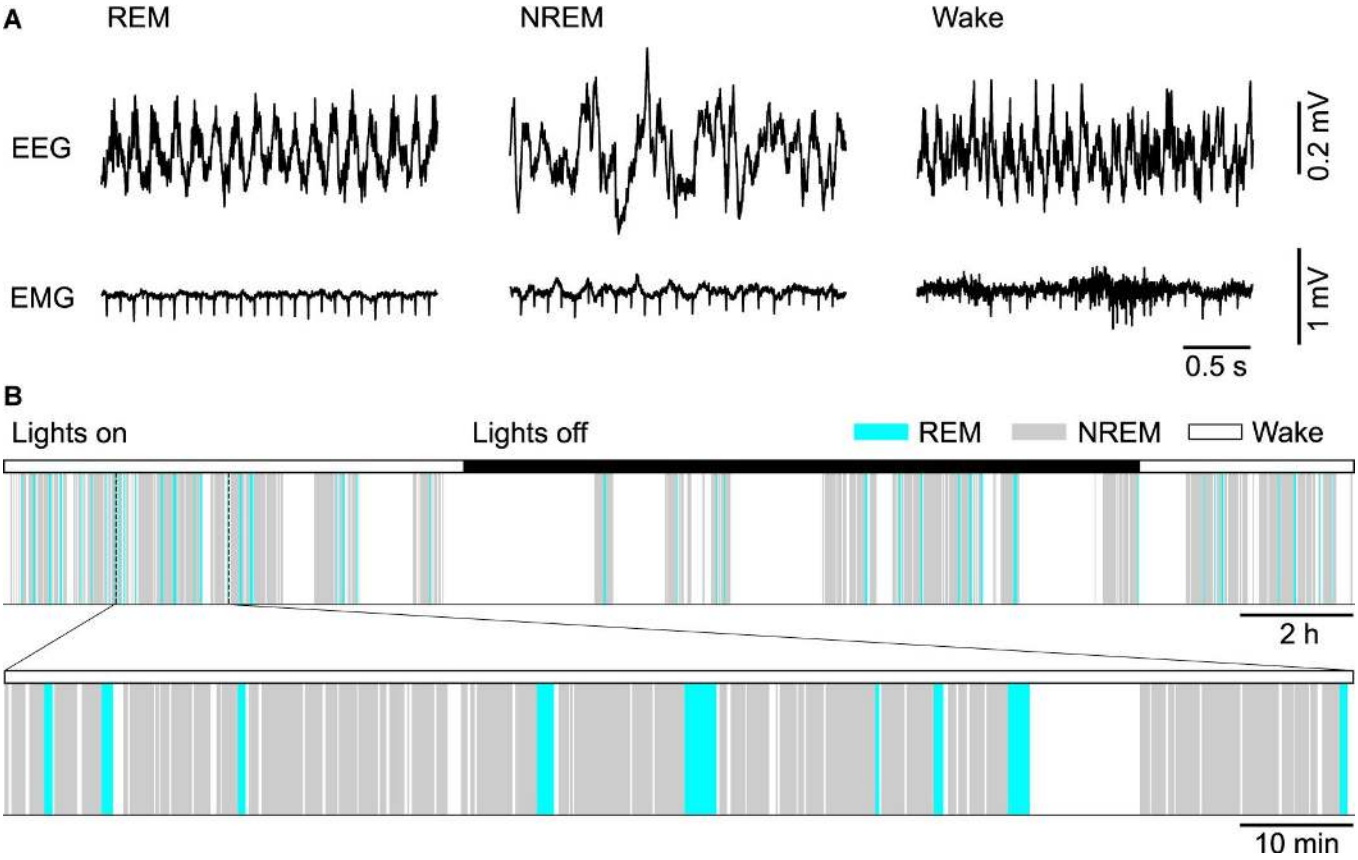


Figure **11.3.6**: Sleep in mice. **(A)** Example electroencephalogram (EEG) and electromyogram (EMG) recordings from a mouse during REM sleep, NREM sleep and wakefulness (Wake). **(B)** Top, color-coded brain state (hypnogram) for a continuous 24 h recording from a mouse during dark and light cycles. Bottom, a 2 h segment of the hypnogram shown at an expanded scale. (Neural and Homeostatic Regulation of REM sleep by Park and Weber, CC-BY)

One of the most singular characteristics of REM sleep—the paralysis of the large muscles of the body that it causes —also is explained by phenomena that occur in certain parts of the brainstem. Intense neural activity of REM sleep excites majority of neurons in the cortex including those in the primary motor cortex. These generate organized sequences of activities that represent commands for bodily movements. But they never reach motor neurons of arms and legs (only the respiratory muscles, those of eye and middle ear actually respond to those commands). See Figure **11.3.6** for the coordination of EEG and EMG in the 24-hour hypnogram of a mouse. Increase of cholinergic activity in the pons will, ultimately, inhibit motor neurons in the spinal cord.

When you are awake, your brain's wakefulness circuits exert controls that prevent you from displaying the forms of brain activity that characterize REM sleep. But in the human fetus, these controls are not yet in place, which may explain why, during the last few months of gestation, the fetus spends such a large proportion of its sleeping time (about 80%) in REM sleep.

11.3.3.5 PGO Waves - (Pons - geniculate - occipital cortex)

PGO waves are among the various phasic events that occur during REM sleep, along with the rapid eye movements and changes in breathing and heart rates. PGO waves can be generated in the absence of REM sleep by stimulation of the pons with acetylcholine. As shown in figure**11.3.7**, the pons send signals to the thalamus which in turn signal the occipital cortex.

REM sleep is triggered by a specialized set of neurons in the pons (Hobson et al., 1975). Increased activity in this neuronal population has two consequences. First, elaborate neural circuitry keeps the body immobile during REM sleep by paralyzing major muscle groups (Chase, 2008). The muscle shut-down allows the brain to simulate a visual experience without moving the body at the same time. Second, we experience vision when waves of activity travel from the pons to the lateral geniculate nucleus and then to the occipital cortex (these are known as ponto-geniculo-occipital waves or PGO waves) (Gott et al., 2017). When the spikes of activity arrive at the occipital pole, we feel as though we are seeing even though our eyes are closed (Nir and Tononi, 2010). The visual cortical activity is presumably why dreams are pictorial and filmic instead of conceptual or abstract.

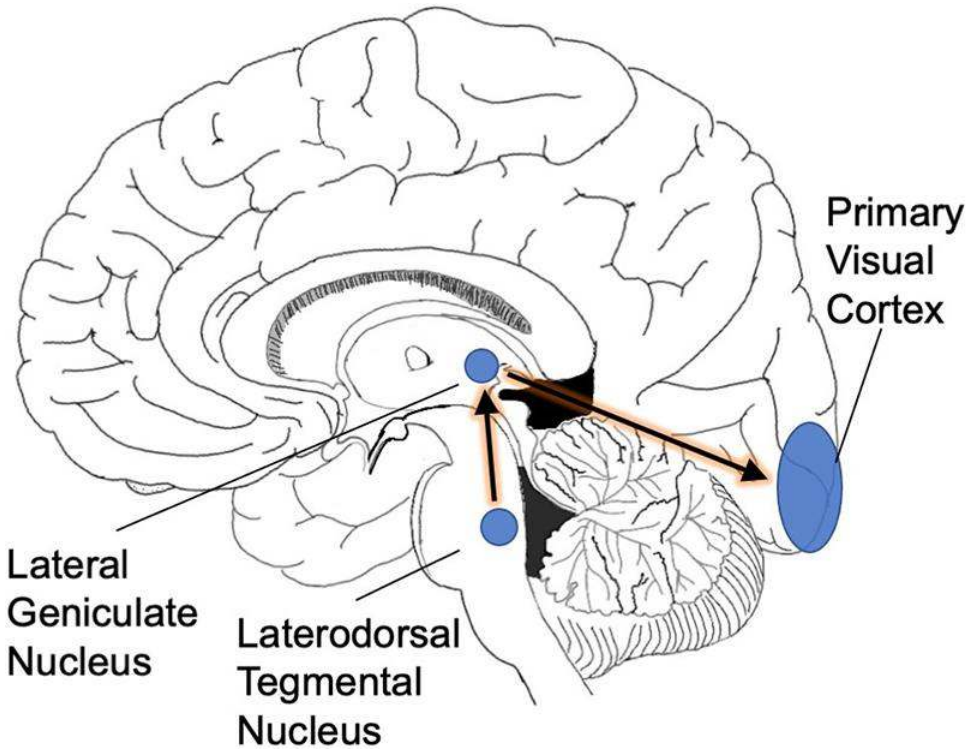


Figure **11.3.7**: PGO waves. During REM sleep, waves of activity move from the brainstem into the occipital cortex. (PGO waves in The Defensive Activation Theory: REM Sleep as a Mechanism to Prevent the Takeover of the Visual Cortex by David Eagleman and Don Vaughn licensed CC-BY 4.0)

11.3.4 Summary

In an 8 hour typical night's sleep we move through distinct stages of sleep, several times through the night. We go from relative wakefulness to the deepest stages of sleep, cycle back out and into REM multiple times. Each of these stages correspond with different EEG wave patterns as well as the involvement of different regions of the brain and neurochemicals. The nature of REM sleep is paradoxical in a number of ways, and it seems to be different from the other stages in terms of quality and quantity.

11.3.5 References

- Aserinsky, E., and Kleitman, N. (1953). Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science* 118, 273–274. doi: 10.1126/science.118.3062.273
- Chase, M. H. (2008). Confirmation of the consensus that glycinergic postsynaptic inhibition is responsible for the atonia of REM sleep. *Sleep* 31, 1487–1491. doi: 10.1093/sleep/31.11.1487
- Datta, S. (1997). Cellular basis of pontine ponto-geniculo-occipital wave generation and modulation. *Cell. Mol. Neurobiol.* 17, 341–365.
- Dement, W., and Kleitman, N. (1957). The relation of eye movements during sleep to dream activity: an objective method for the study of dreaming. *J. Exp. Psychol.* 53, 339–346. doi: 10.1037/h0048189
- Gott, J. A., Liley, D. T. J., and Hobson, J. A. (2017). Towards a functional understanding of PGO Waves. *Front. Hum. Neurosci.* 11:89.
- Hobson, J. A., McCarley, R. W., and Wyzinski, P. W. (1975). Sleep cycle oscillation: reciprocal discharge by two brainstem neuronal groups. *Science* 189, 55–58. doi: 10.1126/science.1094539
- Jouvet, M., and Michel, F. (1959). Electromyographic correlations of sleep in the chronic decorticate & mesencephalic cat. *C. R. Séances Soc. Biol. Ses Fil.* 153, 422–425.
- Karashima, A., Katayama, N., and Nakao, M. (2010). Enhancement of synchronization between hippocampal and amygdala theta waves associated with pontine wave density. *J. Neurophysiol.* 103, 2318–2325. doi: 10.1152/jn.00551.2009
- Nir, Y., and Tononi, G. (2010). Dreaming and the brain: from phenomenology to neurophysiology. *Trends Cogn. Sci.* 14, 88–100. doi: 10.1016/j.tics.2009.12.001

11.3.6 Attributions

- Different types of sleep - intermediate by Bruno Dubuc under a Copyleft license.
- Different types of sleep - advanced by Bruno Dubuc under a Copyleft license.
- Adapted from Information about sleep by NIH in the public domain.
- Information about sleep by NIH in the public domain.
- The Brain structures that keep you up and put you to sleep - intermediate by Bruno Dubuc under a Copyleft license.
- The Brain structures that keep you up and put you to sleep - advanced by Bruno Dubuc under a Copyleft license.
- Neural and Homeostatic Regulation of REM sleep by Park and Weber, in Frontiers licensed CC-BY
- Stages of Sleep by Lumen Learning, authored by OpenStax, licensed CC BY 4.0
- The Defensive Activation Theory: REM Sleep as a Mechanism to Prevent the Takeover of the Visual Cortex by David Eagleman and Don Vaughn licensed CC-BY 4.0
- REM sleep (paradoxical sleep) characteristics by Dr. Mohammad Torabi Nami and Dr. Khsoro Sadeghniaat. Licensed CC BY 3.0 via Webmed Central

---

This page titled 11.3: Stages of Sleep is shared under a mixed license and was authored, remixed, and/or curated by ASCCC OERI & Bakhtawar Bhadha (ASCCC Open Educational Resources Initiative (OERI)) .

11.4: The Brain and Sleep

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe the role of the pons, medulla, hypothalamus (SCN), thalamus, basal forebrain and pineal gland in the initiation and maintenance of sleep.
2. Analyze the role of different neurochemicals, particularly the inhibitory neurotransmitter GABA and the sleep hormone melatonin in sleep initiation and maintenance.
3. Contrast the role of other neurochemicals like excitatory neurotransmitters glutamate and norepinephrine in wakefulness and sleep.

11.4.1 Overview

In general, there are several regions of the brain that work in concert to increase wakefulness and arousal, or to cause us to fall asleep or stay asleep. The primary excitatory neurotransmitters like glutamate and acetylcholine and the primary inhibitory neurotransmitter GABA play a role in these processes.

During the Spanish flu of 1918 some patients went into coma and some were sleepless before they died. Von Economo did autopsies and found different types of lesions. He concluded that the posterior hypothalamus/upper part of midbrain might be a wakefulness center, and the preoptic area of anterior hypothalamus might be a sleep center. These areas are illustrated in Figure 11.4.1. Later countless autopsies suggest that when one goes into a coma, their brainstem has suffered damage.

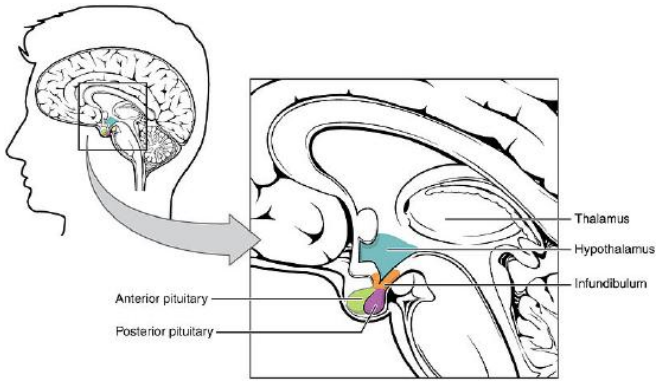


Figure 11.4.1: This sagittal section of the brain shows where the anterior and posterior pituitary, hypothalams and thalamus are located and their relations to each other in the inset - the thalamus is above the hypothalamus and the latter is above the pituitary glands. (CC BY 3.0; OpenStax College via Wikimedia Commons)

11.4.2 Sleep is not a state of neural silence

Work by Moruzzi and Magoun (1949) also suggested that when the reticular formation of cats was destroyed, comas were triggered, and stimulating this part led to awakening from normal sleep. Also, since this part receives incoming messages via sensory pathways, they developed the concept of “ascending activating reticular system” or the prime contender for “wakefulness center.”

But these initial experiments had several issues. Since neurotoxic substances that were used to destroy the neurons/cell bodies of the posterior hypothalamus and reticular formation kept the axons that originated elsewhere intact, the wakefulness function that was diminished initially, quickly returned to normal.

Most importantly, these studies also showed that sleep is not a passive process, in which being deprived of sensory input causes sleep. Many people believe that sleep is something that happens in the absence of anything else like when someone is bored. For example, medical students attending a lecture in a subject that does not totally interest them will fall off to sleep. However, it is

important to keep in mind that that is **not** actually what is happening. If a group of seven year old children who had had enough sleep the previous nights were sitting in the same lecture, they would not fall asleep but rather be bouncing off the walls!

In subsequent studies, application of electrical stimuli to the thalamus of cats when they were awake caused them to fall asleep. This indicates that sleep involves interactions between the thalamus and cortex. Also, when it was discovered that sleep and wakefulness cycles were not disrupted by sensory activation and that during REM there is intense activity in the cortex, the idea of sleep as passivity was further discredited.

There are two major neural circuits in the brainstem that operate in opposition to and alternation with each other. One of these circuits stimulates wakefulness, the other stimulates sleep, and their interaction is regulated by the body's internal clock (SCN).

Some large areas of the cortex are thus under the control of these networks of small groups of neurons that are located in the **brainstem** and that form complex circuits, not single "sleep centers" or "wakefulness centers." Wakefulness, which is indispensable for survival, is thus ensured by a whole network of redundant structures.

11.4.3 Forebrain regions involved in sleep

In the Figure **11.4.1** you can see the hypothalamus where the specific area of the brain, the suprachiasmatic nucleus (SCN- which is involved in the body's daily/circadian rhythm regulation) is located.

11.4.3.1 Anterior hypothalamus/SCN

11.4.3.1.1 SCN

The SCN is considered "body's internal clock." It will be discussed in a later section in more detail. The SCN is a bilateral structure located in the anterior part of the hypothalamus. It is the central pacemaker of the circadian timing system and regulates most circadian rhythms in the body (Hastings et al, 2018). The SCN receives signals from many different places. The major one is the retinohypothalamic tract originating from photosensitive (light-sensitive) ganglion cells of the retina. The SCN sends signals to structures such as the pineal gland, producing melatonin during the night for induction of sleep. Disruptions in the SCN circadian system have been found to correlate with various mood disorders and sleep disorders.

11.4.3.1.2 Preoptic area

The anterior hypothalamus also plays a fundamental role in the process of falling asleep. This structure, and in particular its preoptic area, appears to be sensitive to the serotonin released during waking periods. Apparently, when serotonin stimulates this preoptic area of the anterior hypothalamus, its GABAergic (GABA producing) neurons in turn inhibit the posterior hypothalamus, thus encouraging sleep. Damage to these GABAergic neurons is known to cause insomnia, whereas stimulating them causes experimental subjects to fall asleep rapidly.

11.4.3.1.3 Sleep-wakefulness switch

In vitro experiments have indicated that the wake-promoting neurotransmitters serotonin, norepinephrine, and acetylcholine inhibit identified preoptic GABA neurons (Gallopin et al., 2000); therefore, mutually inhibitory interactions between the sleep-promoting preoptic region and the arousal-related hypothalamic and midbrain structures may provide a substrate for a "sleep–wakefulness switch" (McGinty and Szymusiak, 2000; Saper et al., 2001). Thus, activation of preoptic sleep-promoting cells could lead to sleep onset by inhibiting arousal structures; in turn, activation of arousal hypothalamic and midbrain structures could suppress activity by preoptic NREM sleep-promoting cells as well as REM-promoting neurons (Reinoso-Suárez et al., 2010) and facilitate the switch to wakefulness.

11.4.3.2 Posterior hypothalamus

Stimulating the posterior hypothalamus produces a state of wakefulness comparable to that induced by stimulating the reticular formation in the brainstem. The activity of the posterior hypothalamus diminishes naturally during sleep, when it releases less histamine, a molecule that it uses as a neurotransmitter. The antihistamines that people take for allergy symptoms are known to cause some sleepiness, by reducing the activity of histamine.

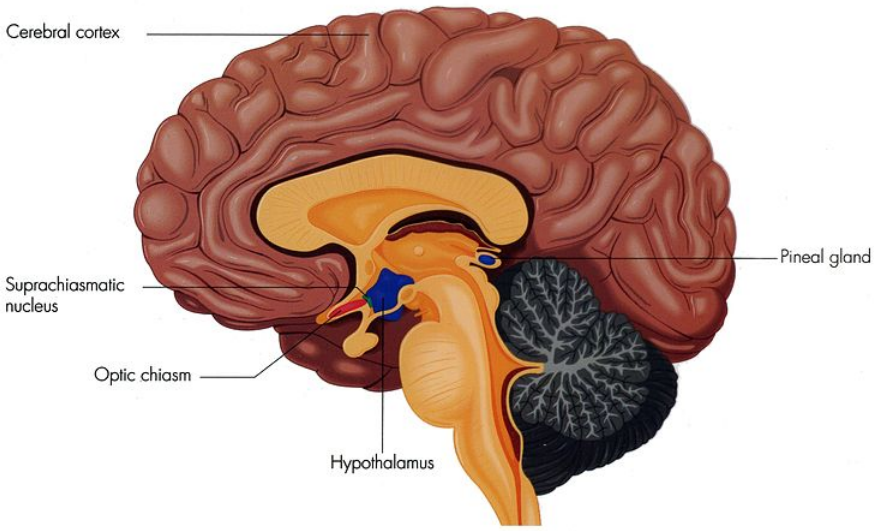


Figure **11.4.2**: Sagittal view of the brain - note the hypothalamus shown in blue, located below the thalamus, and just above the pituitary gland (CC BY-SA 3.0; 黃雨伞 via Wikimedia Commons)

11.4.3.3 Thalamus

The thalamus contains neurons that send projections throughout the cortex. The activation of these thalamocortical neurons causes them to release excitatory amino acids such as aspartate and glutamate, thus contributing to excitation of the cortex and to wakefulness. During wakefulness, these neurons generate single action potentials at regular intervals. However, as the individual falls asleep these neurons begin firing in bursts instead, thus causing the cortex to display the synchronized EEG pattern that is typical of sleep (see section * for sleep stages).

11.4.3.4 Basal forebrain

The system of the basal forebrain is composed of neurons that synthesize acetylcholine and/or GABA. On their own, these neurons account for 70% of the cholinergic innervation of the cortex (where the cortex is activated by acetylcholine), while also sending projections to the thalamic nuclei. Stimulating these neurons causes wakefulness, but destroying them with neurotoxic substances causes wakefulness to decline for only a very short time.

In the situations just described, the cortical activation that causes wakefulness results from the direct stimulation of the cortical neurons by the various components of the wakefulness network. But these cortical neurons can also be activated in another way: by the inhibition of those neurons that naturally inhibit cortical activity. And that is exactly what the GABAergic neurons located in the posterior hypothalamus and the basal forebrain do: they inhibit other, cortical GABAergic neurons.

This executive network for wakefulness is itself activated by other systems arising in the brainstem.

It is thus all of these wakefulness signals that stop reaching the cortex at the onset of non-REM sleep. They are interrupted at the thalamus, which serves as a true gatekeeper to the cortex and is greatly influenced by the diffuse neuromodulatory systems of the brainstem. The complexity of these interactions in NREM sleep is illustrated in the diagram of the cat brain below (See Figure **11.4.2**). This diagram illustrates a complex set of connections between cortical, subcortical, and brainstem structures that mediate various aspects of REM and NREM sleep. The figure is complicated, and the details are not important for our purpose here. It is included to illustrate the complexity of the connections between different brain sub-structures and signals being sent in order to maintain the individual's movement through the different stages of sleep throughout the night.

[A legend indicates that the thalamus–cerebral cortex complex or unit is darker to emphasize that these structures are necessary for the behavioral and bioelectric signs that characterize NREM sleep. AC, anterior commissure; CC, corpus callosum; DCN, deep cerebellar nuclei; Fo, fornix; G7, genu of the facial nerve; IC, inferior colliculus; IO, inferior olive; LV, lateral ventricle; MRF; midbrain reticular formation; MT, medullary tegmentum; OCh, optic chiasm; PGS, periaqueductal gray substance; RPC, caudal

pontine reticular nucleus; RPO, oral pontine reticular nucleus; SC, superior colliculus; SC and PN, spinal cord and peripheral nerves; SN, solitary nucleus; TB, trapezoid body.]

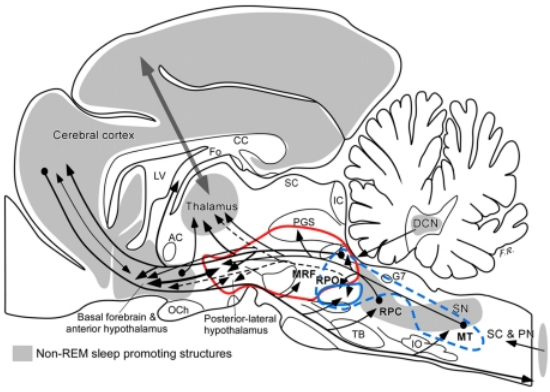


Figure **11.4.3**: Schematic representation of the encephalic NREM sleep-promoting structures. Modified from Reinoso-Suárez et al. (2011). (Functional anatomy of non-REM sleep by Isabel De Andres, Miguel Garzón, and FernandoReinoso-Suárez via ResearchGate licensed CC-BY 4.0)

The rhythmic activity pattern established by the thalamocortical neurons that disconnects the cortex from internal and external signals at the onset of non-REM sleep. In contrast, during the REM phases of sleep, the thalamus probably continues to pass at least some of these signals on to the cortex, at least in some fragmentary, filtered, or distorted form.

The regulation of wakefulness is essential for survival and involves several different redundant structures in the brain. None of these structures, taken in isolation, is indispensable for activation of the cortex. But three of these brain structures that send projections to the cortex are sufficient to maintain the desynchronized EEG pattern that is characteristic of wakefulness. These structures are the 1. posterior hypothalamus, 2. the (intralaminar nuclei of the) thalamus and 3. the basal forebrain. Together they are often referred to as the “executive network”.

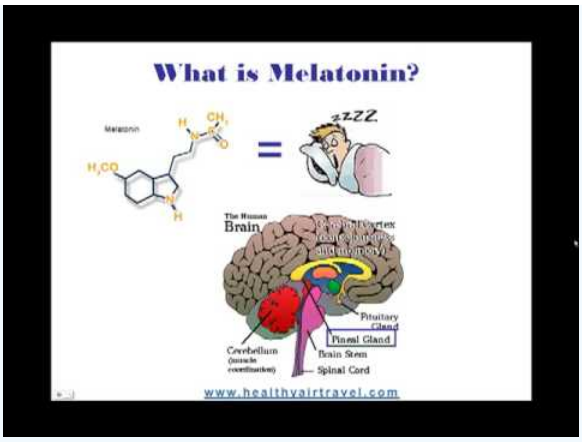
11.4.3.5 Pineal Gland

Inferior and somewhat posterior to the thalamus is the **pineal gland**, a tiny endocrine gland whose functions are not entirely clear. The **pinealocyte** cells that make up the pineal gland are known to produce and secrete the amine hormone **melatonin**, which is derived from serotonin.

The secretion of melatonin varies according to the level of light received from the environment. When photons of light stimulate the retinas of the eyes, a nerve impulse is sent to the SCN. From the SCN, the nerve signal is carried to the spinal cord and eventually to the pineal gland, where the production of melatonin is inhibited. As a result, blood levels of melatonin fall, promoting wakefulness. In contrast, as light levels decline—such as during the evening—melatonin production increases, boosting blood levels and causing drowsiness. The pineal gland and melatonin are discussed in more detail in section *

? Link for video

Watch this video to view an animation describing the function of the hormone melatonin. What should you avoid doing in the middle of your sleep cycle that would lower melatonin?



A YouTube element has been excluded from this version of the text. You can view it online here: pb.libretexts.org/aapii/?p=50

[reveal-answer q="598458"]Show Answer[/reveal-answer]
[hidden-answer a="598458"]Turning on the lights.[/hidden-answer]

The secretion of melatonin may influence the body's circadian rhythms (discussed in section *), the dark-light fluctuations that affect not only sleepiness and wakefulness, but also appetite and body temperature. Interestingly, children have higher melatonin levels than adults, which may prevent the release of gonadotropins from the anterior pituitary, thereby inhibiting the onset of puberty. Finally, an antioxidant role of melatonin is the subject of current research.

Jet lag occurs when a person travels across several time zones and feels sleepy during the day or wakeful at night. Traveling across multiple time zones significantly disturbs the light-dark cycle regulated by melatonin. It can take up to several days for melatonin synthesis to adjust to the light-dark patterns in the new environment, resulting in jet lag. Some air travelers take melatonin supplements to induce sleep.

11.4.4 Hindbrain regions involved in sleep

There are several "lower regions" of the brainstem that seem to be involved in sleep regulation according to early and current research.

11.4.4.1 Pons

The brainstem region known as the pons is critical for initiating REM sleep. During REM sleep, the pons sends signals to the visual nuclei of the thalamus and to the cerebral cortex (this region is responsible for most of our thought processes). The pons also sends signals to the spinal cord, causing the temporary paralysis that is characteristic of REM sleep. This is what happens in the experience of temporary sleep paralysis that many people might have experienced. Since the pons is effectively still "asleep" while the cortex is "awake," people experience the terrifying inability to move their body while they are completely conscious!

Also, the structures like the anterior hypothalamus play an important role in the onset of sleep.

11.4.4.2 Medulla and RAS

The reticular activating system (RAS) plays an important role in conscious awareness. According to Iwanczuk and Guzniczak (2015), "The ascending reticular activating system (ARAS) is responsible for a sustained wakefulness state. It receives information from sensory receptors of various modalities, ... [and those pathways] reach the thalamus directly or indirectly... The reticular activating system begins in the dorsal part of the posterior midbrain and anterior pons, continues into the diencephalon, and then divides into two parts reaching the thalamus and hypothalamus, which then project into the cerebral cortex."

The solitary tract nucleus region in the dorsal medulla is thought to provide a link between visceral activities such as respiratory, cardiovascular and gastrointestinal functions, and the sleep–wakefulness states. The solitary tract nucleus does not directly project to the cerebral cortex (Saper, 1995), although it does project to several brainstem, thalamic, and hypothalamic areas that innervate the cortex and can mediate EEG and sleep responses, such as the lateral hypothalamus, and nuclei of the midline thalamus.

11.4.5 Summary

The main brain regions working together to produce sleep and wakefulness have been discussed here. The neurons of the RAS play a role in arousal and sleep to some degree. The "executive network" seems to play an important role in regulating the responsiveness of the cortex to outside stimulation. The anterior hypothalamus and pineal regulate and fine tune many of these functions as well.

11.4.6 References

- Gallopin, T., Fort, P., Eggermann, E., Cauli, B., Luppi, P. H., Rossier, J., Audinat, E., Mühlethaler, M., & Serafin, M. (2000). Identification of sleep-promoting neurons in vitro. *Nature*, *404*, 992–995.
- Hastings, M.H., Maywood, E.S., & Brancaccio, M. (2018). Generation of circadian rhythms in the suprachiasmatic nucleus. *Nat Rev Neurosci. 19*(8), 453-469.
- Iwańczuk, W., & Guźniczak, P. (2015). Neurophysiological foundations of sleep, arousal, awareness and consciousness phenomena. Part 1. *Anaesthesiol Intensive Ther. 47*(2), 162–167. doi:10.5603/AIT.2015.0015
- McGinty, D., and Szymusiak, R. (2000). The sleep-wake switch: a neuronal alarm clock. *Nat. Med.* 6, 510–511.
- Reinoso-Suárez, F., de la Roza, C., Rodrigo-Angulo, M. L., de Andrés, I., Núñez, A., and Garzón, M. (2010). “GABAergic mechanisms in the ventral oral pontine tegmentum – the REM sleep induction site- in the modulation of sleep-wake states,” in *GABA and Sleep*, ed. J. M. Monti, S. R. Pandi-Perumal, and H. Möhler (Basel: Birkhauser-Verlag), 233–252.
- Saper, C. B. (1995). “Central autonomic system,” in *The Rat Nervous System*, ed. G. Paxinos (San Diego: Academic Press), 107–135.
- Saper, C. B., Chou, T. C., and Scammell, T. E. (2001). The sleep switch: hypothalamic control of sleep and wakefulness. *Trends Neurosci.* 24, 726–731.

11.4.7 Attribution

- The Brain structures that wake you up and put you to sleep by Bruno Dubuc under a Copyleft license.
- Functional Anatomy of Non-REM Sleep by Isabel de Andrés, Miguel Garzón and Fernando Reinoso-Suárez in Frontiers Neuroscience licensed CC-BY 4.0
- Information about sleep by NIH is in the public domain
- The Pineal Gland by Libretexts has no license indicated.
- Neuroanatomy, Nucleus Suprachiasmatic by Melinda A. Ma and Elizabeth H. Morrison licensed CC-BY 4.0

This page titled 11.4: The Brain and Sleep is shared under a mixed license and was authored, remixed, and/or curated by ASCCC OERI & Bakhtawar Bhadha (ASCCC Open Educational Resources Initiative (OERI)) .

- **3.8: The Pineal Gland** has no license indicated.

11.5: Theories of Sleep

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Evaluate the functions of sleep in relation to the biological mechanisms underlying it and affected by it.
- Describe the circadian nature of sleep.
- Describe the homeostatic or recuperative theory of sleep, and other theories of why we sleep.

11.5.1 Overview

When you ask most people why they think they sleep – the answer is "so I can rest" or "because I am tired from the activities of the day." However, the animals that work the hardest do not sleep the longest and even after pulling an all-nighter we don't sleep for the entire following day to make up for the lack of sleep. There are two theories of sleep that work in conjunction to explain why we sleep – recuperative and circadian.

Research indicates that for someone to be able to fall asleep, two bodily processes must be properly synchronized. The first is the **circadian rhythm**, which has a 24-hour period and is governed by your body's biological clock. The circadian rhythm controls the cyclical secretion of several hormones, including melatonin, that are involved in sleep. The second process is the **recuperative function** illustrated by the accumulation of hypnogenic substances in your body for 16 hours every day. These substances induce a desire to sleep that does not go away until you in fact get some sleep.

Thus you can fall asleep only when two conditions have been met: your body's biological clock must have brought it into a hormonal balance conducive to sleep, and it must have been a good while since you last slept, so that your levels of hypnogenic (sleep-producing) substances have built up sufficiently. The following section describes these two theories in greater detail, as well as other functions of sleep.

11.5.1.1 Sleep Regulation and Circadian Rhythms: A Two-Process Model

Sleep is a dynamic process that adjusts to the body's needs every day. What time you fall asleep, how long you sleep, and how well you sleep all result from the combined effects of two forces: the homeostatic debt and the phase of your circadian rhythm. Individuals of course differ as to what time they go to bed and how much sleep they need to function well, but on the whole, the characteristics of sleep can be regarded as the result of complex interactions between two processes.

The first of these processes is called homeostatic debt (see below) or recuperative theory, which increases as a function of how long you have been awake and decreases as you sleep. The process is somewhat like the sand accumulating in one end of an hourglass and having to be emptied into the other after a certain time. See Figure **11.5.1** for a graph of how the pressure to sleep increases as a function of time spent awake, and then decreases as we sleep.

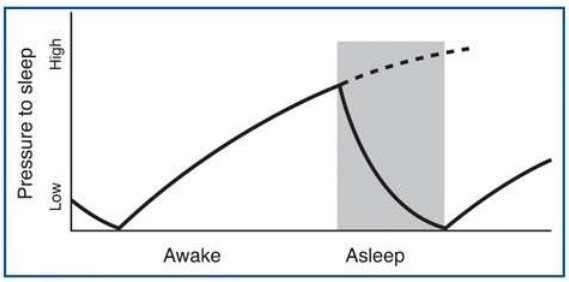


Figure **11.5.1**: Homeostatic regulation of sleep: the pressure to sleep grows stronger across the day as one stays awake and then dissipates when one sleeps at night (shaded area). Sleep pressure increases (dashed line) as one stays awake longer into the normal sleeping hours. (NIH in the public domain)

The second process that greatly influences the onset, duration, and quality of your sleep is the phase of your circadian rhythm. This phase is governed by your biological clock, whose rhythm is endogenous but is reset regularly by daylight. This clock therefore produces a cycle lasting about 24 hours during which the optimal times for falling asleep, dreaming, waking up, and doing work occur over the course of each day. The external cues or zeitgebers (German for "time-givers") - of which sunlight is the most important - entrain your endogenous (internal) body clock to fall in line with the earth's day.

Endogenous sleep rhythms can be depicted graphically. The figure shows a day-by-day representation of one individual's sleep/wake cycle. The black lines indicate periods of sleep, and the gray lines indicate periods of wakefulness. The upper portion of the figure (days 1 through 9) represents this individual's normal sleep/wake cycle. Under these conditions, the individual is exposed to regularly timed exposure to alternating daylight and darkness, which has entrained this person's sleep/wake cycling to a period of 24 hours.

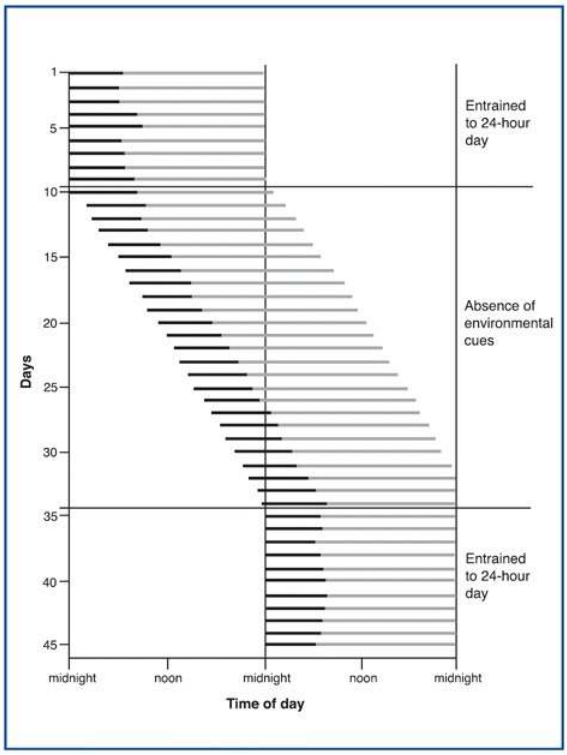


Figure **11.5.2**: Entrainment of the biological clock. Black bars=sleep, gray bars=awake. (NIH in the public domain)

Thus your entire sleep-wake cycle operates as if your circadian oscillator made falling asleep easier at certain times of day, making you appreciably sleepier from 1:00 PM to 4:00 PM, and even sleepier from 2:00 AM to 5:00 AM. These patterns are confirmed by the statistics on workplace and highway accidents.

This two-process model - homeostatic and circadian - for regulating sleep can be diagrammed as a double pendulum.

Under normal circumstances, your periods of activity and rest are in phase with the alternation of day and night—your “circadian” pendulum and your “homeostatic” pendulum are in synch, you sleep well when you fall asleep, and you function well when you wake up.

But when this relationship is disturbed and the two pendulums are no longer in phase, then the quality both of your sleep and of your performance when awake deteriorate significantly. The peaks of activity for several circadian markers occur at inconvenient times in the sleep-wake cycle, which is the source of the problems caused by jet lag and by working night shifts.

In other words, the longer you stay awake, the greater the pressure you will feel to go to sleep. This process of homeostatic debt, or sleep debt, also explains why, if you stay up all night, then the next night, not only will you sleep longer, but your percentage of deep sleep will be higher.

11.5.2 Molecules that build up and make you sleep – Adenosine and Melatonin

As each day draws to a close, you feel the need to lie down and go to sleep. The onset of sleep, which seems like such a simple phenomenon from a behavioral perspective, is actually quite complex from a molecular one.

11.5.2.1 Adenosine and homeostasis/recuperation

It was in the early 1980s that scientists first discovered the chemical mechanism by which drinking coffee helps people to stay awake: caffeine, the psychoactive substance in coffee, prevents adenosine from binding to certain neurons in the brain. Both the caffeine in coffee and the theophylline in tea are examples of such adenosine antagonists and are well known for their stimulant effects.

Once this discovery was made, adenosine became a subject of interest for more and more neurobiologists who were doing sleep research. Numerous animal experiments eventually confirmed that adenosine definitely plays a role in the sleep/wake cycle. Some of the experimental findings that led to this conclusion: a) blocking the effects of adenosine made animals more alert; b) injecting animals with an adenosine agonist (enhancer) caused them to fall asleep; c) in certain parts of the brain, the concentration of adenosine normally increases naturally during the day and decreases at night, but if animals are forced to stay awake at night, this concentration keeps increasing.

These experiments thus showed that adenosine, along with other chemicals such as serotonin and melatonin, is one of the molecules whose concentration in the brain influences the onset of sleep.

Adenosine is produced by the degradation of adenosine triphosphate (ATP), the molecule that serves as the “energy currency” for the body’s various cellular functions. The amount of adenosine produced in the brain thus reflects the activity level of its neurons and glial cells. The brain’s intense activity during periods of wakefulness consumes large amounts of ATP and hence causes adenosine to accumulate.

Glycogen=ATP reserves or energy currency of the brain

Brain uses energy --> ATP degrades into adenosine

More brain energy used = more adenosine accumulates

More adenosine levels --> non-REM sleep triggered

Non-REM sleep = less active brain --> recovery/rebuild glycogen stores

The accumulation of adenosine during waking periods is thus associated with the depletion of the ATP reserves stored as glycogen in the brain. The increased adenosine levels trigger non-REM sleep, during which the brain is less active, thus placing it in a recovery phase that is absolutely essential—among other things, to let it rebuild its stores of glycogen.

But how exactly does adenosine exert this influence? During periods of wakefulness, neuronal activity increases the concentration of adenosine, which has an inhibitory effect on a great many neurons. Among these are the neurons of the hormonal systems that are the most active when we are awake: the norepinephrine, acetylcholine, and serotonin systems. Experiments have shown, for example, that when the levels of adenosine in the basal forebrain are raised artificially, the neurons in this structure that project axons throughout the cortex produce less acetylcholine. As a result, cortical activity slows, and the individual falls asleep.

The synchronized brain activity characteristic of non-REM sleep can then become established. But once non-REM sleep has continued for a while, the adenosine levels begin to decline. The systems responsible for wakefulness can then start becoming more active, causing the individual to awaken and the cycle to begin all over again. Thus we see that the sleep/wake cycle involves a highly efficient negative feedback loop.

11.5.2.2 Circadian rhythms

There are two important structures for the circadian cycle that could be considered our biological clocks.

11.5.2.3 The suprachiasmatic nuclei and the pineal gland and melatonin

Most human bodily functions and behaviors are not “steady-state”. Instead, they fluctuate in 24-hour cycles, such as the sleeping and waking cycle and the cycles for body temperature, hunger, and the secretion of various hormones. The central clock that regulates all of these circadian cycles is located in two tiny structures in the brain, at the base of the left and right hypothalamus.

Each of these structures is no larger than a pencil tip and contains several tens of thousands of neurons. These structures are called the suprachiasmatic nuclei (as mentioned earlier in section *) because they are located just above the optic chiasma, where the left and right optic nerves cross paths.

This strategic position enables the suprachiasmatic nuclei to receive projections from the optic nerve from special retinal ganglion cells that tell them about the intensity of the ambient light entering the eyes. The neurons of these nuclei use this information to resynchronize themselves with daylight every day, because like any clock, the human biological clock is not perfect and does need to be reset periodically. One interesting thing about the retinohypothalamic path is that it is separate from those for vision such that even blind people (and blind mole rats, a species that is otherwise blind) receive information from light to reset their biological clocks.

Despite this need to resynchronize with an external cue, it has been shown that the suprachiasmatic nuclei do in fact constitute a biological clock with its own independent rhythm. First, many experiments have shown that the fluctuations of the human circadian cycle persist even when individuals are cut off from the light of day. Second, in experiments where the suprachiasmatic nuclei were destroyed in animals such as hamsters, their cyclical behaviors, such as their sleep/wake cycles, become completely disorganized. And when suprachiasmatic nuclei were then transplanted from hamster fetuses into these animals, their biological rhythms returned, but with the properties of the donors.

These findings indicate that the mammalian biological clock mechanism is in fact not only endogenous, but also of genetic origin. Scientists have now even determined that these rhythms are the result of the cyclical activity of certain genes.

The circuit of light entering the eye, the signal going to the SCN and to the pineal gland is shown in the figure **11.5.4**. Melanopsin in retinal ganglion cells in the eye respond to light (natural or artificially) and transmit signals to the SCN. Then light-induced activation of the SCN prevents the pineal gland from producing melatonin and; conversely, melatonin production and secretion is increased during the dark period.

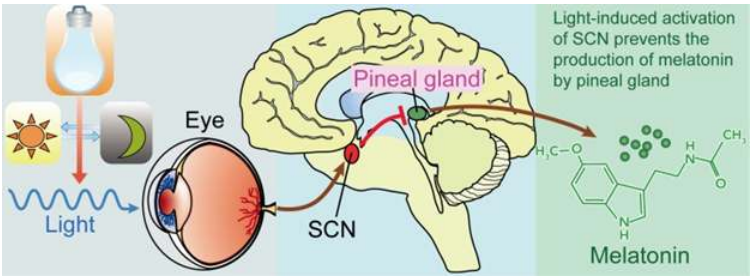


Figure **11.5.3**: Light, suprachiasmatic nuclei (SCN), and the pineal/melatonin circuit. (Zhiqiang Ma, Yang Yang, Chongxi Fan, Jing Han, Dongjin Wang, Shouyin Di, Wei Hu, Dong Liu, Xiaofei Li, Russel J. Reiter, and Xiaolong Yan, CC BY 4.0 <https://creativecommons.org/licenses/by/4.0>, via Wikimedia Commons)

11.5.2.4 Pineal gland

Using these cyclical rhythms, the suprachiasmatic nuclei send signals along their output pathways–for example, to the pineal gland–to regulate the cycles of a number of physiological and behavioral functions. In birds, reptiles, and fish, this small gland located at the top of the brain is sensitive to light and co-ordinates some cyclical phenomena on its own. In mammals, however, though the pineal gland (mentioned earlier in section 12.2) does retain its ability to produce secretions cyclically (specifically, the hormone melatonin, at night), it does not constitute a clock on its own; instead, its cyclical synthesis of melatonin is controlled by timing signals that it receives from the SCN.

Each day, the pineal gland begins to produce melatonin (sometimes called the “sleep hormone”) as night falls. As the level of melatonin in the blood rises, body temperature falls slightly, and the individual feels sleepier and sleepier. The melatonin level remains high for just about 12 hours, then starts falling again in the early morning, as daylight (via the SCN) inhibits this gland’s activity.

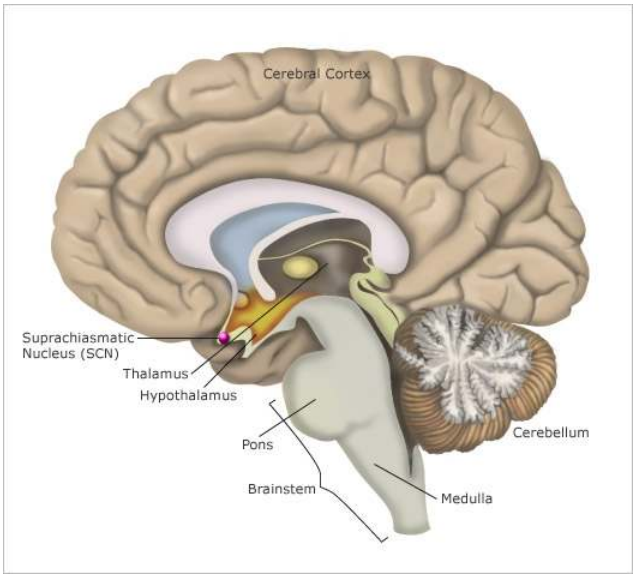


Figure \(\PageIndex{4\): The SCN in the anterior part of the hypothalamus is indicated along with the brainstem - pons and medulla - below (Sagittal image of brain licensed CC BY-SA)

The main neurotransmitter regulating the activity of the pineal gland is norepinephrine. When norepinephrine binds to its receptors, it triggers a cascade of second messengers, including cyclic AMP. This cyclic AMP contributes to the synthesis of melatonin. This melatonin is released into the bloodstream, through which it reaches every organ in the body. That is how it participates in the modulation of the circuits of the brainstem that ultimately control the sleep-wake cycle.

11.5.2.5 Output of SCN

The output pathways from each SCN (see Figure **11.5.4** for another picture of its location) consist of axons that innervate mainly the hypothalamus and nearby structures. Some of these axons also project to other parts of the forebrain, while others project to the midbrain.

Scientists do not yet know the details of how the central biological clock in the SCN regulates so many different human cyclical behaviors. But they do know that it uses the pineal gland to do so, and they have shown that destroying the SCN's output pathways also destroys the body’s circadian rhythms. Because GABA is the essential neurotransmitter for almost all of the SCN's neurons, one would expect an inhibitory effect on the neurons that they innervate. In addition to sending out messages along these axonal pathways, the SCN's neurons seem to secrete a neuropeptide called vasopressin in a cyclical pattern.

Yet another figure **11.5.5** shows the relative location of the different structures involved in the regulation of the circadian rhythms.

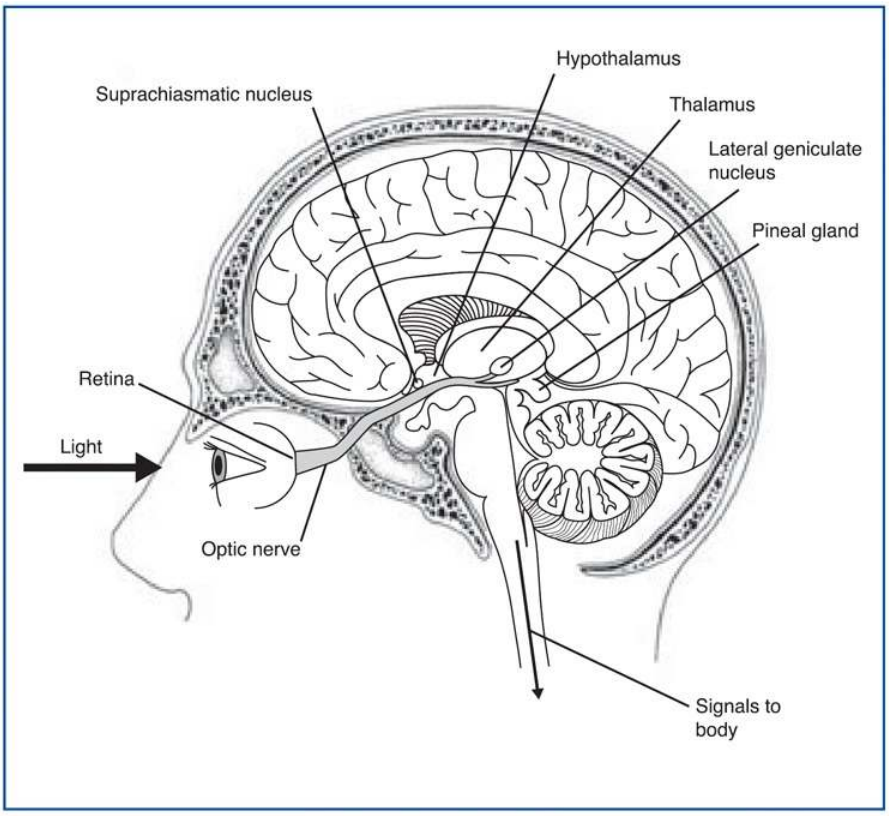


Figure 11.5.5: The biological clock is located in the SCN of the brain (NIH in the public domain)

Scientists have now discovered that the neurotransmitter GABA excites the cells of the dorsal SCN but inhibits those of the ventral SCN. These opposing effects might influence the differing reaction times of these two sub-regions when someone travels across several time zones. This discovery thus opens new insights into the mechanisms behind the disturbing symptoms of jet lag.

Scientists removed neurons from the suprachiasmatic nuclei of rats and isolating these cells in a culture medium in vitro and found that feedback loops inside each of the cells discharge at frequencies varying in cycles lasting about 24 hours. But unlike suprachiasmatic nucleus cells in the brain, which synchronize their activity with the day/night cycle, suprachiasmatic nucleus cells in vitro do not. Like any other clock, the human body's biological clock needs to be reset periodically. For that to happen, every cell in this clock must resynchronize itself daily with external cues that tell it when the day begins and ends. These external cues, also known as Zeitgebers (German for "time givers"), include the ambient temperature, the consumption of meals, ambient noise, and the body's activity level. But the strongest of these cues is undoubtedly the overall intensity of the ambient light.

11.5.3 Resetting the body's internal clocks - Zeitgebers

11.5.3.1 Light-sensitive ganglion cells

Thanks to various tagging methods, scientists now know that a certain sub-population of the ganglion cells in the human retina contain a photosensitive pigment and project their axons directly into the suprachiasmatic nuclei as well as into other brain structures that are concerned with the intensity of ambient light.

These light-sensitive ganglion cells have large receptive fields, because of their long, widely dispersed dendrites. In these cells, accurate reception of information on shape, orientation, and movement is sacrificed to general sensitivity. These cells clearly constitute another light-sensitive system that runs parallel to the visual system but is dedicated to detecting light intensity rather than to forming images.

11.5.3.2 Molecular clockwork

Many human functions, such as alertness, body temperature, and the secretion of certain hormones, work better if they are adjusted according to whether it is day or night. As you might therefore expect, a mechanism has evolved within the human body to co-

ordinate its major functions with the time of day.

The human biological clock is extremely regular, though not exactly 24 hours: it is accurate to within 1%. But just like a watch that is never absolutely accurate on its own and needs to be reset occasionally, this biological clock needs a mechanism to prevent tiny errors from accumulating in each cell. Also, it needs to synchronize itself with external signs that tell it when each new day begins. The growing intensity of the natural light is the first sign of daybreak, and special photopigments in the retina detect this change in light intensity and transmit this information to the human biological clock.

In Figure **11.5.7**, the levels of organization of information and operation of the circadian systems are shown. Circadian systems need to be considered in relation to three differing levels of organization of information and operation. First is the way in which the physical environment communicates (or 'Inputs') key information, particularly related to differentiation of night from day, to the internal 'master' clock (located in the brain's suprachiasmatic nucleus (SCN)). Second are the 'Intrinsic' brain factors, consisting of the master clock and its linked regulatory systems (notably secretion of melatonin from the pineal gland). These contribute to sleep onset, sleep architecture, sleep-wake cycles and other central nervous system (CNS)-dependent behavioral changes. Third is the way in which the circadian system coordinates all other hormonal, metabolic, immune, thermoregulatory, autonomic nervous and other physiological processes to optimize the relationships between behavior and body functions (that is, the 'Outputs').

At the cellular level, almost all individual cells and, hence, organ systems have their own intrinsic clocks. As these cellular (for example, fibroblasts, fat cells, muscles) and organ-based (for example, liver, pancreas, gut) clocks run to intrinsically different period lengths, the differing physiological systems need to be aligned in coherent patterns. Fundamentally, the master circadian clock permits the organism to align key behavioral and intrinsic physiological rhythms optimally to the external 24-hour light–dark cycle.

In the figure, the black arrows from eye (and sunlight), genes, and environment/daily patterns (eating/sleeping/exercising) to the SCN indicate input received by the master clock. The red arrows from the raphe nuclei and pineal gland (via melatonin) to the SCN, and from the SCN to the pituitary indicate intra-CNS modulation of the signals. The blue arrows from melatonin to sleep regulation signals, from the SCN to ANS activity and temperature regulation, and pituitary to hormone regulation (ACTH and TSH) indicate output from the brain controlling other organ systems in the body.

The table at the end illustrates how muscle, liver, fat and pancreas function changes depending on time of day or night. During the day, muscle uptake fatty acid, and increase glycolitic metabolism. For fats, there is lipogenesis and adiponectin production. In the liver, there is glycogen, cholesterol and bile acid synthesis. In the pancreas there is insulin secretion. At night, muscles engage in oxidative metabolism. Fats have lipid catabolism and leptin secretion. The liver engages in gluconeogenesis, glycogenolysis, and mitochondrial biogenesis. The pancreas secrete glucagon.

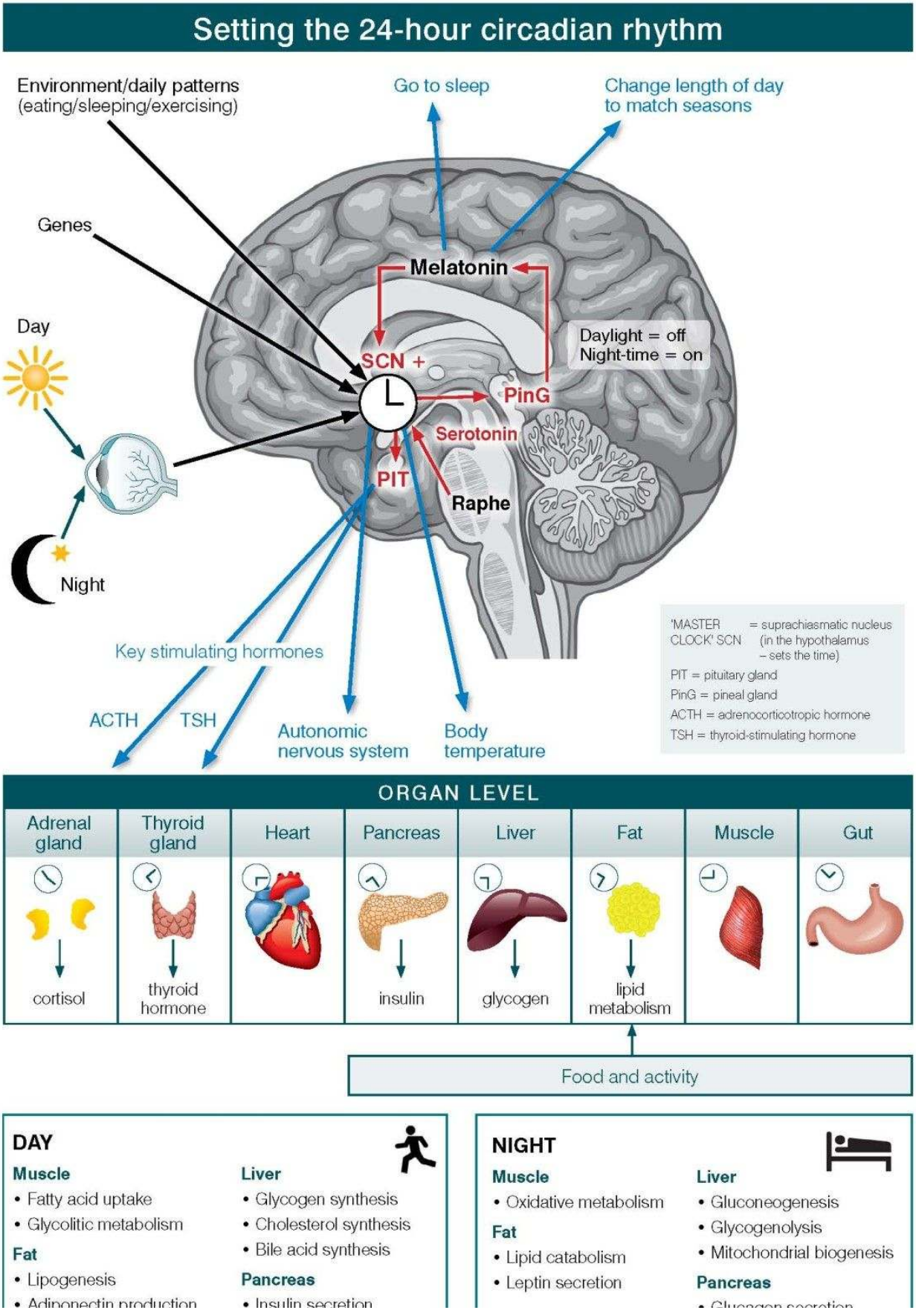

Setting the 24-hour circadian rhythm
Environment/daily patterns
(eating/sleeping/exercising)
Go to sleep
Change length of day
to match seasons
Genes
Melatonin
Day
Daylight = off
Night-time = on
SCN +
PinG
Serotonin
PIT
Raphe
Night
Key stimulating hormones
ACTH
TSH
Autonomic
nervous system
Body
temperature
'MASTER CLOCK' SCN = suprachiasmatic nucleus (in the hypothalamus – sets the time)
PIT = pituitary gland
PinG = pineal gland
ACTH = adrenocorticotropic hormone
TSH = thyroid-stimulating hormone
ORGAN LEVEL
Adrenal gland
Thyroid gland
Heart
Pancreas
Liver
Fat
Muscle
Gut
cortisol
thyroid hormone
insulin
glycogen
lipid metabolism
Food and activity
DAY
Muscle
• Fatty acid uptake
• Glycolitic metabolism
Fat
• Lipogenesis
• Adiponectin production
Liver
• Glycogen synthesis
• Cholesterol synthesis
• Bile acid synthesis
Pancreas
• Insulin secretion
NIGHT
Muscle
• Oxidative metabolism
Fat
• Lipid catabolism
• Leptin secretion
Liver
• Gluconeogenesis
• Glycogenolysis
• Mitochondrial biogenesis
Pancreas
• Glucagon secretion

Figure 11.5.6: The master clock (SCN) receives input from external stimuli, connects with other brain regions (primarily the pineal gland), and regulates/moderates several organ systems and functions in the body. (The master circadian clock in the human brain.jpg by Ian B Hickie, Sharon L Naismith, Rébecca Robillard, Elizabeth M Scott, and Daniel F Hermens, licensed CC BY 3.0, via Wikimedia Commons)

11.5.4 Other theories of why we sleep

As mentioned in section * there are many functions of sleep including memory consolidation, energy conservation, brain development and discharge of emotions. Further, the importance of sleep is apparent in the multiple negative ramifications of sleep deprivation.

11.5.5 Summary

Under normal circumstances, there are two main mechanisms by which our sleep is regulated - homeostatic or recuperative and circadian or rhythmic. Adenosine appears to be the main chemical underlying the homeostatic/recuperative functions of sleep. Indeed most of us are aware of how we feel tired after long periods of wakefulness. Melatonin and the free running rhythms of the SCN seem to underlie the circadian/rhythmic aspects of sleep. Jet lag, shift work, and the overall disruption of people's body clocks during the COVID pandemic provide testimony to how this theory is also an important basis of sleep. Certainly it is clear that humans need sleep for a variety of reasons and that when we are sleep deprived, there are many mental and physical health consequences.

11.5.6 References

11.5.7 Attributions

- Molecules that build up and make you sleep - intermediate by Bruno Dubuc under a Copyleft license.
- Molecules that build up and make you sleep - beginner by Bruno Dubuc under a Copyleft license.
- The suprachiasmatic nuclei and the pineal gland - intermediate by Bruno Dubuc under a Copyleft license.
- Light-sensitive ganglion cells - advanced by Bruno Dubuc under a Copyleft license.
- Cyclic behavior of animals by CK-12: Biology Concepts licensed CC BY-NC 4.0
- Our Molecular Clockwork - intermediate by Bruno Dubuc under a Copyleft license.
- Our Molecular Clockwork - beginner by Bruno Dubuc under a Copyleft license.
- Sleep and Dreaming by Lumen Boundless Psychology licensed CC BY-SA
- Sagittal image of brain licensed CC BY-SA
- Light, suprachiasmatic nuclei (SCN) and the pinealmelatonin ciruit.jpg by Zhiqiang Ma, Yang Yang, Chongxi Fan, Jing Han, Dongjin Wang, Shouyin Di, Wei Hu, Dong Liu, Xiaofei Li, Russel J. Reiter, and Xiaolong Yan, licensed CC BY 4.0, via Wikimedia Commons
- Tool Module: Sleep Regulation and Circadian Rhythms by Bruno Dubuc under a Copyleft license.
- The master circadian clock in the human brain.jpg by Ian B Hickie, Sharon L Naismith, Rébecca Robillard, Elizabeth M Scott, and Daniel F Hermens, licensed CC BY 3.0, via Wikimedia Commons

This page titled 11.5: Theories of Sleep is shared under a mixed license and was authored, remixed, and/or curated by ASCCC OERI & Bakhtawar Bhadha (ASCCC Open Educational Resources Initiative (OERI)) .

11.6: Sleep Disorders

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe the different kinds of sleep disorders.
2. Describe the causes of different sleep disorders.
3. Analyze different treatment options for sleep disorders.
4. Describe current research on the connection between sleep disorders and other physical health conditions.

11.6.1 Overview

Research has demonstrated that good sleep habits result in improved health and psychological functioning (Hyyppa & Kronholm, 1989). In fact poor sleep quality can result in severe psychiatric symptoms. Worldwide, approximately 30 percent of adults report difficulty initiating or maintaining sleep or experiencing poor sleep quality.

About 50 to 70 million Americans have sleep or wakefulness disorders. Sleep deficiency and untreated sleep disorders are associated with a growing number of health problems, including heart disease, high blood pressure, stroke, diabetes, obesity, and certain cancers. Sleep disorders can also be costly. Each year sleep disorders, sleep deprivation, and sleepiness add to the national health care bill. Undiagnosed sleep apnea alone is estimated to cost the US $150 billion annually. Additional costs to society for related health problems, lost worker productivity, and accidents make sleep disorders a serious public health concern.

In this section insomnia, circadian rhythm disorders, sleep apnea and narcolepsy are discussed.

11.6.2 Insomnia

Insomnia is a common sleep disorder. With insomnia, you may have trouble falling asleep, staying asleep, or getting good quality sleep. Chronic (long-term) insomnia occurs three or more nights a week, lasts more than three months, and cannot be fully explained by another health problem or a medicine.

Insomnia can affect your memory and concentration. Chronic insomnia raises your risk of high blood pressure, coronary heart disease, diabetes, and cancer.

In addition to finding out your medical history and having you keep a sleep diary, your doctor may have you take the following tests.

- **A sleep study** to look for other sleep problems, such as circadian rhythm disorders, sleep apnea, and narcolepsy.
- **Actigraphy** to measure how well you sleep. This requires you wear a small motion sensor three to 14 days.
- **Blood tests** to check for thyroid problems or other medical conditions that can affect sleep.

In the short term, insomnia can make it hard to concentrate or think clearly. You may feel irritable, sad, and unrested or have headaches. Insomnia raises your risk of falling, having a road accident, or missing work.

In addition, chronic insomnia can affect how well your brain, heart, and other parts of your body work. It can raise your risk of certain health problems or make existing problems worse. These conditions include:

- **Breathing problems** such as asthma
- **Heart problems** such as arrhythmia, heart failure, coronary heart disease, and high blood pressure
- **Mental health conditions** such as anxiety, depression, and thoughts of suicide. Insomnia can also make it difficult for you to stick to treatment for a substance use disorder.
- **Pain.** People who have chronic pain and insomnia may become more aware of and distressed by their pain.
- **Pregnancy complications** such as having more pain during labor, giving birth preterm, needing a cesarean section (C-section), and having a baby with low birth weight.

- **Problems with your immune system,** your body's natural defense against germs and sickness. These problems can lead to inflammation in your body or make it harder to fight infections.
- **Problems with your metabolism.** Not getting enough sleep can change the levels of hormones that control hunger and how you break down food. This can raise the risk of overweight and obesity, metabolic syndrome, and diabetes.

11.6.2.1 Treatment

Doctors may recommend healthy lifestyle habits such as a regular sleep schedule, cognitive behavioral therapy for insomnia, and medicines to help manage the insomnia.

Medicines

Prescription medicines

Many prescription medicines are used to treat insomnia. Some are meant for short-term use while others are meant for longer-term use. Some insomnia medicines can be habit-forming and all of these medicines may cause dizziness, drowsiness, or worsening of depression or suicidal thoughts. All of the medicines listed below may also cause insomnia.

- **Benzodiazepines,** like Valium, and **Benzodiazepine receptors agonists,** such as zolpidem**,** zaleplon, and eszopiclone can be habit-forming and should be taken for only a few weeks. Benzodiazepines are GABA agonists. Remember, from chapter *, that GABA is one of the most common inhibitory neurotransmitters in the central nervous system. Hence Benzodiazepines (and Benzodiazepine receptor agonists) make it even less likely that neurons will fire. Additional side effects of Benzodiazepine receptor agonists may include anxiety. Rare side effects may include a severe allergic reaction or unintentionally doing activities while asleep such as walking, eating, or driving.
- **Melatonin receptor agonists,** such as ramelteon. Rare side effects may include doing activities while asleep such as walking, eating or driving or a severe allergic reaction.
- **Orexin receptor antagonists,** such as suvorexant. This medicine is not recommended for people who have narcolepsy. Rare side effects may include doing activities while asleep such as walking, eating, or driving or not being able to move or speak for several minutes while going to sleep or waking up.

Off-label medicines

In some special cases healthcare providers may prescribe medicines that are commonly used for other health conditions but are not yet approved by the FDA to treat insomnia. Some of these medicines may include antidepressants, antipsychotics, and anticonvulsants.

Over-the-counter medicines and supplements

Some over-the-counter (OTC) products that contain antihistamines are sold as sleep aids. Although these products might make you sleepy, talk to your doctor before taking them. Antihistamines can be unsafe for some people. Also, these products may not be the best treatment for your insomnia. Your doctor can advise you whether these products will help you.

Melatonin supplements are lab-made versions of the sleep hormone melatonin. Many people take melatonin supplements to improve their sleep. However, research has not proven that melatonin is an effective treatment for insomnia. Talk to your doctor before using these supplements. Dietary supplements can be beneficial to your health, but they can also have health risks.

The U.S. Food and Drug Administration regulates dietary supplements under a different set of regulations than those covering "conventional" foods and medicines. It does not have the authority to review dietary supplement products for safety and effectiveness before they are marketed.

Side effects of melatonin may include daytime sleepiness, headaches, upset stomach, and worsening depression. It can also affect your body's control of blood pressure, causing high or low blood pressure.

Other treatments

Your doctor may recommend that you use light therapy to set and maintain your sleep-wake cycle. With this treatment, you plan time each day to sit in front of a light box, which produces bright light similar to sunlight.

11.6.3 Sleep apnea

Sleep apnea is a common condition in the United States. It can occur when the upper airway becomes blocked repeatedly during sleep, reducing or completely stopping airflow. This is known as obstructive sleep apnea. If the brain does not send the signals needed to breathe, the condition may be called central sleep apnea.

Healthcare providers use sleep studies to diagnose sleep apnea. They record the number of episodes of slow or stopped breathing and the number of central sleep apnea events detected in an hour. They also determine whether oxygen levels in the blood are lower during these events.

Sleep studies can be done in a special center or at home. Studies at a sleep center can detect apnea events, detect low or high levels of activity in muscles that control breathing, monitor blood oxygen levels, brain and heart activity during sleep.

Other medical conditions that can cause sleep apnea are diagnosed in the following manner:

- Blood tests to check the levels of certain hormones and to rule out endocrine disorders that could be contributing to sleep apnea. Thyroid hormone can rule out hypothyroidism. Growth hormone tests can rule out acromegaly. Total testosterone and dehydroepiandrosterone sulphate (DHEAS) tests can help rule out polycystic ovary syndrome (PCOS).
- Pelvic ultrasound to examine the ovaries and detect cysts. This can rule out PCOS.

Doctors may want to know whether there is use of medicines, such as opioids, that could be affecting sleep or causing breathing symptoms of sleep apnea. Doctors may also want to know whether the patient has traveled recently to altitudes greater than 6,000 feet, because these low-oxygen environments can cause symptoms of sleep apnea for a few weeks after traveling.

Undiagnosed or untreated sleep apnea can lead to serious complications such as heart attack, glaucoma, diabetes, cancer, and cognitive and behavioral disorders.

Breathing devices such as continuous positive air pressure (CPAP) machines and lifestyle changes are common sleep apnea treatments.

11.6.4 Narcolepsy

Narcolepsy is a chronic neurological disorder that affects the brain’s ability to control sleep-wake cycles. People with narcolepsy may feel rested after waking, but then feel very sleepy throughout much of the day. Many individuals with narcolepsy also experience uneven and interrupted sleep that can involve waking up frequently during the night.

11.6.4.1 Symptoms

Narcolepsy is a lifelong problem, but it does not usually worsen as the person ages. Symptoms can partially improve over time, but they will never disappear completely. The most typical symptoms are excessive daytime sleepiness, cataplexy, sleep paralysis, and hallucinations. Though all have excessive daytime sleepiness, only 10 to 25 percent of affected individuals will experience all of the other symptoms during the course of their illness.

- **Excessive daytime sleepiness (EDS).** All individuals with narcolepsy have EDS, and it is often the most obvious symptom. EDS is characterized by persistent sleepiness, regardless of how much sleep an individual gets at night. However, sleepiness in narcolepsy is more like a “sleep attack”, where an overwhelming sense of sleepiness comes on quickly.
- **Cataplexy**. This sudden loss of muscle tone while a person is awake leads to weakness and a loss of voluntary muscle control. It is often triggered by sudden, strong emotions such as laughter, fear, anger, stress, or excitement. Some people may only have one or two attacks in a lifetime, while others may experience many attacks a day. Attacks may be mild and involve only a momentary sense of minor weakness in a limited number of muscles, such as a slight drooping of the eyelids. The most severe attacks result in a total body collapse during which individuals are unable to move, speak, or keep their eyes open. But even during the most severe episodes, people remain fully conscious, a characteristic that distinguishes cataplexy from fainting or seizure disorders. The loss of muscle tone during cataplexy resembles paralysis of muscle activity that naturally occurs during REM sleep.
- **Sleep paralysis**. The temporary inability to move or speak while falling asleep or waking up usually lasts only a few seconds or minutes and is similar to REM-induced inhibitions of voluntary muscle activity. Sleep paralysis resembles cataplexy except it occurs at the edges of sleep.

- **Hallucinations**. Very vivid and sometimes frightening images can accompany sleep paralysis and usually occur when people are falling asleep or waking up. Most often the content is primarily visual, but any of the other senses can be involved.

Additional symptoms of narcolepsy include:

- **Fragmented sleep and insomnia**. While individuals with narcolepsy are very sleepy during the day, they usually also experience difficulties staying asleep at night. Sleep may be disrupted by insomnia, vivid dreaming, sleep apnea, acting out while dreaming, and periodic leg movements.
- **Automatic behaviors**. Individuals with narcolepsy may experience temporary sleep episodes that can be very brief, lasting no more than seconds at a time. A person falls asleep during an activity (e.g., eating, talking) and automatically continues the activity for a few seconds or minutes without conscious awareness of what they are doing. This happens most often while people are engaged in habitual activities such as typing or driving. They cannot recall their actions, and their performance is almost always impaired.

In a normal sleep cycle, a person enters rapid eye movement (REM) sleep after about 60 to 90 minutes. Dreams occur during REM sleep, and the brain keeps muscles limp during this sleep stage, which prevents people from acting out their dreams. People with narcolepsy frequently enter REM sleep rapidly, within 15 minutes of falling asleep. Also, the muscle weakness or dream activity of REM sleep can occur during wakefulness or be absent during sleep. This helps explain some symptoms of narcolepsy. Narcolepsy affects both males and females equally. Symptoms often start in childhood, adolescence, or young adulthood (ages 7 to 25), but can occur at any time in life. It is estimated that anywhere from 135,000 to 200,000 people in the United States have narcolepsy. However, since this condition often goes undiagnosed, the number may be higher. Since people with narcolepsy are often misdiagnosed with other conditions, such as psychiatric disorders or emotional problems, it can take years for someone to get the proper diagnosis.

Individuals may be asked by their doctor to keep a sleep journal noting the times of sleep and symptoms over a one- to two-week period. Although none of the major symptoms are exclusive to narcolepsy, cataplexy is the most specific symptom and occurs in almost no other diseases.

A physical exam can rule out or identify other neurological conditions that may be causing the symptoms. Two specialized tests, which can be performed in a sleep disorders clinic, are required to establish a diagnosis of narcolepsy:

- **Polysomnogram (PSG or sleep study)**. The PSG is an overnight recording of brain and muscle activity, breathing, and eye movements. A PSG can help reveal whether REM sleep occurs early in the sleep cycle and if an individual's symptoms result from another condition such as sleep apnea.
- **Multiple sleep latency test (MSLT).** The MSLT assesses daytime sleepiness by measuring how quickly a person falls asleep and whether they enter REM sleep. On the day after the PSG, an individual is asked to take five short naps separated by two hours over the course of a day. If an individual falls asleep in less than 8 minutes on average over the five naps, this indicates excessive daytime sleepiness. However, individuals with narcolepsy also have REM sleep start abnormally quickly. If REM sleep happens within 15 minutes at least two times out of the five naps and the sleep study the night before, this is likely an abnormality caused by narcolepsy.

  Occasionally, it may be helpful to measure the level of hypocretin in the fluid that surrounds the brain and spinal cord. To perform this test, a doctor will withdraw a sample of the cerebrospinal fluid using a lumbar puncture (also called a spinal tap) and measure the level of hypocretin-1. In the absence of other serious medical conditions, low hypocretin-1 levels almost certainly indicate type 1 narcolepsy.

11.6.4.2 Causes

Narcolepsy may have several causes. Nearly all people with narcolepsy who have cataplexy have extremely low levels of the naturally occurring chemical hypocretin, which promotes wakefulness and regulates REM sleep. Hypocretin levels are usually normal in people who have narcolepsy without cataplexy.

Although the cause of narcolepsy is not completely understood, current research suggests that narcolepsy may be the result of a combination of factors working together to cause a lack of hypocretin. These factors include:

- **Autoimmune disorders**. When cataplexy is present, the cause is most often the loss of brain cells that produce hypocretin. Although the reason for this cell loss is unknown, it appears to be linked to abnormalities in the immune system.

- **Family history**. Up to 10 percent of individuals diagnosed with narcolepsy with cataplexy report having a close relative with similar symptoms.
- **Brain injuries**. Rarely, narcolepsy results from traumatic injury to parts of the brain that regulate wakefulness and REM sleep or from tumors and other diseases in the same regions.

11.6.4.3 Treatment

Early recognition and diagnosis of narcolepsy could be significant for certain treatment possibilities. Although under debate, autoimmunity is believed to be responsible for the hypocretin neuron loss. Some case reports and uncontrolled small studies suggest that immunomodulatory treatment, such as intravenous immunoglobulin, is able to ameliorate narcolepsy symptoms and influence hypocretin status (see Giannoccaro, 2020 for an overview). Based on the assumption that immunomodulatory treatment could prevent neuronal death, it should be administered as close to disease onset as possible. However, the presence of cataplexy already indicates the loss of the majority of hypocretinergic cells. Recognition of narcolepsy even before cataplexy onset is therefore crucial in order to modify the course of the pathological process. The emerging evidence of T-cell activation in the blood and CSF of narcolepsy type 1 patients in a very early stage of the disease might become useful in early recognition as well as timing of immunomodulatory treatment.

Although there is no cure for narcolepsy, some of the symptoms can be treated with medicines and lifestyle changes. When cataplexy is present, the loss of hypocretin is believed to be irreversible and lifelong. Excessive daytime sleepiness and cataplexy can be controlled in most individuals with medications.

11.6.4.3.1 **Medications**

- **Modafinil**. The initial line of treatment is usually a central nervous system stimulant such as modafinil, which is less addictive and has fewer side effects than older stimulants.
- **Amphetamine-like stimulants**. In cases where modafinil is not effective, doctors may prescribe amphetamine-like stimulants such as methylphenidate to alleviate EDS. The potential for side effects and abuse is higher so monitoring is necessary.
- **Antidepressants**. Two classes of antidepressant drugs have proven effective in controlling cataplexy in many individuals: tricyclics (including imipramine, desipramine, clomipramine, and protriptyline) and selective serotonin and noradrenergic reuptake inhibitors (including venlafaxine, fluoxetine, and atomoxetine).
- **Sodium oxybate**. Sodium oxybate (also known as gamma hydroxybutyrate or GHB) has been approved by the U.S. Food and Drug Administration to treat cataplexy and excessive daytime sleepiness in individuals with narcolepsy.

11.6.4.4 **Lifestyle changes**

Not everyone with narcolepsy can consistently maintain a fully normal state of alertness using currently available medications. Drug therapy should accompany various lifestyle changes such as taking short naps, maintaining a regular sleep schedule, avoiding caffeine, large meals and alcohol before bed, avoiding smoking, relaxing before bedtime, and exercising daily. Safety precautions while driving, operating heavy machinery and even walking down long flights of stairs are important.

11.6.5 Circadian Rhythm Disorders (Also known as Sleep-Wake Cycle Disorders)

Circadian rhythm disorders are problems that occur when your sleep-wake cycle is not properly aligned with your environment and interferes with your daily activities.

Your body tries to align your suprachiasmatic nuclei and endogenous sleep-wake cycle to cues from the environment, for example, when it gets light or dark outside, when you eat, and when you are physically active. When your sleep-wake cycle is out of sync with your environment, you may have difficulty sleeping, and the quality of your sleep may be poor. Disruptions of your sleep-wake cycle that interfere with daily activities may mean that you have a circadian rhythm disorder.

Disruptions in your sleep patterns can be temporary and caused by external factors such as your sleep habits, job, or travel. Jet lag and shift work are two major contemporary conditions that are a result of technology and our present social and economic realities. Or a circadian rhythm disorder can be long-term and caused by internal factors such as your age, your genes, or a medical condition. Symptoms may include extreme daytime sleepiness, insomnia, tiredness, decreased alertness, and problems with memory and decision-making.

To diagnose a circadian rhythm disorder, your doctor may ask about your sleep habits, suggest sleep tests, a diary to track when and how long you sleep, and test the levels of certain hormones in your blood or saliva. Your treatment plan will depend on the type and cause of your circadian rhythm disorder. Treatment may include light therapy, medicines to help you fall asleep or stay awake, or healthy lifestyle changes including steps to improve your sleep habits. If left untreated, circadian rhythm disorders may increase the risk of certain health problems or lead to workplace and road accidents.

There are many sleep-wake phase disorders which include Advanced sleep-wake phase disorder, Delayed sleep-wake phase disorder, Irregular sleep-wake rhythm disorder, and Non–24-hour sleep-wake rhythm disorder.

Circadian rhythm disorders cause problems with metabolism in shift workers. When shift work triggers a circadian rhythm disorder, it can disrupt your metabolism in a few ways. Normally, your biological clock helps control your hunger hormones. However, when you do not get enough good-quality sleep, your body makes less leptin, the hormone that tells your body when you are full, and more ghrelin, the hormone that tells your body you are hungry. You may respond by eating larger amounts of food than normal, as well as more fatty, sweet, and salty foods.

Improving health with current research

Learn about the following ways the NHLBI continues to translate current research into improved health for people who have circadian rhythm disorders. Research on this topic is part of the NHLBI's broader commitment to advancing sleep science and sleep disorders scientific discovery.

- NHLBI's National Center on Sleep Disorders Research (NCSDR). For 25 years, the NCSDR has led foundational research on sleep and circadian biology across the NIH and has worked with federal and private organizations to disseminate sleep health information. The NCSDR administers sleep and circadian research projects, training, and educational awareness programs, and serves as an NIH point of contact for federal agencies and public interest organizations. The Center also participates in research translation and dissemination of scientific sleep and circadian advances to healthcare professionals, public health officials, and the public.
- Advancing Circadian Rhythm Research. We have organized workshops to help direct future research into circadian rhythms and circadian rhythm disorders. These workshops have brought together experts in the fields of sleep and circadian rhythm research to identify pertinent areas of research into the role of circadian rhythms in the development and progression of several health conditions. Learn more about Developing Biomarker Arrays Predicting Sleep and Circadian-Coupled Risks to Health and Circadian-Coupled Cellular Function and Disease in Heart, Lung, and Blood.
- Investigating the Link between Circadian Rhythms and Lung Diseases. We have hosted workshops to address current gaps in our understanding of acute and chronic lung diseases. Our workshops have helped direct research to build upon preliminary discoveries showing that lung diseases such as asthma may be influenced by circadian rhythms. Learn more about The Circadian Clock's Influence on Lung Health.
- Research Conference on Sleep and the Health of Women. This 2018 conference focused on the importance of sleep for women's health. It showcased a decade of federally funded research advances in understanding the health risks, societal burden, and treatment options associated with sleep deficiency and sleep disorders in women. Topics discussed at this conference included the influence of sleep and circadian rhythms on alcohol consumption and cancer in women, and the social, environmental, and biological factors that affect sleep in women, including during pregnancy, postpartum, and menopause. Learn more from the Research Conference on Sleep and the Health of Women.
- Improving the Quality of Medical School Education on Sleep Disorders. As part of its efforts to ensure that research advances are utilized by healthcare providers, the NCSDR has supported the development of medical school curricula and durable educational materials on sleep disorders, including circadian rhythm disorders.
- Sleep Disorders Research Advisory Board (SDRAB). The NHLBI has administered this specialty program advisory panel since 1993. Board members, including medical professionals, federal partners, and members of the public, meet regularly to provide feedback to the NIH on sleep-related research needs and to discuss how to move sleep research forward. The SDRAB has supported advances to improve our understanding of circadian rhythms and circadian rhythm disorders. Visit the Sleep Disorders Research Advisory Board for more information.

- National Sleep Research Resource (NSRR). This resource was established by the NHLBI to provide biomedical researchers a large, well-characterized data collection from NIH-funded sleep research studies. These data can be used in new research studies to advance sleep research, including research into circadian rhythm disorders. Visit the National Sleep Research Resource external link for more information.

Learn more about how the NHLBI is contributing to knowledge about circadian rhythm disorders.

- Helping to improve the health and well-being of shift workers. Shift work is known to increase the risk of several medical conditions, including cardiovascular disease, diabetes, metabolic syndrome, and obesity. We have funded research to help understand how disruptions in the sleep-wake cycle can cause these conditions. We have also supported research that uses computer analysis to help understand shift work disorder.
- Identifying the genes that control circadian rhythms. We have supported research into discovering the genes that control our circadian clocks and how these genes help us align our circadian rhythms with the environment. In 2017, researchers supported by several NIH Institutes, including the NHLBI, won the Nobel Prize in physiology or medicine for discovering several genes that control circadian rhythms. View Wake-up call: 2017 Nobel Prize in Medicine awarded for studies of the body's internal clock for more information.

11.6.6 Summary of Disorders

Disorder	Causes	Symptoms	Treatments
Insomnia	Risk factors can include age, family history, environmental factors, stress and physical conditions.	Lying awake for a long time, sleeping for only short periods, waking up too early in the morning and/or having poor quality sleep.	Lifestyle changes can include - making bedroom sleep-friendly, going to sleep and waking up at same time everyday, avoiding caffeine, nicotine and alcohol, getting regular physical activity, avoiding daytime naps, eating meals on a regular schedule, limiting fluid intake close to bedtime, learning new ways to manage stress, and avoiding certain over-the-counter meds. Psycholotherapeutic treatments can include –cognitive therapy, relaxation or meditation therapy, sleep education, sleep restriction therapy, stimulus control therapy. Medical treatments can include – Benzodiazepines, Benzodiazepine receptor agonists, Melatonin receptor agonists, Orexin receptor agonists, sometimes antidepressants, antipsychotics and anticonvulsants are used, and Melatonin supplements
Sleep apnea	Obesity, large tonsils, endocrine disorders, neuromuscular conditions, heart or kidney failure, genetic syndromes, and premature birth	Reduced or absent breathing, known as apnea events, frequent loud snoring, gasping for air during sleep, excessive daytime sleepiness and fatigue, decreases in attention, vigilance, concentration, motor skills, and verbal and visuospatial memory, dry mouth or headaches when waking, sexual dysfunction or decreased libido, waking up often during the night to urinate	Healthy lifestyle changes include - Heart healthy eating, regular physical activity, aiming for healthy weight, healthy sleeping habits, and quitting smoking. Breathing devices such as continuous positive airway pressure (CPAP) machine. Oral devices including - Mandibular repositioning mouthpieces, tongue retaining devices Implants including those that detect breathing and nerve stimulator. Therapy for mouth and facial muscles called orofacial therapy. Surgical procedures including – Tonsillectomy, Maxillary/jaw advancement and tracheostomy

Narcolepsy		Unwillingly fall asleep in middle of driving, eating or talking, sudden muscle weakness (cataplexy), vivid dream-like images or hallucinations, total paralysis just before falling asleep or just after waking (sleep paralysis), excessive daytime sleepiness, fragmented sleep and insomnia, automatic behaviors	Medications including Modafinil, Amphetamine-like stimulants, Antidepressants and Sodium oxybate (also called GHB) Lifestyle changes including - taking short scheduled naps, maintaining regular sleep schedules, avoiding caffeine or alcohol before bed, avoiding smoking, exercising daily, avoiding large heavy meals right before bedtime, relaxing before bed and safety precautions particularly while driving.
Circadian rhythm or sleep wake cycle disorders	Genetic conditions, lifestyle issues like jet lag, environment or occupational factors, age, sex and other medical conditions like ASD and blindness	Consistent difficulty falling asleep, staying asleep, or both, excessive daytime sleepiness or sleepiness during shift work, fatigue and exhaustion, lethargy, decreased alertness and difficulty concentrating, impaired judgment and trouble controlling mood and emotions, aches and pains, including headaches, stomach problems, in people who have jet lag disorder	Healthy lifestyle changes, bright light therapy and melatonin

11.6.7 Summary

Since sleep is such a complex biological and psychological process, it should come as no surprise that there are many different kinds of disorders of sleep. These disorders have different symptoms, can be caused by a variety of different situational and biological issues, and can be treated in a variety of ways.

11.6.8 Attributions

- Psychology by Jeffrey C. Levy; licensed CC BY 4.0
- Sleep Science and Sleep Disorders: National Heart, Lung, and Blood Institute; National Institutes of Health; U.S. Department of Health and Human Services. Public Domain
- Insomnia: National Heart, Lung, and Blood Institute; National Institutes of Health; U.S. Department of Health and Human Services. Public Domain
- Sleep Apnea: National Heart, Lung, and Blood Institute; National Institutes of Health; U.S. Department of Health and Human Services. Public Domain

- Narcolepsy Fact Sheet: National Institute of Neurological Disorders and Stroke; National Institutes of Health; U.S. Department of Health and Human Services. Public Domain
- Recognizing the symptom spectrum of Narcolepsy to improve timely diagnosis: A narrative review by Quaedackers, L., Pillen, S., & Overeem, S. licensed under CC-BY NC 3.0

This page titled 11.6: Sleep Disorders is shared under a mixed license and was authored, remixed, and/or curated by ASCCC OERI & Bakhtawar Bhadha (ASCCC Open Educational Resources Initiative (OERI)) .

CHAPTER OVERVIEW

12: Ingestive Behaviors - Eating and Drinking

> This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

12.1: Maintaining Water Balance
12.2: Triggering Drinking Behavior - Osmometric and Volumetric Thirst
12.3: Ingestive Behavior Basics
12.4: Digestion and the Hormonal Control of Metabolism
12.5: Hunger - Theories, Detectors, and the Hypothalamus
12.6: Ingestive Behavior in Context

This page titled 12: Ingestive Behaviors - Eating and Drinking is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

12.1: Maintaining Water Balance

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Discuss the importance of water balance.
- Explain the role of hormones in maintaining water balance.
- Describe the renin-angiotensin-aldoesterone system and how it impacts blood volume and pressure.

12.1.1 Overview

As with eating behavior, environmental factors also impact our intake of fluids. But while we have varied sources of energy and mechanisms for storing that energy, the need to drink is triggered by a disruption in water balance. What makes thirst and drinking more complicated is the fact that water imbalances can be created by either a loss of fluids or an increase in solutes that causes water to move out of calls. In other words, just as it appears that we monitor both short-term and long-term energy stores to trigger hunger, we also monitor two different water stores to determine a need to drink. We'll explore this in more detail following an overview of the biological controls of water balance and the brain mechanisms that trigger thirst.

12.1.2 Maintaining Water Balance - Physiological Mechanisms

Maintaining a proper water balance in the body is important to avoid dehydration or over-hydration. The water concentration of the body is monitored by osmoreceptors in the hypothalamus, which detect the concentration of electrolytes in the extracellular fluid. The concentration of electrolytes in the blood rises when there is water loss caused by excessive perspiration, inadequate water intake, or low blood volume due to blood loss. An increase in blood electrolyte levels results in a neuronal signal being sent from the osmoreceptors in hypothalamic nuclei.

The hypothalamus produces a polypeptide hormone known as antidiuretic hormone (ADH, also known as vasopressin), which is transported to and released from the posterior pituitary gland. The principal action of ADH is to regulate the amount of water excreted by the kidneys. As ADH causes direct water reabsorption from the kidney tubules, salts and wastes are concentrated in what will eventually be excreted as urine. The hypothalamus controls the mechanisms of ADH secretion, either by regulating blood volume or the concentration of water in the blood. Dehydration or physiological stress can cause an increase in solute concentration which leads to ADH secretion and the retention of water, causing an increase in blood pressure. ADH travels in the bloodstream to the kidneys. Once at the kidneys, ADH changes the kidneys to become more permeable to water by temporarily inserting water channels, aquaporins, into the kidney tubules. Water moves out of the kidney tubules through the aquaporins, reducing urine volume. The water is reabsorbed into the capillaries returning the blood's solute levels back toward normal. As the blood solute concentration decreases, a negative feedback mechanism reduces osmoreceptor activity in the hypothalamus, and ADH secretion is reduced. ADH release can be reduced by certain substances, including alcohol, which can cause increased urine production and dehydration.

Chronic underproduction of ADH or a mutation in the ADH receptor results in diabetes insipidus, a condition characterized by the rapid loss of water following intake. If the posterior pituitary does not release enough ADH, water cannot be retained by the kidneys and is lost as urine. This causes increased thirst, but water taken in is lost again and must be continually consumed. If the condition is not severe, dehydration may not occur, but severe cases can lead to electrolyte imbalances due to dehydration.

Another hormone responsible for maintaining electrolyte concentrations in extracellular fluids is aldosterone, a steroid hormone that is produced by the adrenal cortex. In contrast to ADH, which promotes the reabsorption of water to maintain proper water balance, aldosterone maintains proper water balance by enhancing sodium (Na^+) reabsorption and potassium (K^+) secretion from extracellular fluid of the cells in kidney tubules. Because it is produced in the cortex of the adrenal gland and affects the concentrations of minerals Na^+ and K^+, aldosterone is referred to as a mineralocorticoid, a corticosteroid that affects ion and water balance. Aldosterone release is stimulated by a decrease in blood sodium levels, blood volume, or blood pressure, or an increase in

blood potassium levels. It also prevents the loss of Na^+ from sweat, saliva, and gastric juice. The reabsorption of Na^+ also results in the osmotic reabsorption of water, which alters blood volume and blood pressure.

Aldosterone production can be stimulated by low blood pressure, which triggers a sequence of chemical release, as illustrated in Figure **12.1.1**. When blood pressure drops, the renin-angiotensin-aldosterone system (RAAS) is activated. Cells in the juxtaglomerular apparatus, which regulates the functions of the nephrons of the kidney, detect this and release renin. Renin, an enzyme, circulates in the blood and reacts with a plasma protein produced by the liver called angiotensinogen. When angiotensinogen is cleaved by renin, it produces angiotensin I, which is then converted into angiotensin II in the lungs. Angiotensin II functions as a hormone and then causes the release of the hormone aldosterone by the adrenal cortex, resulting in increased Na^+ reabsorption, water retention, and an increase in blood pressure. Angiotensin II in addition to being a potent vasoconstrictor also causes an increase in ADH and increased thirst, both of which help to raise blood pressure.

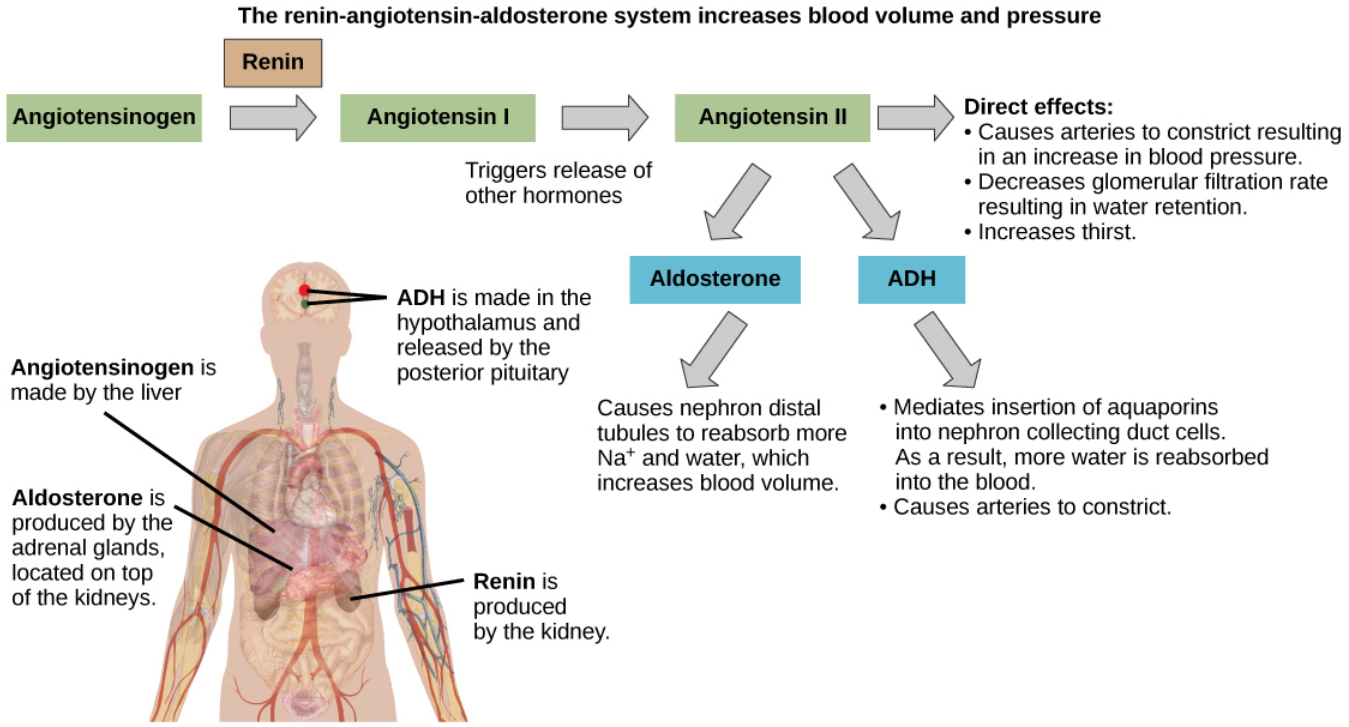


Figure **12.1.1**: ADH and aldosterone increase blood pressure and volume. Angiotensin II stimulates release of these hormones. Angiotensin II, in turn, is formed when renin cleaves angiotensinogen. (credit: modification of work by Mikael Häggström).

12.1.3 Maintaining Water Balance - Triggering Thirst

When your body starts to run low on water, a number of changes take place: for one, the volume of your blood decreases, causing a change in blood pressure. Because the amount of salt and other minerals in your body is staying constant as the volume of liquids decreases, their relative concentration increases (the same number of particles in a smaller volume means that the particles are more concentrated). This concentration of particles in bodily fluids relative to the total amount of liquid needs to be kept in a narrow range to keep the cells in your body functioning properly. Your body also needs a steady supply of fluids to transport nutrients, eliminate waste, and lubricate and cushion joints. To some extent, the body can compensate for water depletion by altering heart rate and blood pressure and by modifying kidney function to retain more water. For you, though, the most noticeable indication that your body is running low on fluids is likely the feeling of thirst, as you increasingly feel like you need to drink some water.

So how does your body know that these responses are necessary, and how are they coordinated across so many different organ systems? Research indicates that a highly specialized part of the brain called the lamina terminalis is responsible for guiding many of these thirst responses (Figure **12.1.2**). Brain cells within the lamina terminalis can sense when the body is running low on water and whether you've had anything to drink recently. When researchers manipulate this brain region, they can also drive animals to seek out or avoid water, regardless of how hydrated that animal might be.

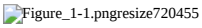


Figure 12.1.2: **Brain regions controlling thirst.** The lamina terminalis (yellow) is a series of interconnected brain structures that act as a central hub to control fluid levels in the body. Some cells in the lamina terminalis are adjacent to large, fluid-filled compartments in the brain, called ventricles (blue). When the body begins to run low on water, the composition of the body's fluids (including the fluid in the brain's ventricles) starts to change. The lamina terminalis neurons that border the ventricles can sense changes in the ventricular fluids, giving a snapshot of whether the body has enough water. These neurons also receive messages from other parts of the brain to give an even more complete picture of the body's water needs.

The lamina terminalis is located towards the front of the brain and occupies a prime location just below a fluid reservoir called the third ventricle. Unlike much of the rest of the brain, many cells in the lamina terminalis aren't guarded by a blood-brain barrier. This barrier prevents many circulating factors in the blood and other fluids from interacting with cells in the brain, offering the brain protection against potentially dangerous invaders like certain bacteria, viruses, and toxins. However, the blood-brain barrier also cuts the brain off from many circulating signals that might hold useful information about the body's overall status. Because certain cells in the lamina terminalis lie outside the blood-brain barrier, these cells can also interact with the fluid in the third ventricle to keep tabs on factors that indicate whether the body needs more or less water. In particular, these cells can monitor the fluid in the ventricle to determine its osmolality (the ratio of salt particles to a given amount of liquid) and the amount of sodium present.

When other parts of the brain detect information that's relevant to understanding the body's water needs, they frequently pass it along to the lamina terminalis, as well (Figure 12.1.2). In this way, the lamina terminalis also collects information about things like blood pressure, blood volume, and whether you've eaten recently (even before food can cause any change in circulating salt or water levels, your body tries to maintain a balance between these factors by encouraging you to drink water every time you eat). Information from the part of the brain that controls the circadian clock (the suprachiasmatic nucleus of the hypothalamus) also gets forwarded to the lamina terminalis, encouraging animals to drink more water before sleeping to avoid becoming dehydrated during long periods of sleep. Collectively, this information gives the lamina terminalis the resources needed to make a call about whether the body needs more or less water. In turn, cells in the lamina terminalis project to many other areas of the brain, sending out their verdict about current water needs. Although scientists are still trying to figure out exactly how information from the lamina terminalis affects other brain regions, it's clear that this output can influence an animal's motivation to seek out water, as well as physiological factors like kidney function and heart rate (Figure 12.1.3).

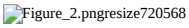


Figure **12.1.3**: **Thirst signals and their effects.** Neurons in the lamina terminalis receive many different messages about the body's water needs. Thanks to their location next to ventricles in the brain, they can directly sense key indicators of water need like sodium levels and osmolality (the ratio of salt particles to a given amount of liquid). They also receive information about what time of day it is from another brain region, as well as cues from the mouth and kidneys. Neurons in the lamina terminalis can pool all of this information to determine whether the body needs more or less water. If it needs more, they can trigger feelings of thirst and appetite suppression. If it needs less, the brain will send signals telling you to stop drinking. The lamina terminalis also sends messages to a brain region called the hypothalamus. In turn, the hypothalamus can affect heart rate or urge the kidneys to retain more or less water.

12.1.3.1 What makes water so refreshing?

After a while standing outside in the hot sun, a cold drink of water tends to feel instantly refreshing. You might also find that drinking a very sugary beverage feels equally refreshing but leaves you feeling thirsty again later. In both cases, it takes tens of minutes for that drink to have any effect on attributes like osmolality or blood pressure, the body's main indicators of hydration status. Instead, the brain must rely on some other cue to tell you to stop drinking and give you that instant feeling of refreshment.

12.1.3.1.1 The Role of the Lamina Terminalis

Researchers have discovered that neurons in the lamina terminalis respond to the physical act of swallowing liquids, even before there are any changes in the amount of water in the blood. Researchers have identified a group of neurons in the lamina terminalis whose activity is required for drinking behaviors: when you artificially turn off the activity in these cells, mice no longer drink water, even when they are water deprived (Zimmerman et al., 2017). When the researchers recorded the activity of these cells as animals drank water, they found that the cellular activity decreased in lockstep with each sip of water, far before any physiological changes in blood pressure or solute concentration could have an effect. They also found that this change in activity only happened when the mice drank water, not when they drank a salt solution. This study suggests that our brains have a built-in mechanism to compare how much water we need with the amount of water we're currently drinking, telling us when we've had enough and leaving us feeling instantly hydrated. Still, scientists don't know exactly how the brain can tell water apart from other liquids, or why drinking some non-water beverages can leave you feeling instantly hydrated, as well.

12.1.3.1.2 The Role of Dopamine

Another group of researchers set out to tackle the problem of why we find drinking water so rewarding when we're thirsty (Augustine at al., 2019). Neuroscientists have long recognized the role of dopamine in determining what we find reinforcing. In order to look at the role that dopamine plays with respect to drinking behaviors, these researchers used a new kind of sensor that glows in the presence of dopamine. By putting this sensor into a mouse's brain, they were able to record dopamine levels in real time as the mouse went about its tasks (Figure **12.1.4**.

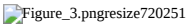


Figure **12.1.4**: **Drinking water is rewarding.** Researchers in Yuki Oka's group at Caltech conducted a study to see why animals find water rewarding. By using a special kind of sensor that glows in the presence of the rewarding molecule dopamine, they could see what kinds of liquids caused dopamine release. They recorded large spikes of dopamine release when thirsty mice drank both water and salty saline solutions, indicating that mice found both of these liquids rewarding. When researchers injected water into thirsty mice, though, they found no changes in dopamine levels, even though the injected water would also hydrate the thirsty animals.

These researchers looked at dopamine levels after thirsty mice drank water and other liquids. They also recorded dopamine levels after they injected water directly into the gastrointestinal system; this procedure hydrated thirsty animals, but meant that the mice didn't actually drink any water. Oka's group found that thirsty mice had a large surge in dopamine levels after drinking either water or saline and that these dopamine changes happened even before drinking would have any effect on blood fluid levels. In contrast, the animals didn't release any dopamine after water was pumped into their gastric systems, suggesting that it's the act of drinking itself that's rewarding—not the feeling of being hydrated. This effect also starts to explain why drinking beverages other than water can be so satisfying, even when they leave you feeling thirsty later: the dopamine spike that comes from drinking liquids when animals are thirsty doesn't depend on what kind of liquid they're drinking, even though not all liquids are equally hydrating.

These two studies highlight the varying strategies the brain uses to monitor essential nutrients like water; because no single sensor can tally current water levels and predict future water needs, the brain relies on a variety of sensations and cues.

12.1.4 Summary

XXXX

12.1.5 References

Augustine, V., Ebisu, H., Zhao, Y., Lee, S., Ho, B., Mizuno, G. O., Tian, L., & Oka, Y. (2019). Temporally and spatially distinct thirst satiation signals. *Neuron*, *103*(2). https://doi.org/10.1016/j.neuron.2019.04.039

Zimmerman, C. A., Lin, Y. C., Leib, D. E., Guo, L., Huey, E. L., Daly, G. E., Chen, Y., & Knight, Z. A. (2016). Thirst neurons anticipate the homeostatic consequences of eating and drinking. *Nature*, *537*(7622), 680–684. https://doi.org/10.1038/nature18950

12.1.6 Attributions

"Maintaining Water Balance - Physiological Mechanisms" adapted from Rye, C., Wise, R., Jurukovski, V., DeSaix, J., Choi, J., & Avissar , Y. (2016, October 21). *37.3 Regulation of Body Processes - Biology*. OpenStax. Retrieved February 6, 2022, from https://openstax.org/books/biology/p...body-processes (CC BY)

"Maintaining Water Balance - Triggering Thirst" adapted from Frank, M. (2019, September 26). *The Neuroscience of Thirst: How your brain tells you to look for water.* Retrieved May 9, 2022 from https://sitn.hms.harvard.edu/flash/2...ls-look-water/ (CC BY-NC-SA)

12.1.7

This page titled 12.1: Maintaining Water Balance is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

- **37.3: Regulation of Body Processes** by OpenStax is licensed CC BY 4.0.

12.2: Triggering Drinking Behavior - Osmometric and Volumetric Thirst

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Discuss the role of solutes in determining the movement of water in the body.
2. Distinguish between intracellular and extracellular fluid.
3. Compare and contrast osmometric and volumetric thirst.

12.2.1 Overview

When talking about thirst, we are talking not just about water in the body, but also the major solute (dissolved substance) we find in this water, salt or sodium chloride (NaCl). Regulating the body's water is necessarily impacted by the substances that are dissolved in that water. Thirst is a signal to the body that there has been a loss of fluid or, more specifically, that there is a fluid imbalance. About two-thirds of the body's water is intracellular, meaning within cells. The other third, the extracellular fluid, consists of interstitial fluid (the fluid bathing cells), cerebrospinal fluid (fluid in the ventricular system), and blood plasma (intravascular).

Because we effectively have two places where water is found (intracellular and extracellular), we have two systems for monitoring the body's fluid levels - and a variety of ways of inducing thirst, as well as different types of thirst. In addition, salt appetite is necessarily associated with one form of thirst as the need created by the fluid loss requires both salt and water to restore homeostasis. One system focuses on the levels of intracellular fluid and triggers osmometric thirst. The other monitors extracellular levels - more specifically, plasma or blood volume - and triggers volumetric thirst. Volumetric thirst is associated wth a neeed for both salt and water.

12.2.2 Intracellular Fluid Volume

Intracellular fluid volume is controlled by the concentration of solutes in the interstitial fluid (the fluid outside of cells). Under normal circumstances, the fluid outside and inside the cell is isotonic (the concentration of solutes is equal). If, however, the solute concentration is increased in the interstitial fluid (due to ingestion of solutes or loss of water), water will leave the cell by the process of osmosis. Osmosis is simply movement of water from an area of low solute concentration to an area of high concentration, water will move as needed to even out the solute concentration (Figure **12.2.1**). If the interstitial fluid (the fluid outside of the cell) is hypertonic (if it has a higher solute concentration), water will leave the cell. If the interstitial fluid is hypotonic, water will enter the cell. Both conditions may be dangerous - disrupting normal neuronal function. When there is a difference between the solute concentration of the interstitial fluid and the intracellular fluid, the movement of water by osmosis causes changes in intracellular volume.

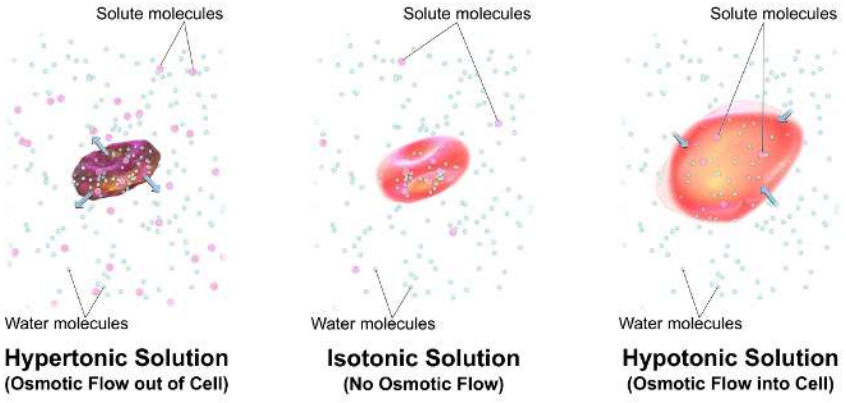


Figure **12.2.1**: Water movement in hypertonic, isotonic and hypotonic solutions. (Images from Blausen.com staff (2014). "Medical gallery of Blausen Medical 2014". WikiJournal of Medicine 1 (2). DOI:10.15347/wjm/2014.010. ISSN 2002-4436., CC BY 3.0 <https://creativecommons.org/licenses/by/3.0>, via Wikimedia Commons)

Generally, there is no need to regulate the volume of the interstitial fluid, although we do regulate its tonicity (solute concentration) through its effects on the intracellular fluid volume. More specifically, we monitor the movement of water.

12.2.3 Plasma Volume (Extracellular)

The other important and closely regulated fluid compartment is the blood plasma. If there is a loss of plasma volume (hypovolemia), this can impair the functioning of the heart. If increased, there may be a dangerous increase in blood pressure. If hypovolemia is severe enough the heart can no longer pump effectively. The vascular system can correct for loss of volume, but only within a limited range.

12.2.4 Two Variables, Two Regulatory Mechanisms

In order to maintain optimal fluid balance, the two variables are monitored:

- Movement of water (intracellular)
- Plasma volume (extracellular)

As there are two variables being monitored, there will be two different regulatory mechanisms that underlie thirst. The monitoring of fluid levels is tied to the monitoring of sodium levels. There are two mechanisms for dealing with a need for fluid and salt or sodium - we see both physiological and behavioral mechanisms for dealing with a need for water and salt:

- Physiologically - At the kidney, the excretion (loss) of water and salt can be modified. And alterations in heart rate and blood pressure can also compensate for losses.
- Behaviorally - ingest water and salt

What is the role of the kidneys and how is kidney activity regulated? Typically, we ingest much more salt and water than we need and what we don't need is excreted. Blood is essentially filtered by the kidneys. The functional units of the kidneys, the nephrons, extract fluid from the blood and carry it (collect it into) the ureter. From the ureter, urine passes to the urinary bladder where urine is stored.

As discussed in the prior section, two hormones are involved in the excretion of sodium and water by the kidney - both hormonal signals increase retention:

- Aldosterone is secreted by the adrenal cortex.
- Vasopressin (Anti Diuretic Hormone or ADH) is secreted by the posterior pituitary, but produced by the hypothalamus. The two names by which this hormone is known reflect two aspects of its effects, "vasopressin" is a reference to its ability to cause contraction of blood vessels while ADH reflects its role in preventing excretion.

What happens without vasopressin? No water is retained. Without vasopressin, diabetes insipidus develops, leading to excessive water loss and increased thirst. Incidentally, the term diabetes insipidus literally means "a tasteless passing through". The urine is so dilute, it has little taste. It is, as you might expect, treated with vasopressin - in the form of a nasal spray. The diabetes you more

commonly hear of (due to a lack of or insensitivity to insulin) is technically diabetes mellitus. Diabetes, again, means "passing through". Mellitus means "sweet" as the urine of the diabetic would be sweet with unused glucose.

12.2.5 Osmometric (Osmotic) Thirst - Detecting Water Movement Out of Cells

So, we can drink to replenish lost fluid and the kidney's function can conserve or allow fluids to be lost. As we have two variables that we are monitoring, we have two different types of thirst. Osmometric thirst derives its name from the word osmosis. Osmometric thirst is triggered by loss of volume from the intracellular fluid stores. This form of thirst is produced when the solute concentration or tonicity of the interstitial compartment (the fluid outside of the cell) is increased, causing water to move out of the cell. Detectors, osmoreceptors, detect the loss of water. Interestingly, it is more than just the change in the solute concentration that triggers this form of thirst. It is the actual effect of this change on the water in the intracellular fluid. If water does not move out of the cell, drinking will not increase. If the concentration of interstitial solutes is increased with something that passes into the cell - so the extracellular and intracellular solute concentrations remain isotonic and there's no net movement of water - drinking will not be triggered. In other words, it isn't just changes in the solute concentration that are detected, it is the loss of water from the intracellular compartment that matters. An increase in fluid intake is only seen when there is an increase in interstitial solute concentration that results in the movement of water out of cells. The movement of water out of cells (cellular dehydration) triggers osmometric thirst.

Where is this change detected? Where are the osmoreceptors? We can find hints of this by looking at the effects of various solutes on thirst.

If urea is injected, the extracellular solute concentration is increased. Urea does pass into cells, but it only crosses the blood brain barrier slowly. In other words, there's no movement of water in the cells of the body, but as you've increased the solute concentration outside of the brain, there will be a net movement of water out of the brain. Not surprisingly, even a slight and temporary decrease in the intracellular fluid concentration of the brain results in thirst. It also makes sense that osmoreceptors would be found in the brain as the brain needs to be protected from changes and it is the brain that triggers behavior. A simpler (and less disturbing) way of demonstrating this is by injecting hypertonic saline directly into the brain. Osmoreceptors can be classified as central or peripheral osmoreceptors based on their location. The central receptors are primarily present in the anterior hypothalamus, including the organum vasculosum laminae terminalis (OVLT) and the subfornical organ (SFO) (Danziger & Zeidel, 2015; Xu et al., 2000; Muhsin & Mount, 2016). These central receptor cells have respond to osmotic changes and they are sensitive to angiotensin II (abbreviated as AII; Bichet, 2012; Benarroch, 2011). Some SFO/OVLT neurons also receive signals from peripheral arterial baroreceptors. Thus, SFO/OVLT neurons sense plasma osmolality, volume, and pressure to control thirst. These cells depolarize due to increased Na^+ concentration, cell shrinkage, angiotensin II or negative suction pressure, and discharge neuronal spikes, which later initiate the sensation of thirst or vasopressin release or both (Muhsin & Mount, 2016; Gizowski & Bourque, 2018).

One of the central receptor organs, SFO, has two distinct types of neurons with opposing actions. A glutamatergic population (SFO-GLUT) that promotes thirst and sodium intake, and a GABAergic population (SFO-GABA) that inhibits thirst (Zimmerman, Leib, & Knight, 2017). The proper agonistic and antagonistic functioning of these cells maintain the optimum level of hydration. As the SFO has access to the systemic blood, it has a markedly different ependymal surface with a flattened appearance, lacks normal cilia count, and has tight junctions between adjacent cells. These features help block the diffusion of substances across the SFO parenchyma into the third ventricle. This peculiar position of the SFO cells is suitable for sensing both plasma and CSF components (Hiyama & Noda, 2016).

OVLT, the other central receptor, is close to the median preoptic nucleus. The peripheral receptors which are present within the upper gastrointestinal tract and portal venous system also detect changes in solute concentration and blood volume via a specific type of receptor (Danziger & Zeidel, 2015; Xu et al., 2000; Muhsin & Mount, 2016; Xu et al., 2000). They act as a supplementary center for osmoregulation in addition to the central osmoreceptors (Muhsin & Mount, 2016).

12.2.6 Volumetric (Hypovolemic) Thirst - Detecting Blood Loss

Volumetric thirst is triggered when there is a loss of volume from the extracellular fluid stores. Volumetric thirst is triggered by a decrease in blood volume referred to as hypovolemia. Hypovolemia is usually accompanied by a drop in blood pressure (hypotension). Hypovolemia and hypotension then ultimately lead to the production of angiotensin II. As discussed in the last

section, a cascade of events is triggered when the kidneys detect a decrease in blood pressure and release the enzyme renin. This activates the renin-angiotensin-aldosterone system. Angiotensinogen, produced by the liver, is acted on by renin to produce angiotensin I, which is then converted into angiotensin II. Angiotensin II functions as a hormone and then causes the release of the hormone aldosterone by the adrenal cortex, resulting in increased sodium reabsorption, water retention, and an increase in blood pressure.

But how do we produce volumetric thirst experimentally? Obviously, we could remove blood. But removing a large volume of blood creates other problems. A better method is to inject a colloid. A colloid is a large molecule that can't cross cell membranes. By placing a colloid under the skin or in the abdominal cavity, extracellular fluid will move out of tissues. As the interstitial fluid decreases, fluid from the plasma fills the vacancy. As a consequence of the resulting drop in blood volume, vasopressin is released and the subject drinks. Early studies using this procedure documented the presence of both thirst (Fitzsimons, 1961) and sodium appetite (Stricker & Wolf, 1966) as indicated by ingestion of both. Later work (Stricker, 1981) explored the effects of colloid treatment on both water and saline drinking in greater detail. Stricker's work explored the effects of various manipulations on thirst and sodium appetite, revealing that sodium appetite accompanies thirst when dietary sodium is controlled. In other words, earlier studies had found that there was a delay in sodium appetite that may be accounted for by the excess of sodium that had previously been ingested. The separate timelines associated with thirst and sodium appetite indicate separate systems for detecting the needs for water and sodium.

Following extracellular fluid loss - loss of plasma volume - there is a need for both salt and water. A rat will now drink the saline solution that it normally rejects. This indicates that there is a mechanism for monitoring the body's sodium concentration, so that needs are detected and then mechanisms are employed to satisfy them. Unlike osmometric thirst, volumetric thirst is accompanied by a salt appetite.

Angiotensin II, the 3rd "step" when the renin-angiotensin-aldosterone system is activated, travels through the blood stream and exerts multiple effects.

- Angiotensin II acts at the adrenal cortex to stimulate aldosterone secretion.
- Angiotensin II acts at the posterior pituitary to stimulate vasopressin (ADH) secretion.
- Angiotensin II acts at the muscles of the the small arteries to cause contraction (increasing blood pressure).
- Angiotensin II triggers drinking and salt ingestion.

Although angiotensin II is thought to produce thirst and act on other brain sites, it does not cross the blood-brain barrier. Thus, in order to cause thirst and other effects, angiotensin II must act at a part of the brain that lacks this barrier. This is characteristic of the circumventricular organs (organs surrounding the ventricles), which includes the organum vasculosum laminae terminalis (OVLT) and the subfornical organ (SFO) mentioned earlier. The subfornical organ (SFO) appears to be the site at which angiotensin II acts to cause volumetric thirst.

Evidence for the role of the SFO in water ingestion:

- Very low doses of angiotensin II at the SFO cause drinking.
- Angiotensin II-induced drinking is blocked by blocking angiotensin II receptors at the SFO.
- Electrical stimulation of the SFO produces drinking.

Furthermore, neurons in the SFO increase their activity in response to angiotensin II, even when neural connections are cut - demonstrating that the neurons are responding to the AII.

The SFO has outputs to several brain regions. And these connections mediate the endocrine, autonomic, and behavioral responses associated with thirst. It is involved in a number of the effects of AII. How do the SFO and angiotensin II work together to produce all the effects associated with angiotensin II?

- Secretion of aldosterone by the kidney is a direct response to AII at the kidney. The other effects of AII are initiated at the brain.
- Endocrine outputs of the SFO communicate with the hypothalamic regions controlling release of vasopressin.
- Autonomic outputs of the SFO act on hypothalamic regions influencing the activity of the autonomic nervous system.
- Behavioral outputs of the SFO act on the basal forebrain. If these efferent connections of the SFO are lesioned, we no longer see angiotensin II-induced drinking.

So, there are receptors in the kidney that detect changes in blood flow. There is a second set of volumetric receptors in the heart. These are baroreceptors, they are stretch receptors that detect the amount of blood in the heart's atria (the part of the heart that is receiving blood from the veins).

12.2.7 Normal Drinking

Just as we typically eat long before a physiological need for fuel would trigger hunger, we drink before change in the body's fluid stores would be detected. As with eating, we have learned elements of our drinking behavior. When do we habitually drink? Normally, we drink with meals. While eating does produce a need for fluid, we typically drink before that need would normally be detected. Food in the digestive system causes water to be diverted to these areas and absorbed food increases the solute concentration of the plasma and produces osmometric thirst.

Why do animals drink before a need is detected? Animals will actually learn to drink more if their diet is changed such that they need to drink more. With a high protein diet, more water is needed. Animals will learn to drink more with a meal in anticipation of the need.

What signals or causes drinking with a meal? The movement of water into the digestive system produces hypovolemia. The hypovolemia causes the kidneys to secrete renin and AII levels are increased. During a normal meal the levels of renin actually double. If we then block AII production, we'll see a decrease in drinking with a meal.

12.2.8 Summary

In order to maintain an appropriate fluid balance, intracellular and extracellular fluid levels are monitored. Intracellular or osmometric thirst is triggered when water moves out of cells. In contrast, extracellular thirst and an appetite for salt is triggered when the loss of blood volume creates a need for salt and water.

12.2.9 Attributions

"Osmometric Thirst - Detecting Water Movement Out of Cells" adapted from Koshy, R. & Jamil, R. (2021). Physiology, Osmoreceptors. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK557510/ (CC BY)

12.2.10 References

Leib, D. E., Zimmerman, C. A., & Knight, Z. A. (2016). Thirst. *Current biology : CB*, *26*(24), R1260–R1265. https://doi.org/10.1016/j.cub.2016.11.019

Stricker E. M. (1981). Thirst and sodium appetite after colloid treatment in rats. *Journal of comparative and physiological psychology*, *95*(1), 1–25. https://doi.org/10.1037/h0077764

This page titled 12.2: Triggering Drinking Behavior - Osmometric and Volumetric Thirst is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

12.3: Ingestive Behavior Basics

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Explain what a drive state is and the properties of a drive state.
2. Discuss drive-reduction theories and their limitations.
3. Define homeostasis and identify its elements.
4. Discuss body temperature as an example of a homeostatic system.
5. Compare and contrast homeotherms and poikilotherms.

12.3.1 Overview

"Ingestive Behavior", eating and drinking, is a fascinating topic when considered from the perspective of a biological psychologist. Why? The connection between the biological need for food (energy, nutrients) and water (hydration) lead to drive states (hunger and thirst) that motivate us to engage in behaviors to meet those needs. In other words, the connection between the body's need for food and water, the drive states of hunger and thirst, and seeking food and drink (the behaviors designed to meet the body's needs) is something that we are all familiar with. At the same time, the impact of learning on what we eat, when we eat, and even how we feel about food is also something that we're aware of. And there even is a link to mental illness - instances where the mind leads to ingestive behaviors (such as overeating or under-eating) that are harmful to the body. While it would be easy to create an entire book (or course) that focuses on ingestive behavior, we are going to explore the topic by focusing on the biological bases for eating and drinking - with an emphasis on how we know what we know about these surprisingly complex behaviors.

What are the biological mechanisms that determine when we need to eat or drink - and how does the detection of a need trigger behaviors intended to meet that need? And how are those hunger and thirst signals turned off? Consider what causes you to "feel" hungry - where would you look for detectors of hunger or satiety (being "full")?

Hunger is a drive state, an affective experience (a mood or a feeling) that motivates organisms to fulfill goals that are generally beneficial to their survival and reproduction. Like other drive states, such as thirst or sexual arousal, hunger has a profound impact on the functioning of the mind. It affects psychological processes, such as perception, attention, emotion, and motivation, and influences the behaviors that these processes generate. Drive states serve to motivate a variety of behaviors.

12.3.2 Drive-Reduction Theories

Drive-reduction theories focus on how motivation originates from biological needs or drives. In these theories, it is proposed that behavior is an external display of the desire to satisfy physiological needs. In other words, a biological need creates a drive and reducing that drive motivates an organism to engage in behaviors to reduce the drive, presumably by meeting the biological need. As this theory focuses on biological needs, it is connected to the principle of homeostasis and the use of behaviors to maintain or restore homeostasis. Homeostasis refers to the stable state across all physiological systems that organisms strive to maintain and will be discussed shortly.

A “drive” is a state of arousal or tension triggered by a physiological or biological need. These needs include those that we are all aware of and that are the focus of this chapter, hunger and thirst. Drive-reduction theory proposes that drives give rise to motivation. When a drive emerges, it creates an unpleasant state of tension that leads to behaviors intended to reduce this tension. To reduce the tension, the organism will begin seeking out ways to satisfy the biological need. For instance, you will look for water to drink if you are thirsty. You will seek food if you are hungry.

According to the theory, any behavior that reduces the drives will be repeated by humans and animals. This is because the reduction of the drive serves as a positive reinforcement for the behavior that caused such drive reduction.

While drive-reduction theory explains how primary reinforcers (those things that are naturally reinforcing because they meet biological needs) are effective in reducing drives, many psychologists argue that the theory is not applicable in the concept of

secondary reinforcers. For example, money is a powerful secondary reinforcer, as it can be used to purchase primary reinforcers like food and water. However, money in itself cannot reduce an individual's drives. Another problem with the theory is that it does not provide an explanation about the reason behind people engaging in behaviors that are not meant to reduce drives, such as a person eating even if they are not hungry.

(Photo by Nikita Tikhomirov on Unsplash)



12.3.3 Key Properties of Drive States

Drive states differ from other affective or emotional states in terms of the biological functions they accomplish. Whereas all affective states are either desirable or undesirable and serve to motivate approach or avoidance behaviors (Zajonc, 1998), drive states are unique in that they generate behaviors that result in specific benefits for the body. For example, hunger directs individuals to eat foods that increase blood sugar levels in the body, while thirst causes individuals to drink fluids that increase water levels in the body.

Different drive states have different triggers. Most drive states respond to both internal and external cues, but the combinations of internal and external cues, and the specific types of cues, differ between drives. Hunger, for example, depends on internal, visceral signals as well as sensory signals, such as the sight or smell of tasty food. Different drive states also result in different cognitive and emotional states, and are associated with different behaviors. Yet despite these differences, there are a number of properties common to all drive states. Most notably, the link between drives and motivation (as already discussed) and the common features of homeostatic mechanisms.

12.3.4 Homeostasis

Humans, like all organisms, need to maintain a stable state in their various physiological systems. For example, the excessive loss of body water results in dehydration, a dangerous and potentially fatal state. However, too much water can be damaging as well. Thus, a moderate and stable level of body fluid is ideal. The tendency of an organism to maintain this stability across all the different physiological systems in the body is called homeostasis.

Homeostasis is maintained by two mechanisms. First, the state of the system being regulated must be monitored and compared to an ideal or optimal level, a set point. Second, there need to be mechanisms for moving the system back to this set point - that is, to restore homeostasis when deviations from it are detected. While our focus here is on homeostatic mechanisms that connect to behavior, there is a wide-range of such systems throughout the body. Most notably, you have already been introduced to the concept of osmosis and the mechanisms that exist to maintain a stable intracellular environment. To better understand the functioning of a homeostatic system, think of the thermostat in your own home. It detects when the current temperature in the house is different from the temperature you have it set at (i.e., the set point). Once the thermostat recognizes the difference, the heating or air conditioning turns on (or off) to bring the overall temperature back to the designated level.

Many homeostatic mechanisms, such as blood circulation and immune responses, are automatic and strictly physiological. Others, however, involve behavioral responses and deliberate action. Most drive states motivate action to restore homeostasis using both "punishments" and "rewards" to modify behavioral responses. Imagine that these homeostatic mechanisms are like molecular parents. When you behave poorly by departing from the set point (such as not eating or being somewhere too cold), they raise their voice at you. You experience this as the bad feelings, or "punishments," of hunger, thirst, or feeling too cold or too hot. However, when you behave well (such as eating nutritious foods when hungry), these homeostatic parents reward you with the pleasure that comes from any activity that moves the system back toward the set point. For example, when body temperature declines below the set point, any activity that helps to restore homeostasis (such as putting one's hand in warm water) feels pleasurable; and likewise, when body temperature rises above the set point, anything that cools it feels pleasurable. While homeostatic mechanisms are often likened to a thermostat that determines what the desired value is and then systems are turned on and off in order to establish the optimal value, both hunger and thirst are not only triggered by need. Both systems anticipate need - they are both proactive and reactive. Before exploring these ingestive behaviors further, we'll consider temperature regulation in more detail as an introduction to how homeostatic mechanisms operate.

12.3.4.1 Elements of a Homeostatic System

Regardless of the homeostatic system we are discussing, it must have specific features.

- System variable - what's being regulated.
- Set point - optimal value.
- Detector - a mechanism must exist for monitoring the levels of the variable of interest.
- Correctional mechanisms - there must be a way for deviations to be altered.

When it comes to body temperature, there is an ideal temperature that an organism strives to maintain. Temperature would be the system variable and that ideal temperature would be the set point. The detector or detectors would be temperature sensors in the skin and brain. What are the correctional mechanisms? Temperature is interesting in that these mechanisms can be behavioral and physiological for some species, while for others they are only behavioral. If a human is cold, they can put on a coat (behavioral) and they may shiver, a physiological response that generates heat. In contrast, some animals do not have the benefit of those physiological mechanisms for adjusting temperature and must rely exclusively on behavioral responses. Mammals and birds are homeotherms, they possess both physiological and behavioral means of altering body temperature. Homeotherms are able to maintain the same body temperature, our temperatures don't fluctuate with changes in environmental temperature. Poikilotherms regulate their body temperature entirely with behavioral means. Reptiles, amphibians, and many fish are poikilotherms.

Regardless of the homeostatic system in question, all have a system variable, a set point, a detector, and at least one correctional mechanism. When we need food or water, physiological and behavioral mechanisms assist us in meeting the need. In contrast, not all species have physiological mechanisms for correction for deviations in body temperature.

12.3.4.2 Negative Feedback

All regulatory mechanisms employ negative feedback - if the optimal value is exceeded, responses will be turned off. But are corrective mechanisms only "turned off" when the desired value is obtained? The answer is no - we stop eating prior to the restoration of blood glucose levels and we start to drink before we actually have a fluid need. So, the systems predict needs and predict when needs have been satisfied.

12.3.5 Hunger and Thirst - The Drives Motivating Eating and Drinking

Hunger is a classic example of a drive state, one that results in thoughts and behaviors related to the consumption of food. In the simplest of terms, hunger is generally triggered by low glucose levels in the blood (Rolls, 2000), and behaviors resulting from hunger aim to restore homeostasis regarding those glucose levels. But various other internal and external cues can also cause hunger. For example, when fats are broken down in the body for energy, this initiates a chemical cue that the body should search for food (Greenberg, Smith, & Gibbs, 1990). External cues include the time of day, estimated time until the next feeding (hunger increases immediately prior to food consumption), and the sight, smell, taste, and even touch of food and food-related stimuli. Note that while hunger is a generic feeling, it can also lead to the eating of specific foods that correct for nutritional imbalances we may not even be conscious of. For example, a couple who was lost adrift at sea found that they began to crave the eyes of fish. Only later, after they had been rescued, did they learn that fish eyes are rich in vitamin C - a very important nutrient that they had been depleted of while lost on the ocean (Walker, 2014).

We'll begin our exploration of ingestive behaviors with eating, the behavior motivated by hunger. While you might think the control of eating is quite simple, the regulated variables involved in eating are not so well described. Why do we eat? We have a whole host of needs that must be satisfied by the ingestion of food.



External cues, like the sight and smell of food, can ignite feelings of hunger. [Image: Marco Verch, https://goo.gl/c4TC5A, CC BY 2.0, https://goo.gl/BRvSA7]

12.3.6 Attributions

Adapted from Bhatia, S. & Loewenstein, G. (2022). Drive states. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology.* Champaign, IL: DEF publishers. Retrieved from http://noba.to/pjwkbt5h (CC BY-NC-SA)

"Drive-Reduction Theories" Adapted from Sincero, S. M. (Jul 10, 2012). Drive-Reduction Theory. Retrieved from https://explorable.com/drive-reduction-theory (CC BY 4.0).

This page titled 12.3: Ingestive Behavior Basics is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

12.4: Digestion and the Hormonal Control of Metabolism

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Distinguish between digestion and metabolism.
- Described the process of digestion, noting the roles of the mouth, stomach, intestines (small and large), and accessory organs.
- Describe how hormones regulate metabolism.
- Describe the three phases of digestion.

12.4.1 Overview

While we know that eating is motivated by the drive state of hunger, what is being regulated? What is the system variable - or variables - that are monitored?

Food:

- Provides energy.
- Provides the building blocks that we use to construct and maintain our bodies - we don't just need calories (energy, fuel), we need to eat certain things - ideally we are monitoring not only the quantity (how much) of food we eat, but the quality (what we eat). In other words, an ideal system for the control of our eating would be one that triggers a hunger that is specific - ensuring that we both get the quantity we need to meet our energy needs, but also the right types of foods to meet our nutritional needs.

The process by which we obtain the substances that the body needs begins with the consumption of food and is achieved when the nutrients provided by food are absorbed. This is the process of digestion, the process of ingesting, breaking down food, and absorbing the nutrients.

Ultimately, food is delivered to the body in three forms:

- Lipids (fats),
- amino acids (proteins), and
- glucose (carbohydrates).

These fuels enter the body by way of the digestive tract, but the digestive tract is often empty. Because of this, and the importance of having fuel available, we have mechanisms that allow us to store energy to support us. This stored energy is primarily fats, but also glycogen and proteins. Before discussing how we store energy (and access stored energy), we need to understand the process of digestion.

12.4.2 **The Human Digestive System**

The process of digestion begins in the mouth with the intake of food (Figure **12.4.1**). The teeth play an important role in masticating (chewing) or physically breaking food into smaller particles. The enzymes present in saliva also begin to chemically break down food. The food is then swallowed and enters the esophagus - a long tube that connects the mouth to the stomach. Using peristalsis, or wave-like smooth-muscle contractions, the muscles of the esophagus push the food toward the stomach. The stomach contents are extremely acidic. This acidity kills microorganisms, breaks down food tissues, and activates digestive enzymes. Further breakdown of food takes place in the small intestine where bile produced by the liver, and enzymes produced by the small intestine and the pancreas, continue the process of digestion. The smaller molecules are absorbed into the blood stream through the epithelial cells lining the walls of the small intestine. The waste material travels on to the large intestine where water is absorbed and the drier waste material is compacted into feces; it is stored until it is excreted through the anus.

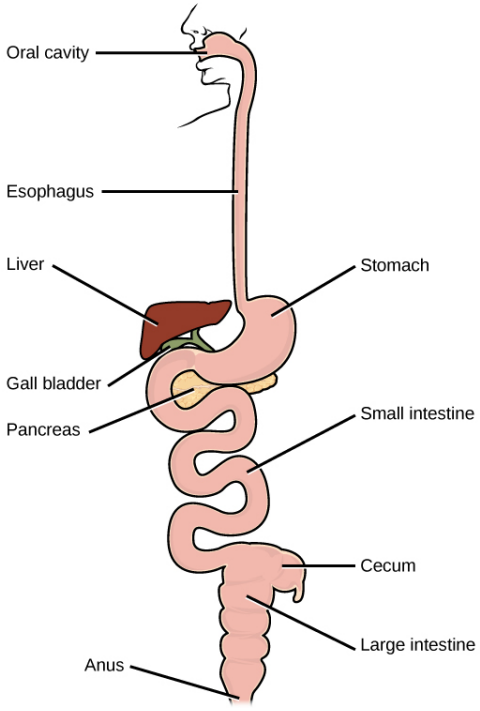


Figure **12.4.1**: The components of the human digestive system are shown.

12.4.2.1 Oral Cavity

Both physical and chemical digestion begin in the mouth or oral cavity, which is the point of entry of food into the digestive system. The food is broken into smaller particles by mastication, the chewing action of the teeth. Almost all mammals have teeth and can chew their food to begin the process of physically breaking it down into smaller particles. (Mammals without teeth have very limited diets - ants or plankton - and include pangolins, anteaters and some species of whales.)

The chemical process of digestion begins during chewing as food mixes with saliva, produced by the salivary glands. Saliva contains mucus that moistens food and buffers the pH of the food. Saliva also contains enzymes that begin the process of breaking down some foods. The chewing and wetting action provided by the teeth and saliva prepare the food into a mass called the bolus for swallowing. The tongue helps in swallowing - moving the bolus from the mouth into the pharynx. The pharynx opens to two passageways: the esophagus and the trachea. The esophagus leads to the stomach and the trachea leads to the lungs. The epiglottis is a flap of tissue that covers the tracheal opening during swallowing to prevent food from entering the lungs.

12.4.2.2 Esophagus

The esophagus is a tubular organ that connects the mouth to the stomach. The chewed and softened food passes through the esophagus after being swallowed. The smooth muscles of the esophagus undergo peristalsis that pushes the food toward the stomach. The peristaltic wave moves food from the mouth to the stomach, and reverse movement is not possible, except in the case of vomiting (initiated by the gag reflex). The peristaltic movement of the esophagus is an involuntary reflex; it takes place in response to the act of swallowing.

Ring-like muscles called sphincters form valves in the digestive system. The gastro-esophageal sphincter (or cardiac sphincter) is located at the stomach end of the esophagus. In response to swallowing and the pressure exerted by the bolus of food, this sphincter opens, and the bolus enters the stomach. When there is no swallowing action, this sphincter is shut and prevents the contents of the stomach from traveling up the esophagus. Acid reflux or “heartburn” occurs when the acidic digestive juices escape into the esophagus.

12.4.2.3 Stomach

A large part of protein digestion occurs in the stomach. The stomach is a saclike organ that secretes gastric digestive juices.

Protein digestion is carried out by an enzyme called pepsin in the stomach chamber. The highly acidic environment kills many microorganisms in the food and, combined with the action of the enzyme pepsin, results in the catabolism (break down) of protein in the food. Chemical digestion is facilitated by the churning action of the stomach caused by contraction and relaxation of smooth muscles. The partially digested food and gastric juice mixture is called chyme. Gastric emptying occurs within two to six hours after a meal. Only a small amount of chyme is released into the small intestine at a time. The movement of chyme from the stomach into the small intestine is regulated by hormones, stomach distension, and muscular reflexes that influence the pyloric sphincter.

The stomach lining is unaffected by pepsin and the acidity because pepsin is released in an inactive form and the stomach has a thick mucus lining that protects the underlying tissue.

12.4.2.4 Small Intestine

Chyme moves from the stomach to the small intestine. The small intestine is the organ where the digestion of protein, fats, and carbohydrates is completed. The small intestine is a long tube-like organ with a highly folded surface containing finger-like projections called the villi. The top surface of each villus has many microscopic projections called microvilli. The epithelial cells of these structures absorb nutrients from the digested food and release them to the bloodstream on the other side. The villi and microvilli, with their many folds, increase the surface area of the small intestine and increase absorption efficiency of the nutrients.

The human small intestine is divided into three parts: the duodenum, the jejunum, and the ileum. The duodenum is separated from the stomach by the pyloric sphincter. The chyme is mixed with pancreatic juices, an alkaline solution rich in bicarbonate that neutralizes the acidity of chyme from the stomach. Pancreatic juices contain several digestive enzymes that break down starches, disaccharides, proteins, and fats. Bile is produced in the liver and stored and concentrated in the gallbladder; it enters the duodenum through the bile duct. Bile contains bile salts, which make lipids accessible to the water-soluble enzymes. The monosaccharides, amino acids, bile salts, vitamins, and other nutrients are absorbed by the cells of the intestinal lining.

The undigested components of the food are sent to the colon from the ileum via peristaltic movements. The ileum ends and the large intestine begins at the ileocecal valve. The vermiform, “worm-like,” appendix is located at the ileocecal valve.

12.4.2.5 Large Intestine

The large intestine reabsorbs the water from indigestible food material and processes the waste material (Figure 12.4.2). Compared to the small intestine, the human large intestine is much smaller in length but larger in diameter. It has three parts: the cecum, the colon, and the rectum. The cecum joins the ileum to the colon and is the receiving pouch for the waste matter. The colon is home to many bacteria or “intestinal flora” that aid in the digestive processes. The colon has four regions, the ascending colon, the transverse colon, the descending colon and the sigmoid colon. The main functions of the colon are to extract the water and mineral salts from undigested food components, and to store waste material.

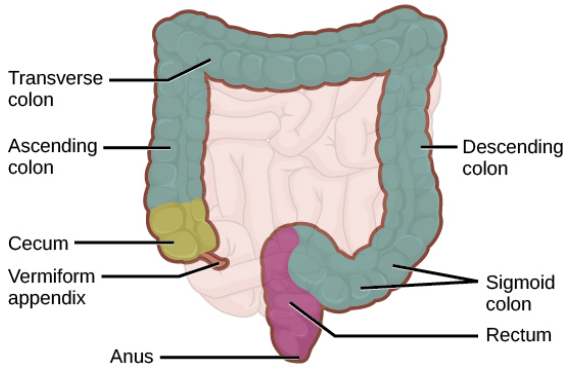


Figure 12.4.2: The large intestine reabsorbs water from undigested food and stores waste until it is eliminated. (credit: modification of work by Mariana Ruiz Villareal)

The rectum (Figure 12.4.2) stores feces until defecation. The feces are propelled using peristaltic movements during elimination. The anus is an opening at the far-end of the digestive tract and is the exit point for the waste material. Two sphincters regulate the exit of feces, the inner sphincter is involuntary and the outer sphincter is voluntary.

12.4.2.6 Accessory Organs

The organs discussed above are the organs of the digestive tract through which food passes. Accessory organs add secretions and enzymes that break down food into nutrients. Accessory organs include the salivary glands, the liver, the pancreas, and the gall bladder (Figure **12.4.3**). The secretions of the liver, pancreas, and gallbladder are regulated by hormones in response to food consumption.

The liver is the largest internal organ in humans and it plays an important role in digestion of fats and detoxifying blood. The liver produces bile, a digestive juice that is required for the breakdown of fats in the duodenum. The liver also processes the absorbed vitamins and fatty acids and synthesizes many plasma proteins. The gallbladder is a small organ that aids the liver by storing bile and concentrating bile salts.

The pancreas secretes bicarbonate that neutralizes the acidic chyme and a variety of enzymes for the digestion of protein and carbohydrates.

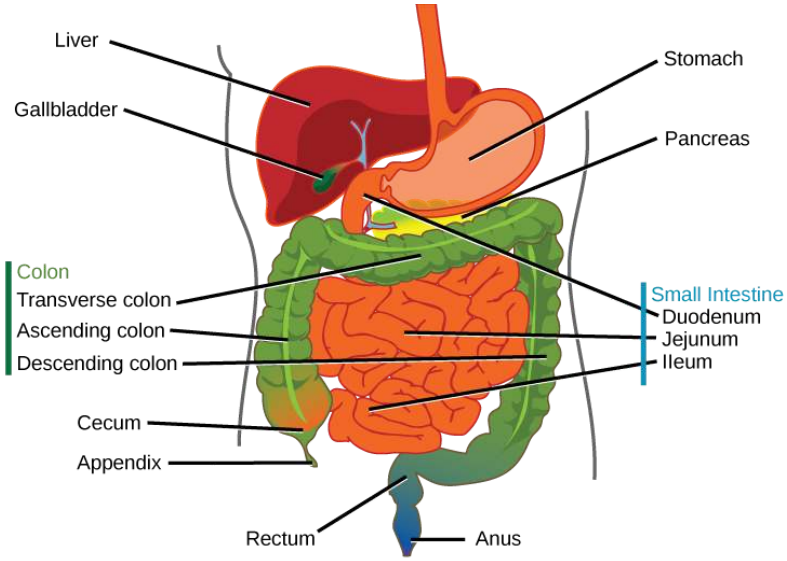


Figure **12.4.3**: The stomach has an extremely acidic environment where most of the protein gets digested. (credit: modification of work by Mariana Ruiz Villareal)

12.4.3 Nutrition

The human diet should be well-balanced to provide nutrients required for bodily function and the minerals and vitamins required for maintaining structure and regulation necessary for good health and reproductive capability. The organic molecules required for building cellular material and tissues must come from food.

During digestion, digestible carbohydrates are ultimately broken down into glucose and used to provide energy to the cells of the body and the brain. Complex carbohydrates can be broken down into glucose through biochemical modification; however, humans do not produce the enzyme necessary to digest fiber. The intestinal bacteria in the human gut are able to extract some nutrition from these plant fibers. These plant fibers are known as dietary fiber and are an important component of the diet. The excess sugars in the body are converted into glycogen and stored for later use in the liver and muscle tissue. Glycogen stores are used to fuel prolonged exertions, such as long-distance running, and to provide energy during food shortage. Fats are stored under the skin of mammals for insulation and energy reserves and provide cushioning and protection for many organs.

Proteins in food are broken down during digestion and the resulting amino acids are absorbed. All of the proteins in the body must be formed from these amino acid constituents; no proteins are obtained directly from food.

Fats add flavor to food and promote a sense of satiety or fullness. Fatty foods are also significant sources of energy, and fatty acids are required for the construction of lipid membranes. Fats are also required in the diet to aid the absorption of fat-soluble vitamins and the production of fat-soluble hormones.

While the human body can synthesize many of the molecules required for function from precursors, there are some nutrients that must be obtained from food. These nutrients are termed essential nutrients, meaning they must be eaten, because the body cannot produce them.

12.4.4 Metabolism - An Overview

Metabolism is the process of making energy available for use. If your metabolism is high, you are breaking stored fuel down at a high rate - you are burning more calories. Note that the different forms of stored energy are found in different places in the body and the flow of energy into and out of storage is under the control of hormones. What hormones are involved directly in this process? Glycogen is the form in which carbohydrate is stored in cells of the liver and muscles. Ingested carbohydrates, in the form of glucose, are converted into glycogen by insulin and then stored. Insulin both promotes the storage of glucose as glycogen and allows body cells to use glucose. In other words, the cells of the body do not have access to glucose if insulin is not present. While insulin promotes the storage and use of glucose, glucagon (a hormone created in the pancreas) promotes the conversion of stored fuels to a form that can be readily used. In other words, if glucose is gone, glucagon makes more fuel available.

12.4.4.1 Regulation of Blood Glucose Levels by Insulin and Glucagon

Blood glucose levels vary widely over the course of a day as periods of food consumption alternate with periods of fasting. Insulin and glucagon are the two hormones primarily responsible for maintaining homeostasis of blood glucose levels. Additional regulation is mediated by the thyroid hormones.

Cells of the body require nutrients in order to function, and these nutrients are obtained through feeding. Excess intake is converted to stores and removed when needed. Hormones moderate our energy stores. Insulin is produced by the beta cells of the pancreas, which are stimulated to release insulin as blood glucose levels rise (for example, after a meal is consumed). Insulin lowers blood glucose levels by enhancing the rate of glucose uptake and use by cells. Insulin also stimulates the liver to convert glucose to glycogen, which is then stored by cells for later use. Some cells, including those in the kidneys and brain, can access glucose without the use of insulin. Insulin also stimulates the conversion of glucose to fat in adipocytes and the synthesis of proteins. These actions mediated by insulin cause blood glucose concentrations to fall, called a hypoglycemic “low sugar” effect, which inhibits further insulin release from beta cells through a negative feedback loop.

Impaired insulin function can lead to a condition called diabetes mellitus, the main symptoms of which are illustrated in Figure 12.4.4. This can be caused by low levels of insulin production by the beta cells of the pancreas, or by reduced sensitivity of tissue cells to insulin. This prevents glucose from being absorbed by cells, causing high levels of blood glucose, or hyperglycemia (high sugar). High blood glucose levels make it difficult for the kidneys to recover all the glucose from the urine, resulting in glucose being lost in urine. High glucose levels also result in less water being reabsorbed by the kidneys, causing high amounts of urine to be produced; this may result in dehydration. Over time, high blood glucose levels can cause nerve damage to the eyes and peripheral body tissues, as well as damage to the kidneys and cardiovascular system. Oversecretion of insulin can cause hypoglycemia, low blood glucose levels. This causes insufficient glucose availability to cells, often leading to muscle weakness, and can sometimes cause unconsciousness or death if left untreated.

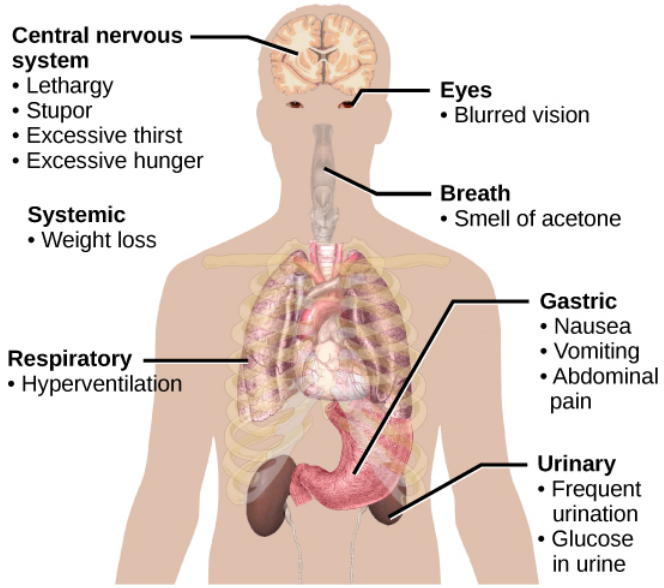


Figure 12.4.4: The main symptoms of diabetes are shown. (credit: modification of work by Mikael Häggström)

When blood glucose levels decline below normal levels, for example between meals or when glucose is utilized rapidly during exercise, the hormone glucagon is released from the alpha cells of the pancreas. Glucagon raises blood glucose levels, causing what is called a hyperglycemic effect, by stimulating the breakdown of glycogen to glucose in skeletal muscle cells and liver cells in a process called glycogenolysis. Glucose can then be utilized as energy by muscle cells and released into circulation by the liver cells. Glucagon also stimulates absorption of amino acids from the blood by the liver, which then converts them to glucose. This process of glucose synthesis is called gluconeogenesis. Glucagon also stimulates adipose (fat) cells to release fatty acids into the blood. These actions mediated by glucagon result in an increase in blood glucose levels to normal homeostatic levels. Rising blood glucose levels inhibit further glucagon release by the pancreas via a negative feedback mechanism. In this way, insulin and glucagon work together to maintain homeostatic glucose levels, as shown in Figure 12.4.5.

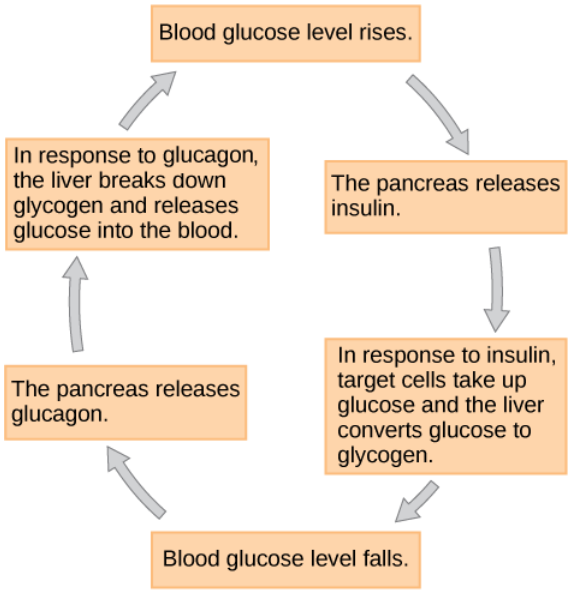


Figure 12.4.5: Insulin and glucagon regulate blood glucose levels. (credit: modification of OpenStax image)

12.4.5 Long-Term Energy Stores

"Fat" is composed of triglycerides, glycerol and fatty acids. Fat, or adipose, tissue is found beneath the skin and in various areas around the abdominal cavity. The cells in this tissue are able to absorb nutrients from the blood and are quite versatile - changing dramatically in size as the levels of stored triglycerides change. Once the short-term carbohydrate reserve is depleted, triglycerides

begin to be converted to a form that cells can use and are released. Free fatty acids can be used for energy by all cells, except those of the central nervous system. Where does the brain get energy in this case? While the cells in the body are using free fatty acids, the liver is taking up the glycerol and converting it to glucose. As discussed earlier, when glucose is gone, we get the release and action of glucagon - a hormone that converts stored fuels to those that can be readily used, like glucose.

What happens when glucagon is high and insulin is low, when stored fuels are being made available? The available free fatty acids (previously stored in adipose tissue) are converted to ketones which can be used by muscles. The breakdown of fat for energy results in a high level of ketones, a phenomenon that can also occur when one is starving. Normally, ketones are excreted in the urine. When the body is not able to process excess ketones, ketoacidosis occurs. Ketoacidosis literally leads to an increase in the acidity of the body which creates a whole host of problems - even death. Dehydration occurs as the body tries to adjust. What you have is also the movement of water out of cells - which will make more sense once we cover thirst.

12.4.6 The Three Phases of Digestion

When we look at the process of digestion, we can recognize three different phases. Each has its own unique hormonal profile.

- Cephalic - During the cephalic or "head" phase, the body is preparing to eat. The expectation of food, the anticipation of food, leads to the release of insulin. If the smell or sight of food leads to a hunger pang, this just might be a consequence of insulin causing any available glucose to be used or stored. This phase ends when those ingested nutrients are absorbed into the blood stream. So the cephalic phase begins when the body prepares for food ingestion and ends when that ingested food impacts fuel levels in the bloodstream.
- Absorptive - During the absorptive phase, insulin is actively promoting the use - and storage - of ingested fuels. This phase ends when the unstored energy from the meal has been used.
- Fasting Phase - When the fuel from the meal is gone, glucagon levels rise and energy stores are converted into usable fuels.

When there is no fuel available in the blood, when glucose levels drop, energy must be mobilized (removed) from energy stores. What part of the nervous system makes energy available? The sympathetic nervous system, the fight or flight component of the autonomic branch of the peripheral nervous system, causes the secretion of glucagon when there is a lack of fuel in the blood.

12.4.7 Summary

The process of digestion begins in the mouth where food begins to be broken down both chemically and physically into the lipids, amino acids, and glucose that are used by the body. Digestion begins in the mouth and ends when fuels are absorbed into the blood stream. Digestion consists of three phases: cephalic, absorptive, and fasting.

Metabolism, a process controlled by the hormones insulin and glucagon, makes energy available for use. Insulin is produced by the pancreas in response to rising blood glucose levels and allows cells to use the available blood glucose and store excess glucose as glycogen for later use. Diabetes mellitus is caused by reduced insulin activity and causes high blood glucose levels, or hyperglycemia. Glucagon is released by the pancreas in response to low blood glucose levels and stimulates the breakdown of glycogen into glucose, which can be used by the body.

12.4.8 Attributions (listed in order of appearance)

"The Human Digestive System" and "Nutrition" adapted from Samantha Fowler, S., Roush, R. & Wise, J, (2013). *16.2 Digestive System - Concepts of Biology.* OpenStax. Retrieved April 24, 2022, from https://openstax.org/books/concepts-biology/pages/16-2-digestive-system (CC BY)

"Hunger, Satiety, and the Brain" adapted from Bhatia, S. & Loewenstein, G. (2022). Drive states. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology.* Champaign, IL: DEF publishers. Retrieved from http://noba.to/pjwkbt5h (CC BY-NC-SA)

"Regulation of Blood Glucose Levels by Insulin and Glucagon" adapted from Rye, C., Wise, R., Jurukovski, V., DeSaix, J., Choi, J., & Avissar , Y. (2016, October 21). *37.3 Regulation of Body Processes - Biology*. OpenStax. Retrieved February 6, 2022, from https://openstax.org/books/biology/p...body-processes (CC BY)

This page titled 12.4: Digestion and the Hormonal Control of Metabolism is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

- **37.3: Regulation of Body Processes** by OpenStax is licensed CC BY 4.0.

12.5: Hunger - Theories, Detectors, and the Hypothalamus

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Explain the concept of a set-point theory and how the detection of blood glucose levels and body fat explains some elements of feeding behavior.
2. Discuss the evidence for and against the existence of hypothalamic hunger and satiety centers.
3. Explain the effects of insulin, CCK, NPY, leptin, and serotonin on eating behavior.

12.5.1 Overview

What determines when we are hungry - how is the need for food detected? How does the detection of a need lead to eating? And what determines when we stop eating? In this section we will explore potential hunger signals and various theories that explain some elements of eating behavior. As previously noted, human eating behavior is complicated by a wide array of environmental factors. As a consequence, much of our understanding of the biological factors involved in eating comes from animal studies.

Eating behavior is controlled by two separate, yet related, systems - one that signals a need to eat and another that signals when that need has been met. The key to these systems may be two hypothalamic nuclei. In addition, various chemical factors, including cholecystokinin (CCK), Neuropeptide Y (NPY), and leptin appear to play a role in eating behavior.

12.5.2 Set-Point Theories

Before we look at the mechanisms that produce hunger, we'll first consider some of the major theories that strive to explain hunger and eating. There are many theories that we could consider and, perhaps, someday a single theory will emerge that is superior to all others, but at the present time we don't have a single theory that accounts for how all the different factors that impact feeding determine what we do. Most significantly, while human eating behavior tends to align with biological expectations over the long-term, the theories that focus on this fail to explain the increase in obesity that we see today (Speakman et al., 2011). In other words, there is a correlation between how much we eat (the calories we ingest) and how much we use (the calories we burn) when we consider our behavior over an extended period of time - but not when we look at individual meal-taking and daily eating behavior.

You've probably been exposed to the notion of a body weight "set-point", the idea that you have an "ideal" weight and it takes a lot of work to move away from that set-point, that optimal value that our body strives to achieve. Perhaps we have some ideal level of energy that is necessary and that we eat to restore this value. In other words, perhaps we have a set-point for available energy. Is hunger an attempt to restore energy reserves and is hunger merely signaled when energy reserves are depleted? This would predict a consistency in energy use and food intake that we simply do not see (Hall et al., 2021). Perhaps the maintenance of blood glucose (readily available fuel in the blood) could be what controls feeding for the short-term and the maintenance of body fat (stored energy) could provide a long-term explanation of eating behavior. These are both variations on the set-point theory. While each of these theories has its problems, they are consistent with some findings. Glucostatic theories propose that we work to maintain some ideal level of blood glucose. In contrast, lipostatic theories propose that we monitor the level of body fat that we have stored and then alter our eating to compensate when fat levels are not at the desired level. Having identified what is being regulated (our system variable, blood glucose or body fat) and presuming some set point, are there detectors to monitor our regulated variables?

12.5.2.1 Hunger Signals

It turns out that the brain responds to two types of hunger signals:

- Short-term. Receptors in the liver and the brain are involved in responding to short-term signals. In other words, they respond to levels of available fuel in the blood.
- Long-term. Long-term hunger signals are a consequence of the monitoring of fat stores. When fat cells are full, they produce a hormone called leptin. Leptin then inhibits the brain mechanisms that control eating - leptin makes the brain less responsive to

short-term signals. Leptin effectively turns hunger signals off (negative feedback).

Consistent with glucostatic theory, the usual trigger for hunger is a drop in blood glucose (hypoglycemia). We can produce hunger, short-term hunger, by one of two mechanisms:

- Decrease blood glucose with the administration of insulin. This causes available glucose to be stored and used by body cells. You should recall that the increase in insulin during the cephalic phase of digestion can increase your pre-meal hunger.
- Inject 2-deoyglucose (2-DG). 2-DG is structurally similar to glucose, but cells are unable to metabolize it. This produces glucoprivation - although glucose is available, cells are unable to use it. Both body and brain cells are deprived of glucose.

Both manipulations deprive body cells of glucose and both will produce eating, consistent with a glucostatic hypothesis (Thompson & Campbell, 1977).

Lipostatic theory proposes that when stored fat drops below an optimal value, hormones are secreted that increase food intake and promote weight gain (Gale et al., 2004). It also proposes that when energy stores exceed that optimal value, adipose (fat) cells release leptin to reduce eating.

12.5.2.2 From Set-Point to Settling Point

If you accept that a set-point theory does not explain our day-to-day eating behavior, what about the notion of a set-point as an explanation for why we can't seem to lose those last 5 pounds? We certainly tend to maintain a relatively constant body weight - why is that? And should we listen to our bodies? Should you eat whenever your body tells you to?

While these observations may suggest that we ultimately seek to defend a particular level of body fat, another notion is that we maintain a weight at which all the factors that influence body weight reach an equilibrium - a so-called "settling point".

12.5.3 Determining What to Eat

Let's look more closely at what determines the specifics of eating behavior. What we find is that, despite the findings with non-human subjects, human eating behavior has little to do with energy stores. What we do see is natural preferences for food that would facilitate survival in times of energy need - what types of food taste good? Sweet and fatty foods are energy dense and our cravings for them make sense from an evolutionary perspective - there was a time when we had to literally work for our food (hunting and gathering) as opposed to opening the refrigerator or placing an order for delivery. At the same time, bitter and sour tastes are typically associated with toxins. So, we have species-typical taste preferences and aversions (dislikes).

In addition to our inherited preferences, we also have those that are learned. We learn to eat what those around us eat - that which is familiar to us is going to be more appealing. Even if you like to try new things, I'm sure that there are limits to what you will eat.

How does an animal in the wild know what to eat in order to get the nutrients it needs? Researchers have looked at what happens when a diet lacks a nutrient and found that animals will respond - they will seek out variety when not getting all the needed nutrients and develop aversions to (dislike of) foods lacking in some nutrient. Most nutrients can't be tasted - especially with our diet today. But we do see cravings and desires at certain times - perhaps because those foods meet a need. Perhaps pickles have some benefit during pregnancy. And maybe ice cream fills a need created by a broken heart. Theoretically, we should have some ability to select the right foods, as we see in animals. But perhaps the learned elements of our eating (eating at meal times, clean your plate) have disrupted the body's ability to direct our eating choices. Further complicating ensuring we consume a balanced diet are personal decisions to restrict the types of foods eaten.

Figure **12.5.1**: Fish eggs and raw fish - an exciting meal to those who appreciate such things. (Photo by Jason Leung on Unsplash)

12.5.4 Detecting Satiety

What determines how much we eat? How do we get satiety, what makes us feel "full"? Food in the body can induce satiety, but so can simply causing the stomach to distend - as was initially demonstrated in the early 1900's by researchers who observed the impact of inflating and deflating a swallowed ballon on hunger pangs (Cannon and Washburn, 1912). The fact is that the size of the stomach impacts how much one eats. Satiety created by food ingestion does seem to be dependent on nutritive density - to some degree food ingestion will be altered when the caloric content is changed. If you increase or decrease the calories per unit volume of food, animals will adjust their diet accordingly. Yet other studies have demonstrated with the use of sham feeding (where an animal feeds, but the food is ultimately not digested or absorbed) that the changes in food ingestion are not merely due to changes in the immediate caloric impact of the food.

We all know that we do respond to how food tastes - and we get tired of eating even those things that we do like - as studies of sensory-specific satiety have demonstrated. Rats (and humans) will eat more when presented with a varied diet of tasty food - a cafeteria diet. Here you see the incentive value of food overriding the normal mechanisms that control food intake.

12.5.5 **Physiological Mechanisms of Hunger and Eating**

There are a number of physiological mechanisms that serve as the basis for hunger. When our stomachs are empty, they contract. Typically, a person then experiences hunger pangs. Chemical messages travel to the brain, and serve as a signal to initiate feeding behaviour. When our blood glucose levels drop, the pancreas and liver generate a number of chemical signals that induce hunger (Konturek et al., 2003; Novin, Robinson, Culbreth, & Tordoff, 1985) and thus initiate feeding behaviour.

For most people, once they have eaten, they feel **satiation**, or fullness and satisfaction, and their eating behavior stops. Like the initiation of eating, satiation is also regulated by several physiological mechanisms. As blood glucose levels increase, the pancreas and liver send signals to shut off hunger and eating (Drazen & Woods, 2003; Druce, Small, & Bloom, 2004; Greary, 1990). The food's passage through the gastrointestinal tract also provides important satiety signals to the brain (Woods, 2004), and fat cells release **leptin**, a satiety hormone.

Controlling the size of the stomach or the fullness of it can be used to help regulate eating. Bariatric surgery involves modifying the digestive system to achieve weight-loss. Approaches to such surgeries are varied, but decreasing the amount of food the stomach can hold is one such approach. Although recognized as an effective long-term approach for many, surgery may not be an option for many with a need for weight loss. Today, intragastric balloons are an alternative to surgical interventions.

The various hunger and satiety signals that are involved in the regulation of eating are integrated in the brain. Research suggests that several areas of the hypothalamus and hindbrain are especially important sites where this integration occurs (Ahima & Antwi, 2008; Woods & D'Alessio, 2008). Ultimately, activity in the brain determines whether or not we engage in feeding behavior.

12.5.6 The Role of the Hypothalamus

The hypothalamus plays a very important role in eating behavior. It is responsible for synthesizing and secreting various hormones. The lateral hypothalamus (LH) is concerned largely with hunger and, in fact, lesions (i.e., damage) of the LH can eliminate the desire for eating entirely - to the point that animals starve themselves to death unless kept alive by force feeding (Anand & Brobeck, 1951). Additionally, artificially stimulating the LH, using electrical currents, can generate eating behavior if food is available (Andersson, 1951).

Activation of the LH can not only increase the desirability of food but can also reduce the desirability of nonfood-related items. For example, Brendl, Markman, and Messner (2003) found that participants who were given a handful of popcorn to trigger hunger not only had higher ratings of food products, but also had lower ratings of nonfood products—compared with participants whose appetites were not similarly primed. That is, because eating had become more important, other non-food products lost some of their value.

While the feeling of hunger gets you to start eating, the feeling of satiation gets you to stop. Hunger and satiation are two distinct processes, controlled by different circuits in the brain and triggered by different cues. Distinct from the LH, which plays an important role in hunger, the ventromedial hypothalamus (VMH) plays an important role in satiety. Though lesions of the VMH can cause an animal to overeat to the point of obesity, the relationship between the LH and the VMH is quite complicated. Rats with VMH lesions can also be quite finicky about their food (Teitelbaum, 1955).

Is it that simple? Is the LH the key to hunger and the VMH the source of satiation (satiety)?

Evidence supporting the LH as a hunger center and the VMH as a satiety center:

- Lesioning (destroying) the lateral hypothalamus leads to anorexia (loss of appetite) and aphagia (absence of eating). At the same time, stimulating this region produces both eating and drinking. If you need a way to remember which is which - if you lesion the LH you get a lean hamster. If you lesion the VMH you get a very meaty hamster.
- Lesioning the ventromedial hypothalamus leads to hyperphagia (excessive eating) and weight gain. And stimulation of the VMH suppresses eating.

Evidence against the LH as a hunger center and the VMH as a satiety center:

- Lesions of the LH interfere with all motivated behaviors, not just feeding. In other words, perhaps the LH is a motivation center as opposed to a hunger center.
- Lesions of the VMH suppress the activity of sympathetic nervous system and increase the activity of the parasympathetic nervous system. In other words, the body is in rest and restore mode. Rats with VMH lesions become picky eaters, preferring carbohydrates. Due to the increased PNS activity, there is no fasting phase of metabolism; they are never living off of their reserves - rats continue to eat as they are not using their energy stores.

Other brain areas, besides the LH and VMH, also play important roles in eating behavior. The sensory cortices (visual, olfactory, and taste), for example, are important in identifying food items. These areas provide informational value, however, not hedonic evaluations. That is, these areas help tell an organism what is good or safe to eat, but they don't provide the pleasure (or hedonic) sensations that *actually* eating the food produces. While many sensory functions are roughly stable across different psychological states, other functions, such as the detection of food-related stimuli, are enhanced when the organism is in a hungry drive state.

After identifying a food item, the brain also needs to determine its reward value, which affects the organism's motivation to consume the food. The reward value ascribed to a particular item is, not surprisingly, sensitive to the level of hunger experienced by the organism. The hungrier you are, the greater the reward value of the food. Neurons in the areas where reward values are processed, such as the orbitofrontal cortex, fire more rapidly at the sight or taste of food when the organism is hungry relative to if it is satiated.

12.5.7 Hunger, Satiety, and the Body

What makes you feel full? Is it when you belly actually feels full? Can a full stomach signal to the brain that you are satiated? Yes. How does the stomach tell the brain that nutrients are there? Energy deficits are not corrected - nutrients are not absorbed at the stomach. But a stomach full of food will still put a stop to feeding, even if there are no nerves connecting it to the brain. And a full stomach - with no nutrients present - can also end feeding. What type of signals are getting to the brain when there are no nerves involved? And how can just filling the stomach lead to satiation?

The stomach and other parts of the gastrointestinal tract produce peptides that can function as hormones, traveling in the blood stream to their site of action. One peptide that has been studied quite extensively is cholecystokinin (CCK). In the absence of a neural connection, a chemical in the blood would be an ideal mechanism for the stomach to use to communicate with the brain. It turns out that CCK does cause food-deprived rats to eat smaller meals - but is that because they are feeling satiety? It turns out that CCK can make one feel ill - which would be likely to decrease eating.

While a whole host of peptides are thought to be involved in satiety, two have been found to play a role in hunger. Neuropeptide Y (NPY) and galanin both cause eating when injected into the paraventricular nucleus of the hypothalamus. NPY creates a hunger for carbohydrates and galanin for fats. Both decrease metabolism and increase fat production, suggesting they play a role in combating starvation. How does this work to regulate eating outside of a lab? Food deprivation results in an increase in NPY levels and eating decreases NPY levels (). If NPY receptors in the hypothalamus are blocked, eating caused by NPY or deprivation is not seen.

What role do neurotransmitters play with respect to eating? It turns out that serotonin inhibits food intake, especially carbohydrate ingestion. Why is this so? Serotonin inhibits NPY.

Leptin also decreases food intake by acting on NPY levels. NPY-secreting neurons in the arcuate nucleus have receptors for leptin.

12.5.8 Summary

Eating behavior is complicated, controlled by two systems, one that controls hunger and one that signals satiety. In other words, we have a system that turns hunger on and another that determines that the need has been met. The lateral hypothalamus plays a role in hunger and the ventromedial hypothalamus appears to play a central role in satiation. Further complicating the control of hunger is the presence of two ways of signaling an energy need, a system that responds to the available levels of fuel and responding to short-term need and one that monitors energy stores. In the short-term, levels of available glucose are monitored, while stored fats are the focus for the long-term system.

12.5.9 References

Cannon, W. B., & Washburn, A. L. (1912). An Explanation of Hunger. *American Journal of Physiology-Legacy Content*, *29*(5), 441–454. https://doi.org/10.1152/ajplegacy.1912.29.5.441

Crossan K, Sheer AJ. Intragastric Balloon. [Updated 2022 Feb 3]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/N...?log$=activity

Gale, S. M., Castracane, V. D., & Mantzoros, C. S. (2004). Energy homeostasis, obesity and eating disorders: Recent advances in endocrinology. *The Journal of Nutrition*, *134*(2), 295. http://dx.doi.org/10.14341/2071-8713-5181

Hall, K. D., Heymsfield, S. B., Kemnitz, J. W., Klein, S., Schoeller, D. A., & Speakman, J. R. (2012). Energy balance and its components: implications for body weight regulation. *The American journal of clinical nutrition*, *95*(4), 989–994. https://doi.org/10.3945/ajcn.112.036350

Speakman, J. R., Levitsky, D. ., Allison, D.B., Bray MS, de Castro JM, Clegg DJ, Clapham JC, Dulloo AG, Gruer L, Haw S, Hebebrand J, Hetherington MM, Higgs S, Jebb SA, Loos RJ, Luckman S, Luke A, Mohammed-Ali V, O'Rahilly S, Pereira M, Perusse L, Robinson TN, Rolls B, Symonds ME, Westerterp-Plantenga MS. Set points, settling points and some alternative models: theoretical options to understand how genes and environments combine to regulate body adiposity. Dis Model Mech. 2011 Nov;4(6):733-45. doi: 10.1242/dmm.008698. PMID: 22065844; PMCID: PMC3209643.

12.5.10 Attributions (listed in order of appearance)

"Overview", "Set-Point Theories", "Determining What to Eat", "Detecting Satiety", "Hunger, Satiety, and the Body", and "Summary" (2022) (CC BY)

"The Human Digestive System" and "Nutrition" adapted from Samantha Fowler, S., Roush, R. & Wise, J, (2013). 16.2 Digestive System - Concepts of Biology. OpenStax. Retrieved April 24, 2022, from https://openstax.org/books/concepts-biology/pages/16-2-digestive-system (CC BY)

Physiological Mechanisms of Hunger and Eating adapted from Hunger and Eating by Edited by Leanne Stevens is licensed under a Creative Commons Attribution 4.0 International License, except where otherwise noted.

"The Role of the Hypothalamus" adapted from Bhatia, S. & Loewenstein, G. (2022). Drive states. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology.* Champaign, IL: DEF publishers. Retrieved from http://noba.to/pjwkbt5h (CC BY-NC-SA)

12.5.11

This page titled 12.5: Hunger - Theories, Detectors, and the Hypothalamus is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

- **37.3: Regulation of Body Processes** by OpenStax is licensed CC BY 4.0.

12.6: Ingestive Behavior in Context

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Type the Learning Objectives here
2.

12.6.1 Overview

Eating and drinking are complicated behaviors that are intended to meet biological needs. As already discussed, complex biological mechanisms serve to initiate and terminate these behaviors. As with many behaviors, humans make choices and they may make choices that are not consistent with their biological needs. We end this chapter with a discussion of diabetes and obesity. Both are conditions that were noted previously in this chapter. Here we consider them - and what we know about them - in greater detail.

12.6.2 Diabetes

Diabetes mellitus (diabetes) is a chronic disorder that can alter carbohydrate, protein, and fat metabolism. It is caused by the absence of insulin secretion due to either the progressive or marked inability of the islet cells of the pancreas to produce insulin, or due to defects in insulin uptake in the peripheral tissue. Diabetes is broadly classified under two categories, which include type 1 and type 2 diabetes (Mathieu & Badenhoop, 2005).

Type 1 diabetes occurs most commonly in children, but it can sometimes also appear in adult age groups, particularly in those in their late thirties and early forties. Patients with type 1 diabetes are generally not obese and frequently present with an emergency status known as diabetes ketoacidosis, a condition that develops when the body begins using fat for energy (American Diabetes Association, 2007).

The development of type 1 diabetes can be explained by damage to the pancreatic cells due to environmental or infectious agents. In individuals who are susceptible to genetic alterations, the immune system is triggered to produce an immune response against the altered islet cells that product insulin, or against molecules in those cells (Hutton & Davidson, 2010). Approximately 80% of patients with type 1 diabetes show circulating islet cell antibodies, and most of these patients have anti-insulin antibodies before receiving insulin therapy (van Belle, Coppieters, & von Herrath, 2011).

The major factor in the pathophysiology of type 1 diabetes is considered to be autoimmunity (Mathieu & Badenhoop, 2005).There is a strong relationship between type 1 diabetes and other autoimmune diseases such as Graves' disease, Hashimoto's thyroiditis, and Addison's disease. When these diseases are present, the prevalence rates of type 1 diabetes increase (Philippe, 2011).

Type 2 diabetes has a different pathophysiology and origin as compared to type 1 diabetes. The existence of many new factors – for example, the increased prevalence of obesity among all age groups and both sexes, physical inactivity, poor diet, and urbanization – means that the number of patients diagnosed with type 2 diabetes is rising (Ershow, 2009) . This finding is significant because it will allow health planners to make rational plans and reallocate health resources accordingly (Wild et al., 2004).

Type 2 diabetes is described as a combination of low amounts of insulin production from pancreatic islet cells and peripheral insulin resistance (Kasuga, 2006). Insulin resistance leads to elevated fatty acids in the plasma, causing decreased glucose transport into the muscle cells, as well as increased fat breakdown, subsequently leading to elevated liver glucose production. Insulin resistance and pancreatic cell dysfunction must occur simultaneously for type 2 diabetes to develop. Anyone who is overweight and/or obese has some kind of insulin resistance, but diabetes only develops in those individuals who lack sufficient insulin secretion to match the degree of insulin resistance. Insulin in those people may be high, yet it is not enough to normalize the level of blood glucose (Røder, Porte, Schwartz, & Kahn, 1998).

Dysfunction of insulin-producing islet cells is a main factor across the progression from prediabetes to diabetes. After the progression from normal glucose tolerance to abnormal glucose tolerance, post-meal blood glucose levels increase initially.

Thereafter, fasting hyperglycemia may develop as the suppression of hepatic (liver) gluconeogenesis fails (Porte, 1991). Despite the fact that the pathophysiology of diabetes differs between type 1 and type 2 diabetes, most of the complications are similar.

12.6.3 Obesity

Overweight and obesity are defined by an excess accumulation of adipose tissue to an extent that impairs both physical and psychosocial health and well-being (Naser, Gruber, & Thomson, 2006). Obesity is considered a health disaster in both Western and non-Western countries (Gallagher, 2000)

The prevalence is escalating significantly in many nations worldwide. This pandemic needs to be stopped if the economic costs, social hazards, morbidity, and mortality of the disease are considered.

12.6.4 Obesity and Type 1 Diabetes

The rising incidence of type 2 diabetes among children and adults is related to the epidemic of obesity. An increase in type 1 diabetes is thought to have similar origins (Arora, 2014). While the underling pathophysiology of type 1 diabetes, which is autoimmune in nature, continues to be investigated and studied, the exact mechanism causing the rise in the incidence of type 1 diabetes remains unclear, particularly in young age groups. One study, which collected data on childhood diabetes from 112 centers around the world, demonstrated an approximate 2.8% annual increase in type 1 diabetes over the period from 1989–2000 (Diamond Project Group, 2006).

The origin of type 1 diabetes, according to twin studies, indicates a joint contribution of environmental and genetic factors (Ershow, 2009). Furthermore, the importance of environmental factors in the development of diabetes is indicated by a significant rise in type 1 diabetes incidence in immigrants from lower to higher incidence regions. Multiple triggers for the development of type 1 diabetes have been investigated, including short-duration or the absence of breastfeeding, exposure to cow's milk protein, and exposure to some kind of infection such as enterovirus or rubella. However, none of these triggering factors has been shown to be the definitive cause (Harder, 2009).

The association between type 1 diabetes and weight gain was first investigated in 1975. The work of Baum and colleagues suggested that there was an association related to overfeeding or to hormonal dysregulation (Baum JD, Ounsted, & Smith, 1975).

The "accelerator hypothesis" proposed by Wilkin (2001) is considered one of the most accepted theories that demonstrates the association between body mass and type 1 diabetes. The authors of this theory suggested that increasing body weight in young age groups increases the risk of developing type 1 diabetes. There is an inverse relationship between body mass index and age at diagnosis. In other words, the higher the body mass index, the younger the age of diagnosis. Furthermore, as young children gain more weight, diabetes can be diagnosed earlier. This is explained by the fact that more weight accelerates insulin resistance, leading to the development of type 1 diabetes in individuals who are predisposed genetically to diabetes. Following this study, many papers were published supporting Wilkin's accelerator hypotheses. One study conducted in the United States in 2003 showed a significant increase in the prevalence of being overweight in children with type 1 diabetes, from 12.6% in the period 1979–1989 to 36.8% in the period 1990–1998. To date, the exact mechanism and relationship between type 1 diabetes and obesity remains inconclusive and needs further explanation (Wilkin, 2001).

12.6.5 Obesity and Type 2 Diabetes

The increased prevalence of obesity these days has drawn attention to the worldwide significance of this problem (Arora, 2014). In the United States, approximately two-thirds of the adult population is considered to be overweight or obese. Similar trends are being noticed worldwide (Tsai, Williamson, & Glick, 2011). Obesity is linked to many medical, psychological, and social conditions, the most devastating of which may be type 2 diabetes. At the start of this century, 171 million people were estimated to have type 2 diabetes, and this figure is expected to increase to 360 million by 2030 (McKeigue, Shah, & Marmot MG, 1991).

Both type 2 diabetes and obesity are associated with insulin resistance. Most obese individuals, despite being insulin resistant, do not develop hyperglycemia. Pancreatic cells release adequate amounts of insulin that are sufficient to overcome insulin level reductions under normal circumstances, thus maintaining normal glucose tolerance (Røder, Porte, Schwartz, & Kahn, 1998).

12.6.6 Obesity and Insulin Resistance

Insulin sensitivity fluctuation occurs across the natural life cycle. For example, insulin resistance is noticed during puberty, in pregnancy, and during the aging process (Kahn, Hull , & Utzschneider, 2006). In addition, lifestyle variations, such as increased carbohydrate intake and increased physical activity, are associated with insulin sensitivity fluctuations (Kasuga, 2006). Obesity is considered the most important factor in the development of metabolic diseases. Adipose tissue affects metabolism by secreting hormones, glycerol, and other substances including leptin, cytokines, adiponectin, and proinflammatory substances, and by releasing nonesterified fatty acids (NEFAs). In obese individuals, the secretion of these substances will be increased (Karpe, Dickmann, & Frayn, 2011).

The cornerstone factor affecting insulin insensitivity is the release of NEFAs. Increased release of NEFAs is observed in type 2 diabetes and in obesity, and it is associated with insulin resistance in both conditions (Jelic, 2007). Shortly after an acute increase of plasma NEFA levels in humans, insulin resistance starts to develop. Conversely, when the level of plasma NEFA decreases, peripheral insulin uptake and glucose monitoring will be improved (Roden et al., 1996).

Insulin sensitivity is determined by another critical factor, which is body fat distribution. Insulin resistance is associated with body mass index at any degree of weight gain. Insulin sensitivity also differs completely in lean individuals because of differences in body fat distribution. Individuals whose fat distribution is more peripheral have more insulin sensitivity than do individuals whose fat distribution is more central (ie, in the abdomen and chest area) (Karpe, Dickmann, & Frayn, 2011).

Differences in adipose tissue distribution help explain, to some extent, how the metabolic effects of subcutaneous and intra-abdominal fat differ. Intra-abdominal fat is more related to the genes that secrete proteins and the specific types of proteins responsible for the production of energy. Adiponectin secretion by omental adipocytes is larger than the amount secreted by subcutaneous-derived adipocytes. Moreover, the quantity secreted from these omental adipocytes is negatively associated with increased body weight (Jelic, 2007). The secretion of NEFAs to different tissue may be affected by their source.

Furthermore, abdominal fat is considered more lipolytic than subcutaneous fat, and it also does not respond easily to the antilipolytic action of insulin, which makes intra-abdominal fat more important in causing insulin resistance, and thus diabetes.[27,28]

Marcial et al[29] further explained the molecular mechanisms of insulin resistance, inflammation, and the development of diabetes. One of the mechanisms of insulin is its effect as an anabolic hormone that enhances glycogen synthesis in liver and muscle. This in turn augments protein synthesis inhibiting the process of proteolysis. Insulin resistance is indeed an important factor in disease process. Fat storage and mobilization are other important factors causing insulin resistance.

Go to:

12.6.7 Obesity and β-cell dysfunction

β-cells play a vital role in regulating insulin release, despite their fragility. The quantity of insulin released by β-cells fluctuates and changes according to the quantity, nature, and route of administration of the stimulus. Therefore, β-cells play a very important role in ensuring that in healthy subjects, concentrations of blood glucose are stable within a relatively normal physiological range. In obesity, insulin sensitivity, as well as the modulation of β-cell function, decreases.[30]

Insulin-resistant individuals, whether slim or fat, have more insulin responses and lower hepatic insulin clearance than those who are insulin sensitive. In a normal healthy subject, there is a continuous feedback relationship between the β-cells and the insulin-sensitive tissues.[10] If the adipose tissue, liver, and muscles demand glucose, this will lead to increased insulin supply by the β-cells. If the glucose levels require stability, changes in insulin sensitivity must be matched by a relatively opposite change in circulating insulin levels. Failure of this process to take place results in a deregulation of glucose levels and the development of DM. If the β-cells are healthy, there is an adaptive response to insulin resistance, which leads to the maintenance of normal levels of glucose. By contrast, when pancreatic β-cells are impaired, abnormal glucose tolerance or abnormal fasting glucose may develop, and it may even be followed by the development of type 2 diabetes.[30]

A continued decline in β-cell function is one of the main causes leading to type 2 diabetes. According to literature, when β-cell dysfunction causes inadequate insulin secretion, fasting blood glucose and postprandial blood glucose will elevate.[31] Subsequently, the decreased efficiency of hepatic and muscle glucose uptake will occur, with the absence or incomplete inhibition of liver glucose

production. Further increases in blood glucose levels will lead to disease severity through glucotoxic effects on the pancreatic β-cells and negative effects on insulin uptake and peripheral tissue sensitivity.[31]

Conversely, in healthy subjects, elevating their blood glucose levels for 20 hours or more has an absolutely inverse action, because it will lead to enhanced β-cell function capacity and improve peripheral insulin uptake.[16] These facts explain that a genetic risk factor is necessary for the occurrence of β-cell function impairment. The progression of time, as well as a pre-existing genetic abnormality in insulin secretion and a subsequently continuous elevation of blood glucose levels, lead to complete β-cell failure.[24]

A second factor that might contribute to a continuous loss of function of β-cells is increasing plasma NEFA levels. Despite the fact that NEFAs play a major role in insulin release, the continuous exposure to NEFAs is related to significant malfunction in glucose-stimulated insulin secretion pathways and reduced insulin biosynthesis. Moreover, the occurrence of insulin resistance in vivo and a failure of the compensatory mechanism of β-cells in humans contributes to increase amounts of NEFA levels produced by lipids.[32]

The two actions of NEFA contribute to a significant etiology that links β-cell dysfunction and insulin resistance in people with type 2 diabetes, and those who are at risk for the disease. The effect of lipotoxic increases in plasma NEFA levels and the rise of glucose levels might produce a more harmful effect known as glucolipotoxicity.[33,34]

Go to:

12.6.8 Conclusion

Diabetes and obesity are chronic disorders that are on the rise worldwide. Body mass index has a strong relationship to diabetes and insulin resistance. In an obese individual, the amount of NEFA, glycerol, hormones, cytokines, proinflammatory substances, and other substances that are involved in the development of insulin resistance are increased. Insulin resistance with impairment of β-cell function leads to the development of diabetes. Gaining weight in early life is associated with the development of type 1 diabetes. NEFA is a cornerstone in the development of insulin resistance and in the impairment of β-cell function. New approaches in managing and preventing diabetes in obese individuals must be studied and investigated based on these facts.

12.6.9 Attributions

FROM Al-Goblan, A. S., Al-Alfi, M. A., & Khan, M. Z. (2014). Mechanism linking diabetes mellitus and obesity. *Diabetes, metabolic syndrome and obesity : targets and therapy*, *7*, 587–591. https://doi.org/10.2147/DMSO.S67400 (CC BY-NC)

This page titled 12.6: Ingestive Behavior in Context is shared under a mixed license and was authored, remixed, and/or curated by multiple authors (ASCCC Open Educational Resources Initiative (OERI)) .

CHAPTER OVERVIEW

13: Sexuality and Sexual Development

> This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

13.1: Sex Hormones, Sexual Motivation, and Reproductive Behaviors
13.2: Sexual Development, Anatomy and Physiology
13.3: Differences of Sexual Development
13.4: Gender
13.5: Sexual Orientation
13.6: Sex Differences and Evolutionary Interpretations

This page titled 13: Sexuality and Sexual Development is shared under a mixed license and was authored, remixed, and/or curated by Naomi Bahm (ASCCC Open Educational Resources Initiative (OERI)) .

13.1: Sex Hormones, Sexual Motivation, and Reproductive Behaviors

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Distinguish between the organizational and activational effects of sex hormones
2. Describe some reciprocal interactions between sex hormones and reproductive behavior
3. Understand basic biological mechanisms regulating sexual behavior and motivation
4. Identify the role of the brain in responding to sexual stimuli
5. Explain the role of hormones in parental behavior

13.1.1 Overview

This module discusses the relationship between sex and hormones (including organizational and activational effects, maturation of the reproductive systems, interactions between hormones and behavior, and anabolic steroids), sexual behavior as a form of motivation (including physiological mechanisms and human sexual behavior and motivation), sex and the brain (including erogenous zones, and the role of the hypothalamus and the pituitary gland), and parental behavior (including hormones and rodent maternal behavior, maternal aggression, and hormones and human maternal behavior).

13.1.2 Human Sexuality in Context

Human sexuality refers to people's sexual interest in and attraction to others, as well as their capacity to have erotic experiences and responses. Sexuality may be experienced and expressed in a variety of ways, including (but not limited to) thoughts, desires, practices, roles, and behaviors. This chapter mostly focuses on the biological and physical aspects of sexuality- human reproductive anatomy and functions, including the human sexual response cycle and the basic biological drive that exists in all species- but also addresses the psychological domains of gender identity and sexual orientation. We begin with hormones, as they play such an integral role in the development and expression of human sexuality.

13.1.3 Sex and Hormones

Because of their inextricable link to the development and maturation of reproductive anatomy (both prenatally and during puberty), as well as the promotion and maintenance of adult reproductive functions, Biological Psychology textbooks often include hormones in the chapter on sex and reproductive behaviors. However, many other bodily functions (such as growth, sleep, hunger, satiation, and maintenance of blood pressure, to name a few) are regulated by hormones, so we have opted to cover basics of the endocrine system (the organs which produce and secrete hormones) in the Nervous System Anatomy chapter. Please refer to that chapter if you need a brief overview of hormones and the endocrine system.

13.1.3.1 Organizational and Activational Effects of Sex Hormones

Although the dichotomy cannot be strictly applied, the effects of steroid sex hormones have typically been characterized as organizational versus activational. Organizational effects result in permanent changes that usually occur early in development, such as the prenatal development of human reproductive structures (also known as **primary sex characteristics**- differences in male and female bodies that are present at birth). Activational effects are usually temporary and occur throughout life, dependent on specific conditions. Thus, structures in the body, brain, and nervous system are thought to be *organized* (in a male-typical or female-typical manner) through the action of steroid hormones early in development, and then later on *activated* by steroid hormones, resulting in structures and behaviors that differ between the sexes (Arnold and Breedlove, 1985). While this simple dichotomy works well with animals with very distinct sexual dimorphism in behavior (such as rats, where only males attempt to mount other rats and only females exhibit a sexually receptive posture called lordosis), it cannot be applied in such a straightforward manner to people. For one thing, humans do not exhibit any sexual behaviors that are strictly stereotyped and unique to either males or females.

Nonetheless, some of the differences between human males and females can be attributed, at least in part, to either the organizational or activational effects of sex hormones.

One example of a permanent organizational effect is the difference between males and females in the secretion of certain hormones (related to reproductive function) by the hypothalamus; adult females follow a pattern of cyclical hormone release (related to their fertility cycles), whereas males do not. Changing the hormones that are present in the body later in life will not change the capacity of the hypothalamus to cycle the release of these hormones. Some **secondary sex characteristics** (changes that occur in male and female bodies during the sexual maturation that occurs with puberty) are also permanent changes- requiring surgery to alter- such as the development of breasts and wider hips in women or a prominent Adam's apple (cartilage around the thyroid in the neck area) and broader shoulders in men. In contrast, some secondary sex characteristics, like the presence of facial hair in men (but not in women), can be modified with changes to hormone levels in the adult body of either sex.

One example of a temporary activational effect is lactation, or the production and secretion of milk from the mammary glands (breasts). Normally this only occurs in females after childbirth, when the necessary combination of the hormones estrogen, progesterone, and prolactin are present. "However, if a man is treated with this hormone combination, his mammary glands not only can produce milk, but he can nurse a baby! Thus, lactation is an "activated-only" action of hormones, and there is no organized sex difference in the mammary gland tissue itself." (Jones and Lopez, 2006, page 466). So, while modern science cannot yet assist transgender women (whose biological birth sex was male) in carrying and giving birth to an infant, they can nonetheless participate in nursing the baby.

13.1.3.2 Sex Hormones and Maturation

The main categories of sex hormones are **androgens** (more prevalent in biological males), of which **testosterone** is a primary example, and **estrogens** (more prevalent in biological females), of which **estradiol** is a primary example. It is important to note that *all* individuals have both androgens ("male hormones") and estrogens ("female hormones"), but the relative proportion of the hormones present in a given body varies, usually correlated with biological sex. Testosterone is secreted by the testes in males and in small amounts by the ovaries in females (although most is converted to estradiol). A small amount of testosterone is also secreted by the adrenal glands (endocrine glands positioned on top of the kidneys) in both sexes. Estrogens are secreted by the ovaries in females and by fat cells in both sexes.

Male and female reproductive systems are different at birth (due to differences in primary sex characteristics), but the gonads (testes in males and ovaries in females) are immature and incapable of producing gametes (sperm in males and eggs in females) or sex hormones. Maturation of the reproductive system occurs during puberty when hormones from the hypothalamus and pituitary gland stimulate the testes or ovaries to start producing sex hormones again. Sex hormones, in turn, lead to the growth and maturation of the reproductive organs, rapid body growth, and the development of secondary sex characteristics, such as pubic and underarm hair in both sexes, facial hair in males and breasts in females.

13.1.3.3 Interactions Between Sex Hormones and Reproductive Behaviors

As discussed in Chapter 4.6, the interaction between hormones and behavior is bidirectional: hormones can influence behavior, and behavior can sometimes influence hormone concentrations. Hormones travel through the blood, influencing the nervous system to regulate an individual's behaviors, some of which are related to sexuality and reproduction (such as aggression, mating, and parenting). Some hormone-behavior interactions that are related specifically to reproductive behaviors are described below.

13.1.3.3.1 Hormonal Influence on Reproductive Behaviors

Hormones coordinate the physiology and behavior of individuals. Over evolutionary time, hormones have been co-opted by the nervous system to influence behavior to ensure reproductive success. For example, the same hormones, testosterone and estradiol, that cause gamete (egg or sperm) maturation also promote mating behavior. This dual hormonal function ensures that mating behavior occurs when animals have mature gametes available for fertilization. Similarly, during pregnancy estrogens and progesterone concentrations are elevated, and these hormones are also involved in maternal behavior in the mothers.

How might hormones affect behavior? While hormones do not *cause* behavioral changes, they influence the components that interact to produce behavior- sensory systems (input), the central nervous system (integration), and muscles and glands (output). In this manner, specific stimuli are more likely to elicit certain responses in the appropriate behavioral or social context. In other words, hormones change the *probability* that a particular behavior will occur in the appropriate situation (Nelson, 2011). This is a

critical distinction that can affect how we think of hormone-behavior relationships. In most cases, hormones can be considered to affect behavior by influencing any or all of these components.

An example of the influence of hormones on a simple behavior is singing in zebra finches. Only male zebra finches sing. If the testes of adult male finches are removed, then the birds reduce singing, but castrated finches resume singing if the testes are reimplanted, or if the birds are treated with either testosterone or estradiol. Although we generally consider androgens to be “male” hormones and estrogens to be “female” hormones, it is common for testosterone to be converted to estradiol in nerve cells. Thus, many male-like behaviors are associated with the actions of estrogens! Singing behavior is most frequent when blood testosterone or estrogen concentrations are high. Males sing to attract mates or ward off potential competitors from their territories.

13.1.3.3.2 Behavioral Influence on Sex Hormones

How might behaviors affect hormones? If a male mouse or rhesus monkey loses a fight, blood testosterone levels decrease for several days or even weeks afterward. Comparable results have also been reported in humans. Testosterone concentrations are affected not only in humans involved in physical combat, but also in those involved in simulated battles. For example, testosterone concentrations were elevated in winners and reduced in losers of regional chess tournaments.

People do not have to be directly involved in a contest to have their hormones affected by the outcome of the contest. Male fans of both the Brazilian and Italian teams were recruited to provide saliva samples to be assayed for testosterone before and after the final game of the World Cup soccer match in 1994. At the end of the game, Brazil and Italy were tied for regulation and overtime, but then Brazil won in penalty shots. The Brazilian fans were elated and the Italian fans were crestfallen. When the samples were assayed, 11 of 12 Brazilian fans who were sampled had increased testosterone concentrations, and 9 of 9 Italian fans had decreased testosterone concentrations, compared with pre-game baseline values (Dabbs, 2000).



Figure **13.1.1**: The expectation of events can influence one’s hormonal activity. How do you think yours is affected if you anticipate going on a date with a romantic interest soon?

In some cases, hormones can be affected by anticipation of behavior (Figure **13.1.1**). For example, testosterone concentrations also influence sexual motivation and behavior in women. In one study, the interaction between sexual intercourse and testosterone was compared with other activities (cuddling or exercise) in women (van Anders, Hamilton, Schmidt, & Watson, 2007). On three separate occasions, women provided a pre-activity, post-activity, and next-morning saliva sample. After analysis, the women’s testosterone was determined to be elevated prior to intercourse as compared to other times. Thus, an anticipatory relationship exists between sexual behavior and testosterone. Testosterone values were higher post-intercourse compared to exercise, suggesting that engaging in sexual behavior may also influence hormone concentrations in women.

13.1.3.4 Anabolic Steroids

The endocrine system can be exploited for illegal or unethical purposes. A prominent example of this is the use of steroid drugs by professional athletes. Commonly used for performance enhancement, anabolic steroids are synthetic versions of the sex hormone testosterone. By boosting natural levels of this hormone, athletes experience increased muscle mass. Synthetic versions of human growth hormone are also used to build muscle mass.

The use of performance-enhancing drugs is banned by all major collegiate and professional sports organizations in the United States because they impart an unfair advantage to athletes who take them. In addition, the drugs can cause significant and dangerous side effects. For example, anabolic steroid use can increase cholesterol levels, raise blood pressure, and damage the liver. Altered testosterone levels (both too low or too high) have been implicated in causing structural damage to the heart, and increasing the risk for cardiac arrhythmias, heart attacks, congestive heart failure, and sudden death. Paradoxically, steroids can cause shriveled testes and enlarged breast tissue in men. In females, their use can cause analogous effects such as an enlarged clitoris and growth of facial hair. In both sexes, their use can promote increased aggression (commonly known as “roid-rage”), depression, sleep disturbances, severe acne, and infertility.

13.1.4 Sexual Behavior as a Form of Motivation

Sex is an important part of most people's lives. From an evolutionary perspective, the reason is obvious—perpetuation of the species. Sexual behavior in humans, however, involves much more than reproduction. Sexual arousal is the drive state that results in thoughts and behaviors related to sexual activity. It is generated by a large range of internal and external mechanisms that are triggered either after the extended absence of sexual activity or by the immediate presence and possibility of sexual activity (or by cues commonly associated with such possibilities). This section provides an overview of some research that has been conducted on sexual behavior and motivation.

Sexual motivation, often referred to as **libido**, is a person's overall sexual drive or desire for sexual activity. This motivation is determined by biological, psychological, and social factors. In most mammalian species, sex hormones control the ability to engage in sexual behaviors. However, sex hormones do not directly regulate the ability to have sexual intercourse in primates (including humans); rather, they are only one influence on the motivation to engage in sexual behaviors. Social factors, such as work, family, and relationship issues also have an impact, as do internal psychological factors like personality and lifestyle stress. Sex drive may also be affected by medical conditions (including illness or injury), medications, and pregnancy.

13.1.4.1 Physiological Mechanisms of Sexual Motivation and Behavior

Much of what we know about the physiological mechanisms that underlie sexual motivation and behavior comes from animal research. Research on male rats suggests that limbic system structures such as the amygdala and nucleus accumbens are especially important for sexual motivation. Damage to these areas results in a decreased motivation to engage in sexual behavior, while leaving the ability to do so intact (Everett, 1990; Figure **13.1.2**). Similar dissociations of sexual motivation and sexual ability have also been observed in the female rat (Becker, Rudick, & Jenkins, 2001; Jenkins & Becker, 2001).

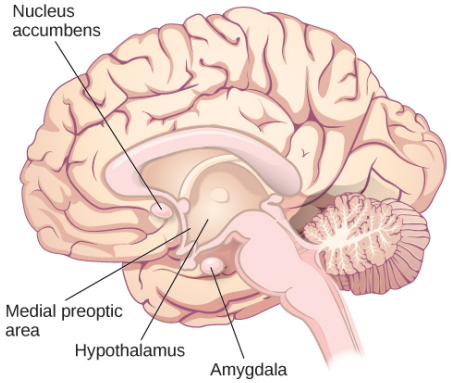


Figure **13.1.2**: The medial preoptic area, an area of the hypothalamus, is involved in the ability to engage in sexual behavior, but it does not affect sexual motivation. In contrast, the amygdala and nucleus accumbens are involved in motivation for sexual behavior, but they do not affect the ability to engage in it.

The hypothalamus plays an important role in motivated behaviors, and sex is no exception. In fact, lesions to an area of the hypothalamus called the medial preoptic area completely disrupt a male rat’s ability to engage in sexual behavior, but, surprisingly, do not change how hard a male rat is willing to work to gain access to a sexually receptive female (Figure **13.1.3**). This suggests that the ability to engage in sexual behavior and the motivation to do so may be mediated by neural systems distinct from one another.

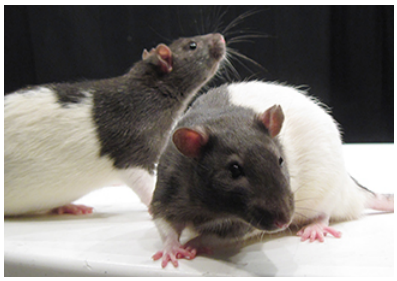


Figure 13.1.3: A male rat that cannot engage in sexual behavior still seeks receptive females, suggesting that the ability to engage in sexual behavior and the motivation to do so are mediated by different systems in the brain.

13.1.4.2 Human Sexual Motivation and Behavior

Although human sexual behavior is much more complex than that seen in rats, some parallels between animals and humans can be drawn from this research. The worldwide popularity of drugs used to treat erectile dysfunction (Conrad, 2005) speaks to the fact that sexual motivation and the ability to engage in sexual behavior can also be dissociated in humans. Moreover, disorders that involve abnormal hypothalamic function are often associated with hypogonadism (reduced function of the gonads) and reduced sexual function (e.g., Prader-Willi syndrome). Given the hypothalamus's role in endocrine function, it is not surprising that hormones secreted by the endocrine system also play important roles in sexual motivation and behavior. For example, many animals show no sign of sexual motivation in the absence of the appropriate combination of sex hormones from their gonads. While this is not the case for humans, there is considerable evidence that sexual motivation for both men and women varies as a function of circulating testosterone levels (Bhasin, Enzlin, Coviello, & Basson, 2007; Carter, 1992; Sherwin, 1988). In other words, testosterone maintains libido (sex drive) in both males and females.

Vasopressin is also involved in the male arousal phase, and the increase of vasopressin during erectile response may be directly associated with increased motivation to engage in sexual behavior. The relationship between hormones and female sexual motivation is not as well understood, largely due to the overemphasis on male sexuality in Western research. Estrogen and progesterone typically regulate motivation to engage in sexual behavior for females, with estrogen increasing motivation and progesterone decreasing it. The levels of these hormones rise and fall throughout a woman's menstrual cycle. Research suggests that testosterone, oxytocin, and vasopressin are also implicated in female sexual motivation in similar ways as they are in males, but more research is needed to understand these relationships.

13.1.5 Sex and the Brain

The brain is the organ that translates the nerve impulses from the skin into pleasurable sensations. It controls nerves and muscles used during sexual behavior. The brain regulates the release of hormones, which are believed to be the physiological origin of sexual desire. The cerebral cortex (the outer layer of the brain that allows for thinking and reasoning) is believed to be the origin of sexual thoughts and fantasies. Deep to the cortex is the limbic system, a collection of structures believed to be the origin of emotions and feelings, also important for sexual behavior. The limbic system includes the amygdala, hippocampus, cingulate gyrus, and septal nucleus. The septal nucleus, an area that receives reciprocal connections from many other brain regions (including the hypothalamus and the amygdala), seems to play an important role in sexual pleasure. This region shows rhythmic spiking activity during sexual orgasm, and is also one of the brain regions that rats will most reliably voluntarily self-stimulate (Olds & Milner, 1954). In humans, placing a small amount of acetylcholine into this region, or stimulating it electrically, has been reported to produce a feeling of imminent orgasm (Heath, 1964; Heath, 1972).

13.1.5.1 Erogenous Zones

At first glance- or touch for that matter- the glans of the clitoris and penis are the parts of our anatomies that seem to bring the most pleasure. However, these two organs pale in comparison to our central nervous system's capacity for pleasure. Extensive regions of the brain and brainstem are activated when a person experiences pleasure, including the insula, temporal cortex, limbic system, nucleus accumbens, basal ganglia, superior parietal cortex, dorsolateral prefrontal cortex, and cerebellum (Ortigue et al., 2007; Figure 13.1.4). Neuroimaging techniques show that these regions of the brain are active when patients have spontaneous orgasms involving no direct stimulation of the skin (e.g., Fadul et al., 2005) and when experimental participants self-stimulate erogenous zones (e.g., Komisaruk et al., 2011). Erogenous zones are particularly sensitive areas of skin (which may be different across

individuals), and (like all body areas with sensation) are connected (via the nervous system) to the somatosensory cortex in the brain (refer to the section on the central nervous system).



Figure **13.1.4**: Some of the many regions of the brain and brainstem activated during pleasure experiences.

A study by Nummenmaa and his colleagues (2016) used a unique method to test the hypothesis that the more sensitive areas of our bodies have greater potential to evoke pleasure. The Nummenmaa research team showed experimental participants images of same- and opposite-sex bodies. They then asked the participants to color the regions of the body that, when touched, they or members of the opposite sex would experience as sexually arousing while masturbating or having sex with a partner. Nummenmaa found the expected "hotspot" erogenous zones around the external sex organs, breasts, and anus, but also reported areas of the skin beyond these hotspots: "[T]actile stimulation of practically all bodily regions trigger sexual arousal…." Moreover, he concluded, "[H]aving sex with a partner…"—beyond the hotspots—"…reflects the role of touching in the maintenance of…pair bonds." This also underlines the fact that individuals are different, and erogenous zones vary correspondingly.

13.1.5.2 The Role of the Hypothalamus and the Pituitary Gland

One brain structure that is particularly important for sexual functioning is the **hypothalamus** (Figure **13.1.5**). This is a small area at the base of the brain consisting of several groups of neuron cell bodies that receive input from the limbic system. One of the reasons for the importance of the hypothalamus is that it controls the pituitary gland, which secretes hormones that control the other glands of the body.

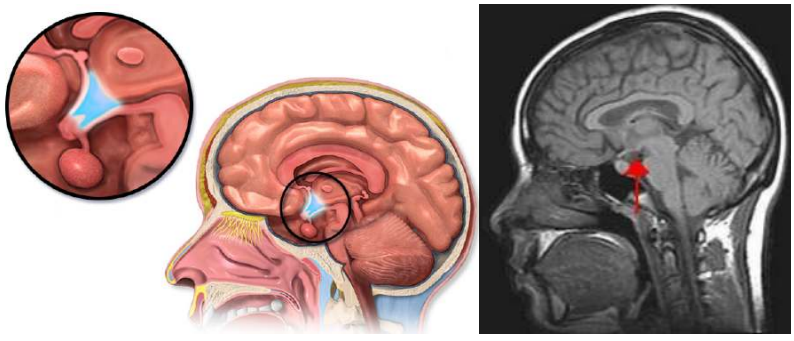


Figure **13.1.5**: The hypothalamus controls the pituitary gland. Midsagittal views of the human brain: First, a drawing with the roughly triangular-shaped hypothalamus colored blue, and the small, oval-shaped pituitary gland visible attached below it (circle inset shows a magnified view). In the next image, an arrow points to the location of the hypothalamus on a brain scan.

Several important hormones related to sexual function are secreted by the pituitary gland, which is divided into anterior and posterior sections. The anterior pituitary secretes follicle-stimulating hormone (FSH), luteinizing hormone (LH), and prolactin. FSH and LH are responsible for ovulation in females and sperm production in males. Prolactin and oxytocin (which is released by the posterior pituitary) stimulate milk production in lactating females. Oxytocin is sometimes called the "love" hormone and is

believed to be involved with maintaining close relationships. It is released with sexual activity during orgasm, causing rhythmic contractions in the uterus and penis. Oxytocin is also released in females during childbirth (causing the contractions of the uterus that push the baby out) and during breast-feeding (as part of the milk let-down reflex).

13.1.6 Parental Behavior

Parental behavior can be considered to be any behavior that contributes directly to the survival of fertilized eggs or offspring that have left the body of the female. There are many patterns of parental care in mammals. The developmental status of the newborn is an important factor driving the type and quality of parental care in a species- **altricial** young are born in an underdeveloped state and require extensive parental care to survive, whereas **precocial** young are born in a more advanced and mature state, already mobile and able to feed themselves. Maternal care is much more common than paternal care.

13.1.6.1 Hormones and Rodent Maternal Behavior

The vast majority of research on the hormonal correlates of mammalian parental behavior has been conducted on rats. Rats bear altricial young, and mothers perform a cluster of stereotyped maternal behaviors, including nest building, crouching over the pups to allow nursing and to provide warmth, pup retrieval, and increased aggression directed at intruders. If you expose nonpregnant female (or male) rats to pups, their most common reaction is to huddle far away from them in fear, since rats avoid new things (neophobia). However, if you expose adult female rats to pups every day, they soon begin to behave maternally.

Of course a new mother needs to act maternally as soon as her offspring arrive- not a week later. Hormones trigger the onset of maternal behavior in rats, and several methods of study (such as hormone removal and replacement therapy) have been used to investigate rat maternal behavior. A fast decline of blood concentrations of progesterone in late pregnancy (after sustained high concentrations of this hormone), in combination with high concentrations of estradiol (and probably prolactin and oxytocin), induces female rats to behave maternally almost immediately in the presence of pups. This pattern of hormones during the delivery of pups overrides the usual fear response of adult rats toward pups, and permits the onset of maternal behavior.

The medial preoptic area (of the hypothalamus) is critical for the expression of rat maternal behavior, and the amygdala appears to inhibit the expression of maternal behavior. The fearful response of adult rats towards pups is apparently mediated by chemosensory information, and lesions of the amygdala (or sensory pathways to the amygdala) allow the expression of maternal behavior. Hormones or sensitization likely act to disinhibit the amygdala, thus permitting the occurrence of maternal behavior. Although correlations have been established, direct evidence of brain structural changes in human mothers remains unspecified (Fleming & Gonzalez, 2009).

13.1.6.2 Maternal Aggression

Laboratory rats are usually docile, but mothers can be quite aggressive toward animals that venture too close to their litter. Progesterone appears to be the primary hormone that induces this maternal aggression in rodents, but species differences exist. The role of maternal aggression in women’s behavior has not been adequately described or tested.

13.1.6.3 Hormones and Human Maternal Behavior

A series of elegant experiments by Alison Fleming and her collaborators studied the endocrine correlates of human mothers' behavior and maternal attitudes, as expressed in self-report questionnaires. Responses such as patting, cuddling, or kissing the baby were called affectionate behaviors; talking, singing, or cooing to the baby were considered vocal behaviors. Both affectionate and vocal behaviors were considered approach behaviors. Basic caregiving activities, such as changing diapers and burping the infants, were also recorded. In these studies, no relationship between hormone concentrations and maternal responsiveness (as measured by attitude questionnaires) was found. For example, most women showed an increasing positive self-image during early pregnancy that dipped during the second half of pregnancy, but recovered after parturition (childbirth). A related dip in feelings of maternal engagement occurred during late pregnancy, but rebounded substantially after birth in most women



Figure **13.1.6**: Although cortisol may not directly increase maternal behaviors, the next time your mom gives you a hug you know one hormone to thank.

However, when behavior (rather than questionnaire responses) was compared with hormone concentrations, a different story emerged. Blood plasma concentrations of cortisol were positively associated with approach behaviors. In other words, women who had high concentrations of blood cortisol (in samples obtained immediately before or after nursing) engaged in more physically affectionate behaviors (Figure **13.1.6**) and talked more often to their babies than mothers with low cortisol concentrations. Additional analyses from this study revealed that the correlation was even greater for mothers who had reported positive maternal regard (feelings and attitudes) during gestation (pregnancy). Indeed, nearly half of the variation in maternal behavior among women could be accounted for by cortisol concentrations and positive maternal attitudes during pregnancy. Presumably, cortisol does not induce maternal behaviors directly, but it may act indirectly by increasing the mother's general level of arousal, thus increasing her responsiveness to infant-generated cues. New mothers with high cortisol concentrations were also more attracted to their infant's odors, were superior in identifying their infants, and generally found cues from infants highly appealing (Fleming, Steiner, & Corter, 1997).

13.1.7 Summary

Hormones are inextricably linked to the development and maturation of reproductive anatomy (both prenatally and during puberty), as well as the promotion and maintenance of adult reproductive functions. The effects of steroid sex hormones are typically characterized as organizational versus activational. Organizational effects result in permanent changes that usually occur early in development, such as the prenatal development of primary sex characteristics (differences in male and female bodies that are present at birth). Activational effects are usually temporary and occur throughout life, dependent on specific conditions.

The main categories of sex hormones are androgens (more prevalent in biological males, e.g. testosterone), and estrogens (more prevalent in biological females, e.g. estradiol). All individuals have both androgens and estrogens, but the relative proportion of the hormones present varies, usually correlated with biological sex. Testosterone is secreted by the testes in males and in small amounts by the ovaries in females. A small amount of testosterone is also secreted by the adrenal glands in both sexes. Estrogens are secreted by the ovaries in females and by fat cells in both sexes.

The interaction between hormones and behavior is bidirectional: hormones can influence behavior, and behavior can sometimes influence hormone concentrations. Hormones do not cause behavioral changes, but they influence the components that interact to produce behavior- sensory systems, the central nervous system, and muscles and glands. In this manner, specific stimuli are more likely to elicit certain responses in the appropriate behavioral or social context. For behavioral effects on hormones, in animal models, losing a fight results in decreased blood testosterone levels. Comparable results have been reported in humans.

Research on rats suggests that the motivation to engage in sexual behavior is separate from the ability to engage in sexual behavior. The hypothalamus plays an important role in motivated behaviors, including sex. In humans, testosterone maintains libido (sex drive) in both males and females.

Although the glans of the clitoris and penis are known to be associated with pleasure, extensive regions of the brain and brainstem are also activated when a person experiences pleasure. Erogenous zones are particularly sensitive areas of skin, which vary across individuals.

The hypothalamus is important for sexual functioning, particularly because it controls the pituitary gland. Hormones related to sexual function secreted by the pituitary gland include follicle-stimulating hormone (FSH), luteinizing hormone (LH), and prolactin from the anterior pituitary, and oxytocin from the posterior pituitary. FSH and LH are responsible for ovulation in females and sperm production in males. Prolactin and oxytocin stimulate lactation. Oxytocin causes rhythmic contractions in the uterus and penis (associated with orgasm), as well as uterine contractions during childbirth. Oxytocin is also believed to be involved in maintaining close relationships.

Hormones trigger the onset of maternal behavior in rats. A fast decline of progesterone in late pregnancy, combined with high concentrations of estradiol, induce female rats to behave maternally (overriding the usual fear response of adult rats toward pups). The medial preoptic area is critical for the expression of rat maternal behavior, and the amygdala appears to inhibit the expression of maternal behavior. Progesterone appears to be the primary hormone that induces maternal aggression (in rats) toward animals that venture too close to their litter.

In human studies, no relationship between hormone concentrations and maternal responsiveness (on attitude questionnaires) was found. However, women who had high concentrations of blood cortisol engaged in more physically affectionate behaviors and talked more often to their babies than mothers with low cortisol concentrations. The correlation was even greater for mothers who had reported positive maternal feelings and attitudes during pregnancy.

13.1.8 References

Arnold, A.P. and Breedlove, S.M. (1985). Organizational and activational effects of sex steroids on brain and behavior: A reanalysis. *Hormones and Behavior*, 19(4), 469-498.

Fleming, A. S., & Gonzalez, A. (2009). Neurobiology of human maternal care. In P. T. Ellison & P. B. Gray (Eds.), *Endocrinology of social relationships* (pp. 294–318). Cambridge, MA: Harvard University Press.

Fleming, A. S., Steiner, M., & Corter, C. (1997). Cortisol, hedonics, and maternal responsiveness in human mothers. *Hormones and Behavior*, 32, 85–98.

Heath, R.G. (1972). Pleasure and brain activity in man. *The Journal of Nervous and Mental Disease*, *154*, 3-18. (As cited in Panksepp, 2004, page 397.)LeVay, S. (1991). A difference in hypothalamic structure between heterosexual and homosexual men. *Science*, *253*(5023), 1034-1037. doi: 10.1126/science.1887219

Jones, R.E. & Lopez, K.H. (2006) *Human Reproductive Biology, 3rd edition*. Academic Press.

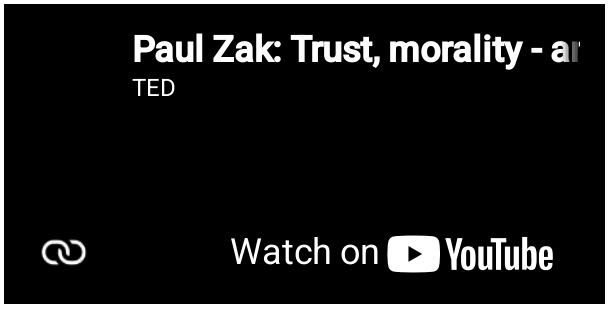
Panksepp, J. (2004). *Affective neuroscience: The foundations of human and animal emotions*; *Chapter 12: The varieties of love and lust: Neural control of sexuality*. Oxford University Press.

13.1.9 Additional Resources

Video: Endocrinology Video (Playlist) - This YouTube playlist contains many helpful videos on the biology of hormones, including reproduction and behavior. This would be a helpful resource for students struggling with hormone synthesis, reproduction, regulation of biological functions, and signaling pathways.

https://www.youtube.com/playlist?list=PLqTetbgey0aemiTfD8QkMsSUq8hQzv-vA

Video: Paul Zak: Trust, morality - and oxytocin- This Ted talk explores the roles of oxytocin in the body. Paul Zak discusses biological functions of oxytocin, like lactation, as well as potential behavioral functions, like empathy.

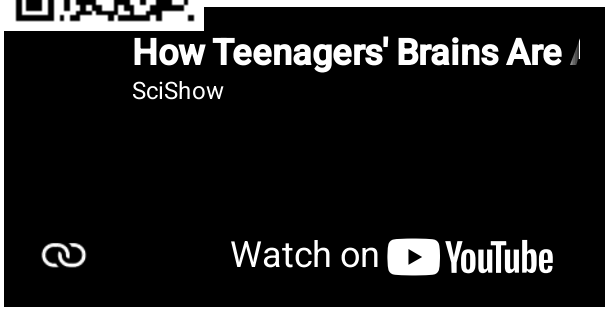




Video: Sex Differentiation- This video discusses gonadal differentiation, including the role of androgens in the development of male features.



Vid rain Explained- This is a great video explaining the roles of hormones during puberty.





Web: Society for Behavioral Neuroendocrinology - This website contains resources on current news and research in the field of neuroendocrinology.

http://sbn.org/home.aspx

13.1.10 Attributions

1. Figures:
    1. Male/Female "date", goo.gl/m25gce, licensed CC0 Public Domain, sourced from NOBA Hormones & Behavior by Nelson
    2. Midsagittal brain figure, sourced from OpenStax Psychology 2e Sexual Behavior by Spielman et al. (no author or licensing details given)
    3. WT and TK rat photo by Jason Snyder from Washington, DC, United States, licensed CC BY 2.0 via Wikimedia Commons
    4. Colors related to brain arterial supply (mostly) removed by Naomi Bahm, figure originally from Brain areas by Frank Gaillard, https://goo.gl/yCKuQ2, CC-BY-SA 3.0, sourced from NOBA Human Sexual Anatomy and Physiology by Lucas & Fox (who added Identifying asterisks to original image)
    5. Left image: Hypothalamus location by Blausen.com staff (2014). "Medical gallery of Blausen Medical 2014". WikiJournal of Medicine 1 (2). DOI:10.15347/wjm/2014.010. ISSN 2002-4436., licensed CC BY 3.0 via Wikimedia Commons; Right image: "Hypothalamus controls the pituitary gland", image is in the public domain, sourced from Physical changes in Adolescence by Paris et al.
    6. Left image: Mother cuddling daughter by Maria Grazia Montagnari, https://goo.gl/LY1Tq0, licensed CC BY 2.0, sourced from NOBA Hormones & Behavior by Nelson; Right image: Mother and young child with a disability by AnikaMeyer, CC BY-SA 4.0 via Wikimedia Commons
2. Text adapted from:
    1. Hormones & Behavior by Randy J. Nelson, licensed CC BY-NC-SA 4.0 via Noba Project.
    2. "Physical Changes in Adolescence" by Paris, Ricardo, Raymond, & Johnson, LibreTexts is licensed under CC BY.
    3. " Development of Sexual Identity" by Paris, Ricardo, Raymond, & Johnson, LibreTexts is licensed under CC BY.
    4. 10.3 Sexual Behavior (Psychology 2e) by Rose M. Spielman, William J. Jenkins, and Marilyn D. Lovett, licensed CC BY 4.0 via OpenStax.
    5. "Introduction to the Reproductive System" by Suzanne Wakim & Mandeep Grewal, LibreTexts is licensed under CC BY-NC.
    6. Human Sexual Anatomy and Physiology by Don Lucas and Jennifer Fox, licensed CC BY-NC-SA 4.0 via Noba Project.
    7. Anabolic Steroids section: "Gonadal and Placental Hormones" by Whitney Menefee, Julie Jenks, Chiara Mazzasette, & Kim-Leiloni Nguyen, LibreTexts is licensed under CC BY.
    8. Drive States by Sudeep Bhatia and George Loewenstein, licensed CC BY-NC-SA 4.0 via Noba Project. ***NOTE:*** Although a few sentences of general text were curated from this source (specifically located in the sections Sexual Behavior as a Form of Motivation, and Sex and the Brain), most of the content it contains (related to the drive state of sexual arousal) is based on rodent research that is presented as though it has been verified to be equivalent to human physiology. In fact, the source they site for male sexuality overlapping with areas associated with aggression specifically states "In many mammals, the vigor of male sexuality and male assertiveness (i.e., social dominance) tend to go together...The fact that male sexuality and aggression interact to a substantial extent in subcortical areas of the brain is now a certainty ... The meaning of this interaction for human sexuality remains regrettably pregnant with ambiguities." (Panksepp 2004, page 239)
3. Changes: Text (and images) from above eight sources pieced together with some modifications, transitions and additional content added (particularly the sections: Sex and Hormones, Organizational and Activational Effects of Sex Hormones, and parts of Sex Hormones and Maturation) by Naomi I. Gribneau Bahm, PhD., Psychology Professor at Cosumnes River College, Sacramento, CA.
4. (particularly the Sex and Hormones section, Organizational and Activational Effects of Sex Hormones section, and part of the Sex Hormones and Maturation section)

This page titled 13.1: Sex Hormones, Sexual Motivation, and Reproductive Behaviors is shared under a mixed license and was authored, remixed, and/or curated by Naomi Bahm (ASCCC Open Educational Resources Initiative (OERI)) .

- **22.2: Introduction to the Reproductive System** by Suzanne Wakim & Mandeep Grewal is licensed CK-12. Original source: https://www.ck12.org/book/ck-12-human-biology/.
- **13.3: Physical Changes in Adolescence** by Paris, Ricardo, Rymond, & Johnson is licensed CC BY 4.0. Original source: https://open.umn.edu/opentextbooks/textbooks/child-growth-and-development.

13.2: Sexual Development, Anatomy and Physiology

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe the process of prenatal sexual differentiation in humans.
2. Identify male and female reproductive structures that originate from homologous embryonic tissues versus those that are formed from separate (Wolffian and Müllerian) duct systems.
3. Identify the functions of the main reproductive structures for both males and females.
4. Understand the basic events occurring in the ovaries and uterus during the human menstrual cycle.
5. Explain the four phases in the human sexual response cycle (Masters and Johnson).

13.2.1 Overview

This section covers development of the male and female reproductive systems (including prenatal sex differentiation, homologous structures and reproductive ducts, and sexual maturation during puberty), and adult sexual anatomy and physiology (including the male reproductive system, the female reproductive system, the menstrual cycle, and the sexual response cycle.

13.2.2 Development of the Reproductive Systems

The **reproductive system** is the human organ system responsible for the production of gametes (sperm or eggs), the meeting of gametes during fertilization, and the carrying of an embryo/fetus. The reproductive system is the only body system that differs substantially between individuals, typically divided into male and female forms (Figure **13.2.1**), although there is actually a range of biological sex (see the differences of sexual development section). Many embryonic structures that will develop into the reproductive system start out the same in males and females, but by birth, the reproductive systems have differentiated. This section will describe how male and female reproductive structures typically form during development and the role of several hormones.

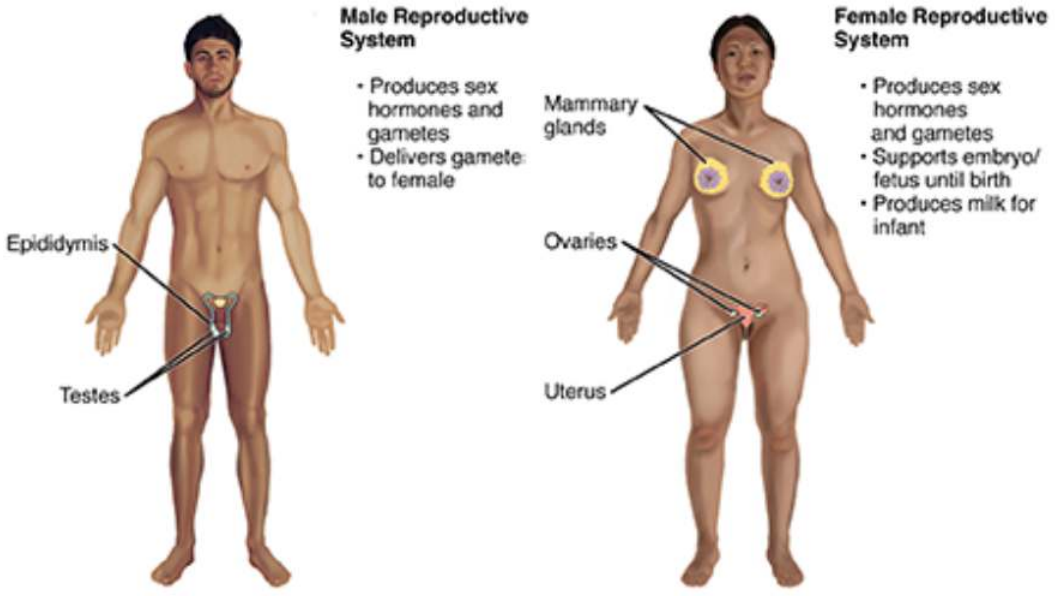


Figure **13.2.1**: Reproductive Systems in the Human Male and Female. Both systems produce sex hormones and gametes. The male reproductive system additionally delivers gametes to the female; the epididymis and testes are labeled. The female reproductive system additionally supports an embryo/fetus until birth and produces milk for the infant; the mammary glands, ovaries, and uterus are labeled.

The **gonads** (the testes in males and the ovaries in females) produce gametes (sperm in males and eggs in females). A **gamete** is a haploid cell- a cell that contains one full set of chromosomes, or half the number needed to create a new individual. When two haploid gametes combine during fertilization, they form a single diploid cell, which now contains two full sets of chromosomes (the necessary number to create a new individual) and is called a zygote (fertilized egg). In mammals, the ovum always contains an X chromosome, which can be fertilized by a sperm bearing either an X or a Y chromosome; this process is called sex determination. XX is the chromosomal sex of a female mammal and XY is the chromosomal sex of a male mammal.

Besides producing gametes, the gonads also produce sex hormones. **Sex hormones**, such as androgens and estrogens, are endocrine hormones that control the development of sex organs before birth, sexual maturation at puberty, and reproduction once sexual maturation has occurred. Other reproductive system organs have various functions, such as maturing gametes, delivering gametes to the site of fertilization, and providing an environment for the development and growth of offspring.

13.2.2.1 Prenatal Sex Differentiation

The process of becoming female or male is called sexual differentiation. Although both the X and the Y chromosomes are called sex chromosomes, only the Y chromosome contains genes that determine sex. A single gene on the Y chromosome, called SRY (for "sex-determining region Y gene"; Figure **13.2.2**), triggers male development in the embryo. Without a Y chromosome, the embryo develops a female body plan, so you can think of a female as the default sex of the human species (as is the case for all mammals).

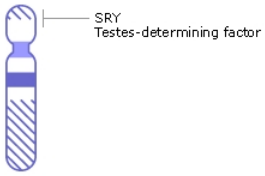


Figure **13.2.2**: The SRY gene on the short arm of the Y chromosome causes the undifferentiated gonads of an embryo to develop into testes. Otherwise, the gonads develop into ovaries.

The gonads of the embryo are initially undifferentiated, meaning that they are indeterminate- identical in XX and XY embryos- and able to become either testes or ovaries. Starting around the sixth week after conception in genetically male (XY) embryos, the SRY gene initiates the production of a protein called testes determining factor. Testes determining factor causes the undifferentiated gonads to develop into testes. The testes secrete hormones — including testosterone — that trigger other changes in the developing offspring (now called a fetus), causing it to develop a complete male reproductive system. Without a Y chromosome, an embryo will develop ovaries that will then produce estrogens. Estrogens, in turn, enable the formation of the other organs of a female reproductive system.

13.2.2.2 Homologous Structures and Reproductive Ducts

Undifferentiated embryonic tissues develop into different structures in male and female fetuses. Structures that arise from the same tissues in males and females are called **homologous structures**. The testes and ovaries, for example, are homologous structures that develop from the undifferentiated gonads of the embryo. Likewise, the penis and clitoris are homologous structures that develop from the same embryonic tissues, as are the scrotum and the skin folds of the labia majora. The formation of typical male external anatomy (the penis and scrotum) requires the conversion of testosterone into a more potent androgen called dihydrotestosterone, whereas the formation of typical female external anatomy (the clitoris and the labia) does not require hormones. In this manner, the same embryonic tissue develops into one structure or another in a given individual, or occasionally forms a structure that is intermediate between the typical male or female form.

Not all tissues in the prenatal reproductive tract have the potential to develop into either male or female forms. The internal reproductive structures (for example the uterus, uterine tubes, and part of the vagina in females; and the epididymis, vas deferens, and seminal vesicles in males) form from one of two rudimentary duct systems in the embryo. For full reproductive function in the adult, one set of these ducts must develop properly, and the other must degrade. In males, cells in the testes secrete **Müllerian inhibiting substance**, which causes the **Müllerian duct** (the female duct) to degenerate. At the same time, testosterone secretion stimulates growth of the male tract, the **Wolffian duct**. Without Müllerian inhibiting substance, the Müllerian duct will develop; without testosterone, the Wolffian duct will degrade. Thus, the developing offspring will be female (Figure **13.2.3**), following the default female body form.

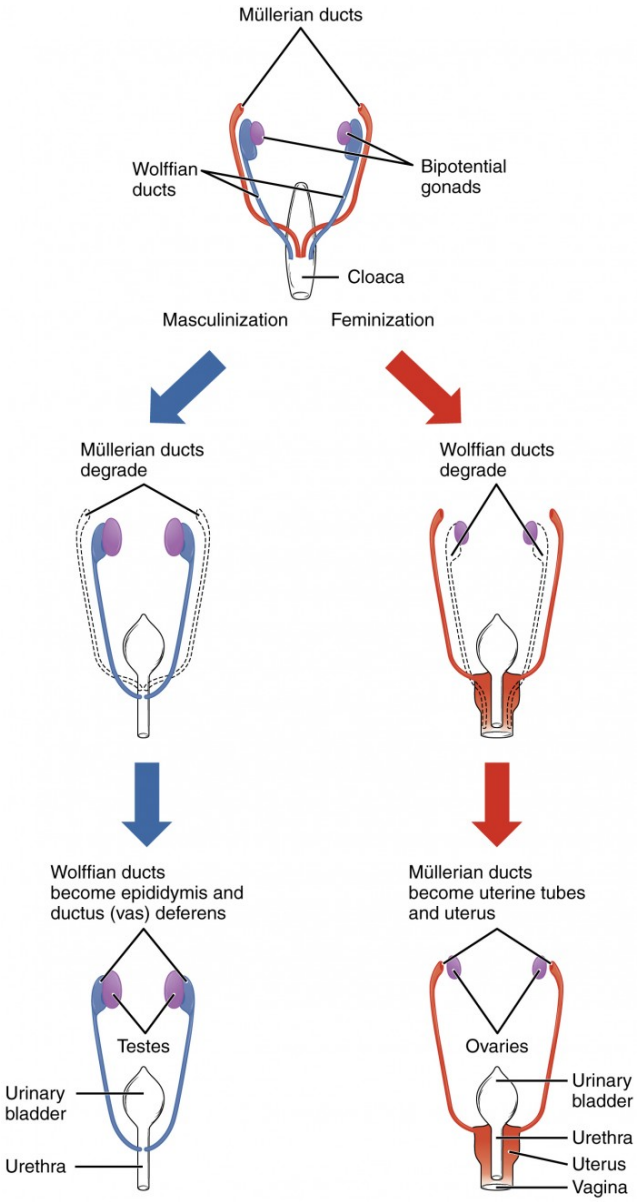


Figure **13.2.3**: Differentiation of the male and female reproductive systems occurs during the fetal period of development. Initially, the cloaca, bipotential gonads, and both Wolffian and Mullerian ducts are present. In both sexes, the cloaca becomes the urinary bladder and urethra.With masculinization (shown on the left), the Mullerian ducts degrade and the Wolffian ducts become the epididymis and the ductus (vas) deferens. The bipotential gonads become testes. With feminization (shown on the right), the Wolffian ducts degrade and the Mullerian ducts become the uterine tubes and the uterus. The bipotential gonads become ovaries.

13.2.2.3 Sexual Maturation during Puberty

Puberty is a *period of rapid growth and sexual maturation*. These changes typically begin sometime between the ages of eight and fourteen, and start with an overall physical growth spurt, often most noticeable in the adolescent's increased height. Generally, girls begin puberty at around ten years of age and boys begin approximately two years later. Pubertal changes usually take around three to four years to complete.

Typically, the physical growth spurt is followed by the development of sexual maturity. Changes in the primary sexual characteristics (reproductive organs) in males includes growth of the testes, penis, and scrotum, and in females the growth and maturation of the ovaries, uterus and external genitalia (vulva). Important signals of sexual maturity include **spermarche** *or first ejaculation of semen* in males, and **menarche** *or the first menstrual period* in females. Stress and higher percentage of body fat are correlated with menstruation at younger ages.

Secondary sexual characteristics *are visible physical changes not directly linked to reproduction, but that signal sexual maturity*. For males this includes broader shoulders and a lower voice as the larynx grows. Body hair becomes coarser and darker, and hair growth occurs in the pubic area, under the arms and on the face. For females breast development occurs around age 10, although full development takes several years. Hips broaden and pubic and underarm hair develops and also becomes darker and coarser.

13.2.3 Adult Sexual Anatomy and Physiology

13.2.3.1 The Male Reproductive System

The main structures of the male reproductive system are external to the body (Figure **13.2.4**). The two testes (singular, testis) hang between the thighs in a sac of skin called the scrotum. The testes produce both sperm and testosterone. Testosterone production is under the control of luteinizing hormone (LH) from the pituitary gland, which stimulates cells in the testes to secrete testosterone. Both follicle stimulating hormone (FSH) from the pituitary gland and testosterone are needed for normal spermatogenesis to be maintained in the testes. Additionally, testosterone maintains libido (sex drive) and plays a role in erection, allowing sperm to be deposited within the female reproductive tract. Testosterone is necessary for the proper functioning of the prostate gland, and is also important for muscle development and bone growth.

Resting atop each testis is a coiled structure called the epididymis (plural, epididymes). The function of the epididymes is to mature and store sperm. The penis is a tubular organ that contains the urethra and has the ability to stiffen during sexual arousal. Sperm passes out of the body through the urethra during a sexual climax (orgasm). This release of sperm is called ejaculation.

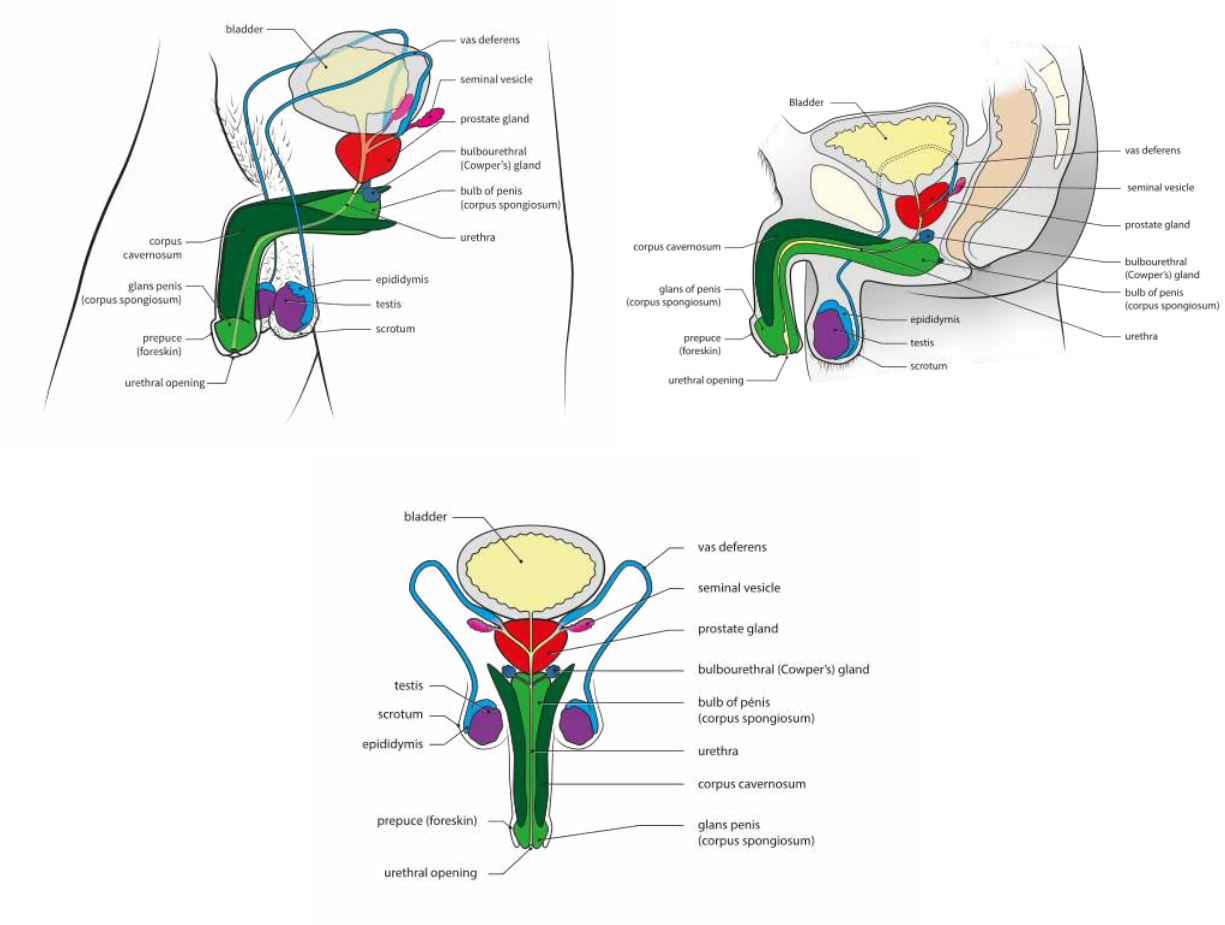


Figure **13.2.4**: The main structures of the male reproductive system in 3D view, sagittal view, and frontal view. Note that structures that are homologous to female structures (originating from the same embryonic tissue, such as the testes) are color-coded to match the female reproductive system diagrams in Figure **13.2.5**.

In addition to these external organs, there are several ducts and glands that are internal to the body. The ducts, which include the vas deferens (also called the ductus deferens), transport sperm from the epididymis to the urethra. The glands, which include the

prostate gland and seminal vesicles, produce fluids that become part of semen. Semen is the fluid that carries sperm through the urethra and out of the body. It contains substances that control pH and provide sperm with nutrients for energy.

13.2.3.2 The Female Reproductive System

The main structures of the female reproductive system are internal to the body (Figure **13.2.5**). They include the paired ovaries, which are small, oval structures that produce eggs and secrete estrogens. The two uterine tubes (also known as Fallopian tubes or oviducts) start near the ovaries and end at the uterus. Their function is to transport eggs from the ovaries to the uterus. If fertilization of an egg occurs, it usually happens while it is traveling through the Fallopian tube. The uterus is a pear-shaped muscular organ that functions to carry a fetus until birth. It can expand greatly to accommodate a growing fetus, and its muscular walls can contract forcefully during labor to push the baby into the vagina. The vagina is a tubular tract connecting the uterus to the outside of the body. The vagina is where sperm are usually deposited during sexual intercourse (via ejaculation). The vagina is also called the birth canal because a baby travels through the vagina to leave the body during birth.

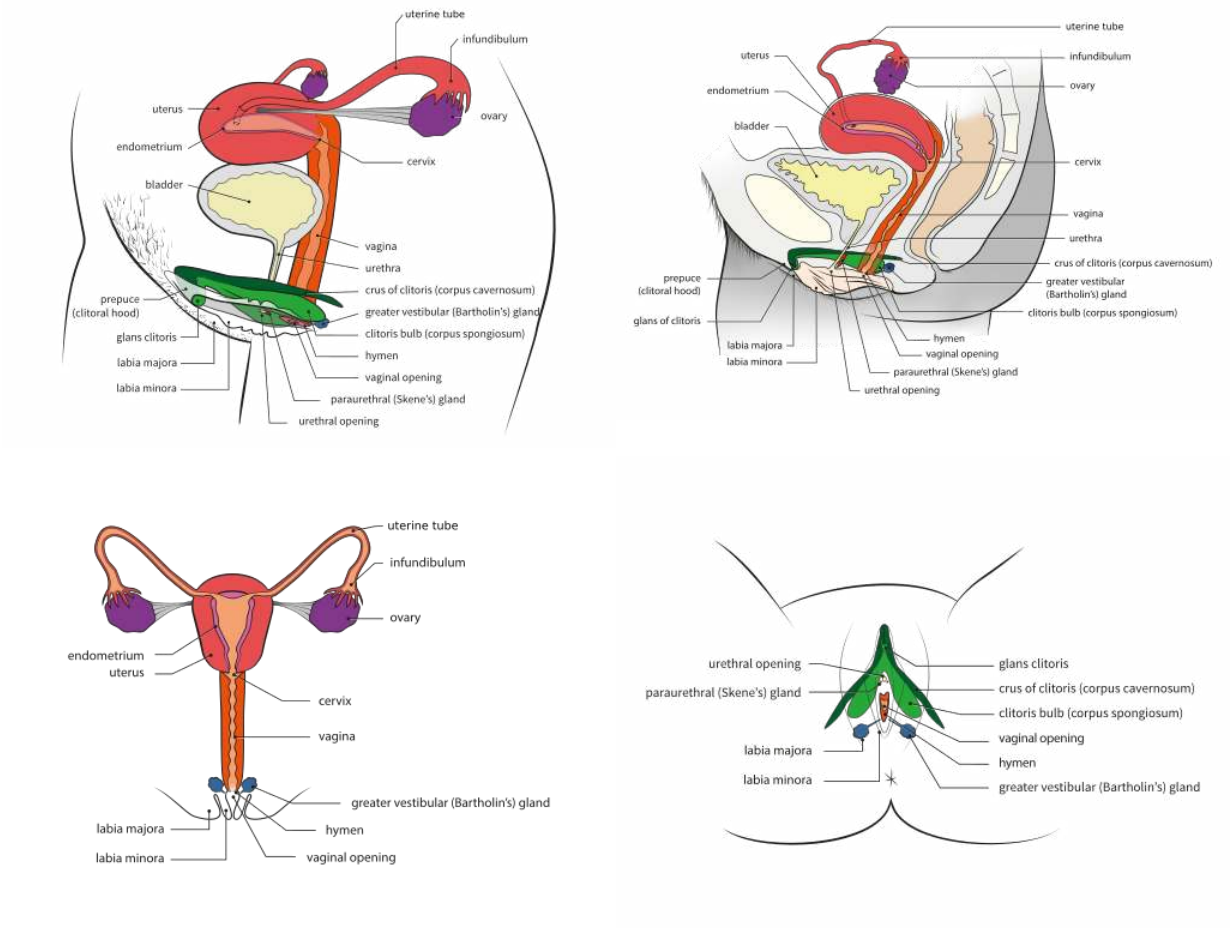


Figure **13.2.5**: The main structures of the female reproductive system in 3D view, sagittal view, frontal view, and inferior view (between the legs). Note that structures that are homologous to male structures (originating from the same embryonic tissue, such as the ovaries) are color-coded to match the male reproductive system diagrams in Figure **13.2.4**.

The external structures of the female reproductive system are referred to collectively as the vulva. The vulva includes the clitoris, which is homologous to the male penis (the erectile tissues of the penis and clitoris are colored dark green and light green in Figure **13.2.4** and Figure **13.2.5**), and the two pairs of labia (singular, labium), which surround and protect the openings of the urethra and vagina.

13.2.3.3 The Menstrual Cycle

The **menstrual cycle** refers to natural changes that occur in the female reproductive system each month during the reproductive years. The cycle is necessary for the production of eggs and the preparation of the uterus for pregnancy. It involves changes in both the ovaries and the uterus and is controlled by pituitary and ovarian hormones. Day 1 of the cycle is the first day of the menstrual period**,** when bleeding from the uterus begins as the built-up endometrium lining the uterus is shed. The endometrium builds up again during the remainder of the cycle, only to be shed again during the beginning of the next cycle if pregnancy does not occur. In the ovaries, the menstrual cycle includes the development of a follicle, ovulation of a secondary oocyte (the name given to an egg before maturation), and the degeneration of the follicle if pregnancy does not occur. Both uterine and ovarian changes during the menstrual cycle are generally divided into three phases, although the phases are not the same in the two organs.

13.2.3.3.1 Ovarian Cycle

The events of the menstrual cycle that take place in the ovaries make up the ovarian cycle. It consists of changes that occur in the follicles of one of the ovaries. The ovarian cycle is divided into the following three phases: 1) the follicular phase (several follicles begin maturing, but only one is selected to mature completely), 2) ovulation (the mature egg is released), and 3) the luteal phase (progesterone maintains the lining of the uterus for potential implantation of the fertilized egg).

Follicle-stimulating hormone (FSH), secreted by the pituitary gland, rises and causes several follicles to begin to mature. One maturing follicle becomes dominant and starts releasing estrogen. The resulting rising levels of estrogen prevent multiple follicles from fully developing; non-dominant follicles undergo atresia, or degeneration. The continued rise in estrogen triggers a luteinizing hormone (LH) surge from the pituitary gland, which in turn stimulates ovulation. During the luteal phase, progesterone is secreted by the corpus luteum (a structure formed from the remnants of the follicle that released the egg) until the end of the cycle, when the corpus luteum either degenerates (if the egg was not fertilized) or is maintained by the new pregnancy.

13.2.3.3.2 Uterine Cycle

The events of the menstrual cycle that take place in the uterus make up the uterine cycle. This cycle consists of changes that occur mainly in the endometrium, which is the layer of tissue that lines the uterus. The uterine cycle is divided into the following three phases: 1) menstruation (shedding of the unfertilized egg and the endometrium lining that has built up), 2) the proliferative phase (the lining of the uterus grows again), and 3) the secretory phase (endometrium is prepared to receive a fertilized egg).

The ovarian cycle, the uterine cycle, and the changes in hormone levels that occur during the menstrual cycle are illustrated in Figure 13.2.6. Note that the changes in the lining of the endometrium are illustrated in the part showing the uterine cycle phases.

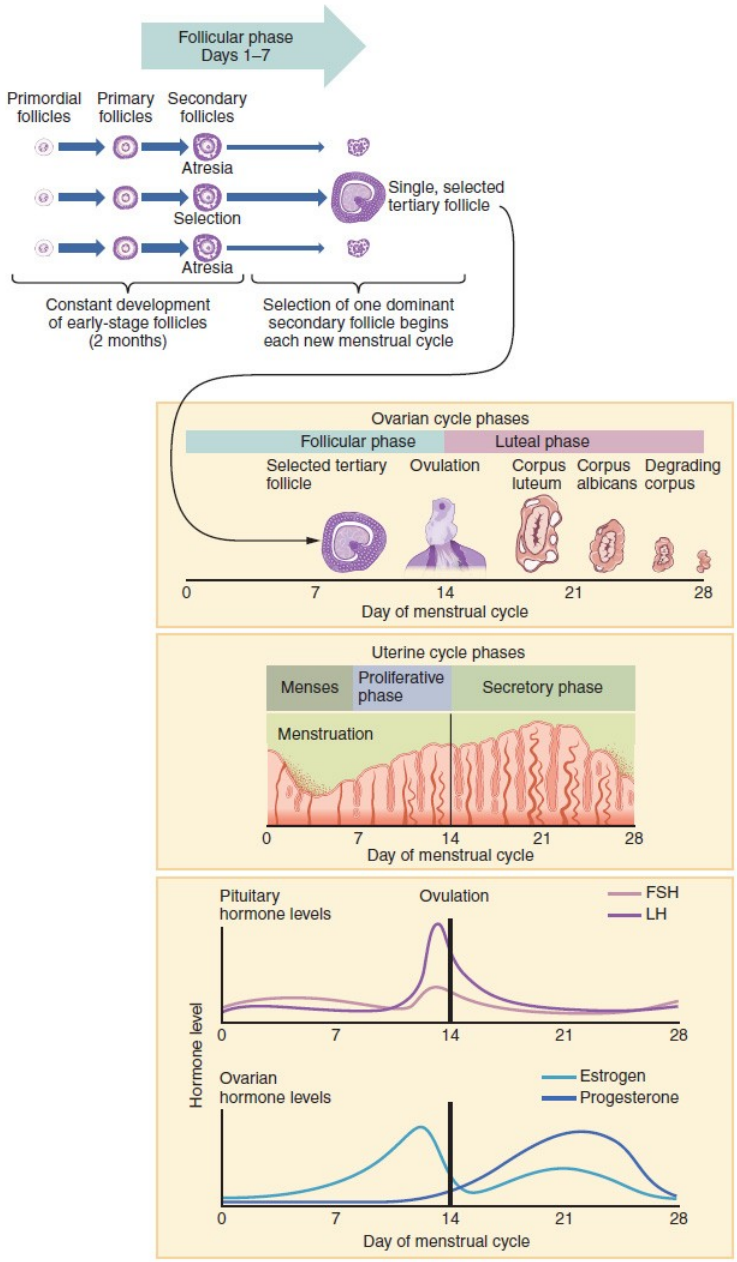


Figure **13.2.6**:

The correlation of the hormone levels and their effects on the female reproductive system is shown in this timeline of the ovarian and menstrual cycles. The menstrual cycle begins at day one with the start of menses. Ovulation occurs around day 14 of a 28-day cycle, triggered by the LH surge.

13.2.3.4 The Sexual Response Cycle

In 1966, William Masters and Virginia Johnson published a book detailing the results of their observations of nearly 700 people who agreed to participate in their study of physiological responses during sexual behavior. Masters and Johnson observed people having intercourse in a variety of positions, as well as people masturbating, manually or with the aid of a device. Measurements of physiological variables, such as blood pressure and respiration rate, as well as measurements of sexual arousal, such as vaginal lubrication and penile tumescence (swelling associated with an erection) were recorded. In total, Masters and Johnson observed nearly 10,000 sexual acts as a part of their research (Hock, 2008).

Based on these observations, Masters and Johnson divided the sexual response cycle (Masters & Johnson, 1966) into four phases that are fairly similar in men and women:

1. Excitement: Activation of the sympathetic branch of the autonomic nervous system defines the *excitement phase*; heart rate and breathing accelerates, along with increased blood flow to the penis, vaginal walls, clitoris, and nipples. Involuntary muscular movements (**myotonia**), such as facial grimaces, also occur during this phase.
2. Plateau: Blood flow, heart rate, and breathing intensify during the *plateau phase*. During this phase females experience an **orgasmic platform**—the outer third of the vaginal walls tightening—and males often exhibit a release of pre-ejaculatory fluid.
3. Orgasm: The shortest but most pleasurable phase is the *orgasm phase*. After reaching its climax, neuromuscular tension is released and the hormone **oxytocin** floods the bloodstream—facilitating emotional bonding. Although the rhythmic muscular contractions of an orgasm are temporally associated with ejaculation in males, orgasm and ejaculation are actually two separate physiological processes (and can thus occur independently).
4. Resolution: The body returns to a pre-aroused state in the *resolution phase*.

The final phase, resolution, is where the main differences between males and females occur. Males enter a **refractory period** of being unresponsive to sexual stimuli. The length of this period depends on age, frequency of recent sexual behavior, level of intimacy with a partner, and novelty. Because females do not have a refractory period, they have a greater potential—physiologically—of having multiple orgasms. Ironically, females are also more likely to "fake" having orgasms (Opperman et al., 2014). As such, whereas the graph for the male sexual response cycle shows only the most common progression through the four phases including orgasm (with a potential partial second cycle starting after the refractory period ends), the graph for the female sexual response cycle shows three alternative patterns- A. progression through all four stages with multiple orgasms, B. prolonged time in the plateau phase with no orgasm, and C. progression through all four stages with a single orgasm (Figure **13.2.7**).

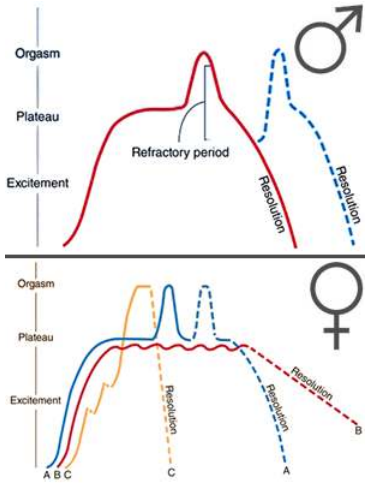


Figure **13.2.7**:

Diagram of the common human sexual response cycles as described by Masters and Johnson

Of interest to note, the sexual response cycle is a universal response to sexual stimuli, and occurs regardless of the type of sexual behavior—whether the behavior is masturbation; romantic kissing; or oral, vaginal, or anal sex (Masters & Johnson, 1966). Further, a partner (of any sex/gender) or environmental object is sufficient, but not necessary, for the sexual response cycle to occur.

Feature: Human Biology in the News

Lung, heart, kidney, and other organ transplants have become relatively commonplace, so when they occur, they are unlikely to make the news. However, when America's first penis transplant took place, it was considered very newsworthy.

In 2016, Massachusetts General Hospital in Boston announced that a team of its surgeons had performed the first penis transplant in the United States. The patient who received the donated penis was a 64-year-old cancer patient. During the 15-hour procedure, the intricate network of nerves and blood vessels of the donor penis were connected with those of the penis recipient. The surgery went well, but doctors reported it would be a few weeks until they would know if normal urination would be possible, and even longer before they would know if sexual functioning would be possible. At the time that news of the surgery was reported in the media, the patient had not shown any signs of rejecting the donated organ. The surgeons also

reported they were hopeful that such transplants would become relatively common, and that patient populations would expand to include wounded warriors and transgender males seeking to transition.

The 2016 Massachusetts operation was not the first penis transplant ever undertaken. The world's first successful penis transplant was actually performed in 2014 in Cape Town, South Africa. A young man who had lost his penis from complications of a botched circumcision at age 18 was given a donor penis three years later. That surgery lasted nine hours and was highly successful. The young man made a full recovery and regained both urinary and sexual functions in the transplanted organ.

In 2005, a man in China also received a donated penis in a technically successful operation. However, the patient asked doctors to reverse the procedure just two weeks later, because of psychological problems associated with the transplanted organ for both himself and his wife.

13.2.4 Summary

The reproductive system is responsible for the production of gametes (sperm or eggs), the meeting of gametes during fertilization, and the carrying of an embryo/fetus. The gonads (the testes in males and the ovaries in females) produce gametes and sex hormones. Sex hormones (androgens and estrogens) control the development of sex organs before birth, sexual maturation at puberty, and reproduction once sexual maturation has occurred.

Sexual differentiation is the process of becoming male or female. A gene on the Y chromosome called SRY is critical for stimulating testis development and subsequently the development of male reproductive structures. Without a Y chromosome, the embryo develops a female body plan.

Undifferentiated embryonic tissues develop into different homologous structures in male and female fetuses, such as the testes and ovaries, or the penis and clitoris. However, the tissue that forms the internal reproductive structures stems from ducts that will develop into only male (Wolffian) or only female (Müllerian) structures. To be able to reproduce as an adult, one of these systems must develop properly and the other must degrade.

Further development of the reproductive systems occurs at puberty. The increase in sex steroid hormones leads to maturation of the gonads and other reproductive organs. Important signals of sexual maturity include spermarche (first ejaculation of semen) in males, and menarche (first menstrual period) in females. Increases in sex steroid hormones also lead to the development of secondary sex characteristics such as breast development in girls and facial hair and larynx growth in boys.

The main male reproductive structures are external to the body: the penis and the paired testes. The penis contains the urethra and is able to stiffen during sexual arousal. Sperm passes out of the body through the urethra during orgasm, called ejaculation. The testes produce sperm and testosterone, and are located in the scrotum. Luteinizing hormone (LH) stimulates cells in the testes to secrete testosterone. Both follicle stimulating hormone (FSH) and testosterone are needed for spermatogenesis. Additionally, testosterone maintains libido (sex drive), plays a role in erection, is necessary for prostate gland function, and is important for muscle development and bone growth. Internal ducts and glands include the vas deferens (which transports sperm from the epididymis to the urethra), the prostate gland and the seminal vesicles (which produce fluids that become part of semen).

The external structures of the female reproductive system are referred to as the vulva: the clitoris and the two pairs of labia surrounding and protecting the openings of the urethra and vagina. The main female reproductive structures are internal to the body and include the paired ovaries (which produce eggs and secrete estrogens), the two Fallopian tubes (which transport eggs from the ovaries to the uterus), the uterus (a muscular organ that carries a fetus until birth), and the vagina. The vagina is a tubular tract connecting the uterus to the outside of the body. Sperm are usually deposited in the vagina during sexual intercourse (via ejaculation), and a baby travels through the vagina (also called the birth canal) to leave the body.

The menstrual cycle in females is necessary for the production of eggs and the preparation of the uterus for pregnancy. It involves changes in both the ovaries and the uterus and is controlled by pituitary and ovarian hormones. The ovarian cycle is divided into the follicular phase (several follicles begin maturing, but only one is selected to mature completely), ovulation (the mature egg is released), and the luteal phase (progesterone maintains the lining of the uterus for potential implantation of the fertilized egg). The uterine cycle is divided into menstruation (shedding of the unfertilized egg and the endometrium lining that has built up), the proliferative phase (the lining of the uterus grows again), and the secretory phase (endometrium is prepared to receive a fertilized egg).

Masters and Johnson (1966) divided the sexual response cycle into four phases that are fairly similar in men and women: excitement (heart rate and breathing accelerates, increased blood flow to the penis, vaginal walls, clitoris, and nipples), plateau (blood flow, heart rate, and breathing intensify; females experience an orgasmic platform—the outer third of the vaginal walls tightening—and males often exhibit a release of pre-ejaculatory fluid), orgasm (shortest but most pleasurable phase), and resolution (the body returns to a pre-aroused state). The main differences between males and females occur during resolution. Males enter a refractory period of being unresponsive to sexual stimuli. Females do not have a refractory period, so they have a greater potential for having multiple orgasms.

13.2.5 References

Note: These are no references added to the content of this page. Refer to the text attributions to locate citations for articles from outside sources used.

13.2.6 Additional Resources

13.2.7 Attributions

Figures:

1. Reproductive systems figure- cropped from **Organ Systems of the Human Body:** Organs that work together are grouped into organ systems. (https://open.oregonstate.education/aandp); Reproductive figure https://open.oregonstate.education/a...ody/#fig_1_2_2; originally found in "Human Organs and Organ Systems" by Suzanne Wakim & Mandeep Grewal, LibreTexts is licensed under CC BY-NC.
2. Human chromosome Y by National Center for Biotechnology Information, Public domain, via Wikimedia Commons
3. Sexual differentiation by OpenStax College, CC BY 3.0, via Wikimedia Commons
4. Male reproductive system: 3D view, Sagittal view, and Frontal view by R. Dewaele (Bioscope, Unige), J. Abdulcadir (HUG), C. Brockmann (Bioscope, Unige), O. Fillod, S. Valera-Kummer (DIP), www.unige.ch/ssi, CC BY-SA 4.0, via Wikimedia Commons
5. Female reproductive system: 3D view, Sagittal view, Frontal view 1, and Frontal view 2 by R. Dewaele (Bioscope, Unige), J. Abdulcadir (HUG), C. Brockmann (Bioscope, Unige), O. Fillod, S. Valera-Kummer (DIP), www.unige.ch/ssi, CC BY-SA 4.0, via Wikimedia Commons
6. Menstrual cycle timeline- from " Anatomy and Physiology of the Female Reproductive System" by LibreTexts is licensed under notset .
7. Sexual response cycle by Avril1975, Public domain, via Wikimedia Commons

Text adapted from:

1. "Introduction to the Reproductive System" by Suzanne Wakim & Mandeep Grewal, LibreTexts is licensed under CC BY-NC.
2. "Anatomy of the Male Reproductive System" by Whitney Menefee, Julie Jenks, Chiara Mazzasette, & Kim-Leiloni Nguyen, LibreTexts is licensed under CC BY.
3. "Human Organs and Organ Systems" by Suzanne Wakim & Mandeep Grewal, LibreTexts is licensed under CC BY-NC.
4. Hormones & Behavior by Randy J. Nelson, licensed CC BY-NC-SA 4.0 via Noba Project.
5. "Development of the Male and Female Reproductive Systems" by Whitney Menefee, Julie Jenks, Chiara Mazzasette, & Kim-Leiloni Nguyen, LibreTexts is licensed under CC BY.
6. Sexual Maturation during Puberty section (heavily edited): " Growth in Adolescence" by Martha Lally & Suzanne Valentine-French, LibreTexts is licensed under CC BY-NC-SA.
7. "Physical Changes in Adolescence" by Paris, Ricardo, Raymond, & Johnson, LibreTexts is licensed under CC BY .
8. "Functions of the Male Reproductive System" by Suzanne Wakim & Mandeep Grewal, LibreTexts is licensed under CC BY-NC.
9. "Menstrual Cycle" by Suzanne Wakim & Mandeep Grewal, LibreTexts is licensed under CC BY-NC .
10. 10.3 Sexual Behavior (Psychology 2e) by Rose M. Spielman, William J. Jenkins, and Marilyn D. Lovett, licensed CC BY 4.0 via OpenStax.
11. Human Sexual Anatomy and Physiology by Don Lucas and Jennifer Fox, licensed CC BY-NC-SA 4.0 via Noba Project.

12. Human Biology in the News feature: "Structures of the Male Reproductive System" by Suzanne Wakim & Mandeep Grewal, LibreTexts is licensed under CC BY-NC.

Changes: Text (and images) from above twelve sources pieced together with some modifications, transitions and additional content added by Naomi I. Gribneau Bahm, PhD., Psychology Professor at Cosumnes River College, Sacramento, CA.

This page titled 13.2: Sexual Development, Anatomy and Physiology is shared under a mixed license and was authored, remixed, and/or curated by Naomi Bahm (ASCCC Open Educational Resources Initiative (OERI)) .

- **22.2: Introduction to the Reproductive System** by Suzanne Wakim & Mandeep Grewal is licensed CK-12. Original source: https://www.ck12.org/book/ck-12-human-biology/.
- **22.7: Menstrual Cycle** by Suzanne Wakim & Mandeep Grewal is licensed CK-12. Original source: https://www.ck12.org/book/ck-12-human-biology/.
- **22.3: Structures of the Male Reproductive System** by Suzanne Wakim & Mandeep Grewal is licensed CK-12. Original source: https://www.ck12.org/book/ck-12-human-biology/.

13.3: Differences of Sexual Development

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Identify the basic characteristics of the most common sex chromosome abnormalities (Turner Syndrome and Klinefelter Syndrome).
2. Understand the term intersex and the various conditions it encompasses.
3. Explain the role of hormones in the development of anomalous sexual differentiation for XX individuals with congenital adrenal hyperplasia (CAH) and XY individuals with androgen insensitivity syndrome (AIS).
4. Explain how David Reimer's case challenged Dr. John Money's theory of psychosexual neutrality.

13.3.1 Overview

This section discusses some of the many ways in which sexual development can differ from typical XX/male and XY/female formats. Note that while the scientific literature on this topic typically labels these conditions "Disorders of Sexual Development (DSD)", we prefer to refer to them as "Differences of Sexual Development". Chromosomal abnormalities (such as Triple X syndrome, Turner syndrome, and Klinefelter syndrome), intersex conditions, anomalous female differentiation (such as congenital adrenal hyperplasia) and anomalous male sexual differentiation (such as 5α-reductase deficiency and androgen insensitivity syndrome) demonstrate some of the diverse variations of biological sex found in humans. The case of David Reimer illustrates that raising a typically-developed biological male (who lost his penis in infancy due to a circumcision accident) as a girl can fail catastrophically.

13.3.2 Chromosomal Abnormalities

13.3.2.1 Aneuploidy

As discussed in the chapter on genetics, a typical human individual has 22 pairs of autosomes ("self" chromosomes) and one pair of sex chromosomes (either XX for females or XY for males), resulting in a total of 46 chromosomes (Figure **13.3.1**).

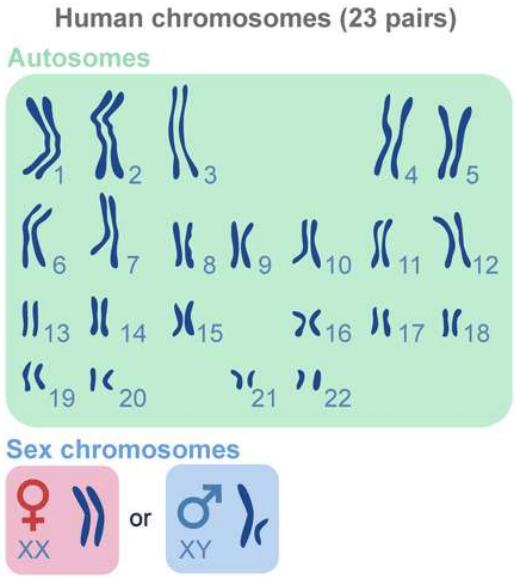


Figure **13.3.1**:Humans have 23 pairs of chromosomes. Pairs 1-22 are autosomes. Females have two X chromosomes, and males have an X and a Y chromosome

Sometimes errors occur during the formation of the gametes (eggs in females and sperm in males), resulting in missing chromosomes or extra chromosomes. When an individual inherits too many or two few chromosomes, it is called a **chromosomal abnormality**. (Or, more specifically, **aneuploidy**- the presence of an abnormal number of chromosomes in a cell.) The most common cause of chromosomal abnormalities is older age of the mother. As a female ages, the immature eggs in her ovaries (which have been present since before birth, unlike the sperm cells in males) are more likely to suffer damage due to longer term exposure to environmental factors. Consequently, some gametes do not divide evenly when they are forming, and some eggs have more (or less) than 23 chromosomes. Often when the error includes one of the autosomes, the resulting zygote (fertilized egg) is not viable. In fact, it is believed that close to half of all zygotes have an abnormal number of chromosomes. Most of these zygotes fail to develop and are spontaneously aborted by the body of the mother (often without her knowledge that conception ever occurred).

One of the few chromosomal abnormalities that more commonly survives (to birth and decades beyond) is Down syndrome (trisomy 21- three copies of chromosome number 21). In addition to some degree of intellectual disability and characteristic facial features, affected individuals often have heart defects and other health problems. The overall occurrence of Down Syndrome is one in every 691 births, but it increases to one in every 300 births for women aged 35 and older.

Interestingly, chromosomal errors involving the sex chromosomes (the 23rd pair), called **sex chromosome aneuploidy**, are less lethal. The exception is having only a single Y chromosome- there are many important genes on the X chromosome, so a zygote with only one Y chromosome (and no X chromosome) will not survive. At least 1 in every 1,000 conceptions results in a variation of chromosomal sex beyond the typical XX or XY sets. Some of these variations include XXX (Figure **13.3.2**), XYY, XXY, or even a single X (Dreger, 1998).



Figure **13.3.2**: 9 year old girl with Triple X Syndrome.

Some individuals with atypical sex chromosomes may have unusual physical characteristics, such as being taller than average, having a thick neck, or being sterile (unable to reproduce); but in many cases, these individuals have no cognitive, physical, or sexual issues (Wisniewski et al., 2000). These sex-linked chromosomal disorders are briefly described in Table **13.3.1**. It is even possible to have four or more sex chromosomes (XXXX, XXYY, XXXYY, etc.).

Table **13.3.1**: Sex-Linked Chromosomal Disorders

Disorder	Sex Appearance & Fertility	Characteristics
Turner Syndrome (XO)	Female, but immature (infertile)	Affects cognitive functioning and sexual maturation; short stature may be noted.
Klinefelter Syndrome (XXY)	Male, may have some breast development (often infertile)	The Y chromosome stimulates the growth of male genitalia, but the additional X chromosome inhibits this development (may have low levels of testosterone).
Supermale Syndrome (XYY)	Male (usually normal fertility)	Few symptoms, which may include being taller than average, acne, and an increased risk of learning problems. Generally otherwise normal.
Triple X Syndrome (XXX)	Female (usually normal fertility)	May result in being taller than average, having learning difficulties, decreased muscle tone, seizures, and kidney problems.

The more common of these sex chromosome disorders are Turner Syndrome and Klinefelter Syndrome. **Turner Syndrome** occurs when one of the X chromosomes is missing or damaged and the resulting zygote has XO sex chromosomes. This occurs in 1 of every 2,500 live female births (Carroll, 2007) and affects the individual's cognitive functioning and sexual maturation. The external genitalia appear normal, but breasts and ovaries do not develop fully and the woman does not menstruate. Turner's syndrome also results in short stature and other physical characteristics, such as a webbed neck, shield-shaped thorax, and cardiac defects. Figure **13.3.3** illustrates the symptoms of Turner's syndrome and Figure **13.3.4** shows pictures of 18 individuals of Asian descent with Turner's syndrome.

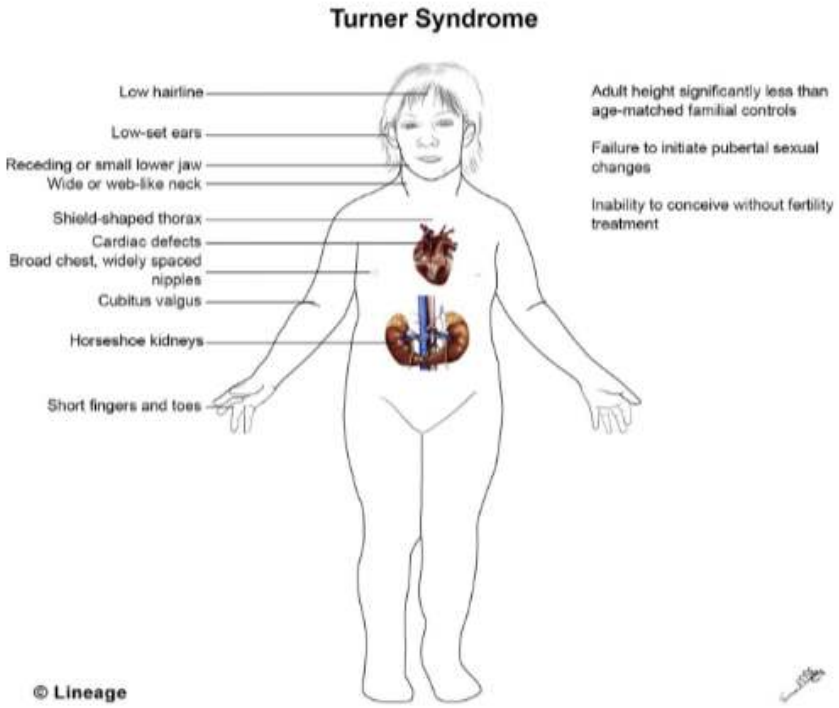


Figure **13.3.3**: Symptoms of Turner's syndrome.



Figure **13.3.4**: Eighteen Asian girls and women with Turner's syndrome.

Klinefelter Syndrome (XXY) *results when an extra X chromosome is present in the cells of a male* and occurs in 1 out of 700 live male births. The Y chromosome stimulates the growth of male genitalia, but the additional X chromosome inhibits this development. An individual with Klinefelter Syndrome often has some breast development, wide hips, infertility (this is the most common cause of infertility in males), small testicular size, and low levels of testosterone (National Institutes of Health, 2019). Figure **13.3.5** illustrates the symptoms of Klinefelter's syndrome on the left and a 29 year old male with Klinefelter`s syndrome who is undergoing testosterone treatment on the right. Regular testosterone injections can promote strength and facial hair growth, build a more muscular body type, increase sexual desire, and enlarge the testes.

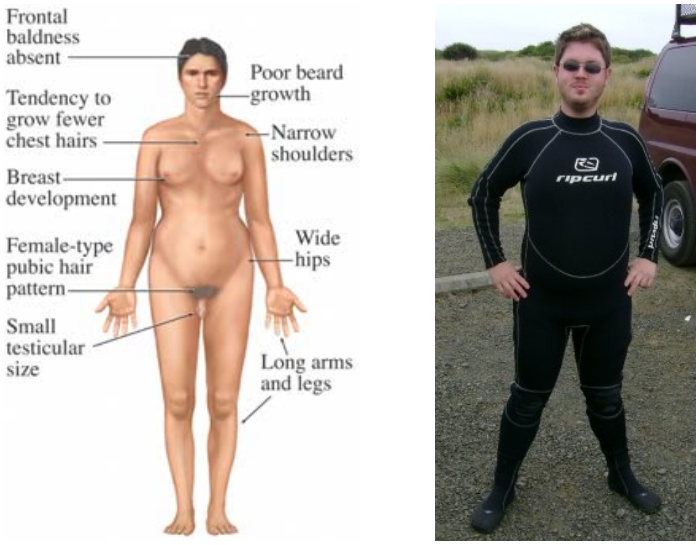


Figure **13.3.5**: Symptoms of Klinefelter's syndrome [left] and 29 year old male undergoing testosterone treatment [right].

13.3.2.2 Other Chromosomal Alterations

As mentioned in Section 13.2, the SRY gene on the Y chromosome is the gene that directs development of the testes (typically leading to male reproductive structure development). Occasionally during sperm formation the SRY gene can cross over from the Y chromosome to the X chromosome. (Gene cross over regularly occurs between the 22 autosomes during the formation of gametes, but usually does *not* occur between the X and Y chromosomes, since they do not have the same genes.) If a sperm carrying a Y chromosome that has lost its SRY gene fertilizes an egg, although chromosomally genetically male (XY), the resulting embryo will not form testes and consequently will not develop male reproductive structures. (For mammals, female development is the "default" in the absence of hormones.) Likewise, if a sperm carrying an X chromosome that has an SRY gene attached to it fertilizes an egg, although chromosomally genetically female (XX), the resulting embryo will develop testes (which will then produce testosterone) and reproductive structures will develop in a male pattern. The individual will be infertile, however, since other genes on the Y chromosome are necessary for viable sperm production. This is another way in which abnormalities on the X or Y chromosomes (rather than having entire chromosomes that are extra or missing) produce variations in human sex characteristics.

13.3.3 Variations in Reproductive Structures

13.3.3.1 Intersex Conditions

In cases where hormones are not produced (or responded to) following the two most common patterns, a fetus may develop biological characteristics in between typical male and typical female structures. These people are considered to have intersex conditions. According to the Intersex Society of North America, intersex is a general term used for a variety of conditions in which a person is born with a reproductive or sexual anatomy that doesn’t fit the typical definitions of male or female. For example, a person might be born appearing to be male on the outside, but having mostly female-typical anatomy on the inside. Or a person may be born with genitals that seem to be intermediate between the usual male and female types—for example, a girl may be born with a notably large clitoris, or lacking a vaginal opening, or a boy may be born with a micropenis, or with a scrotum that is divided so that it has formed more like labia. Some people have mosaic genetics, meaning that two separate sets of chromosomes merged into one individual early in embryonic development. Any given body cell will contain one set or the other, and if all cells have either XX chromosomes or XY chromosomes, the condition may never even be detected. However, if some of their cells have XX chromosomes and some of them have XY chromosomes, separate areas of the body may have developed differently as a result. For example, a person may have an ovary on one side of the body and a testis on the other side. This is the true meaning of the word "hermaphrodite"- having both male and female gonadal tissue in the same body. (It is also possible to have a mixture of ovarian and testicular tissue in the gonads.) The intersex conditions discussed below are sometimes called "pseudohermaphrodites"- having something similar to both male and female anatomical structures in the same body (Figure **13.3.6**).



Figure 13.3.6: Wax figure of human intersex genitalia from the Panoptikum House of Wax, Hamburg.

Though intersex is commonly thought of as an inborn condition, intersex anatomy isn't always apparent at birth. Some intersex traits are not recognizable until puberty or later in life (interACT 2021), or never even realized at all. Sometimes a person isn't found to have intersex anatomy until they reach the age of puberty, or find themselves to be infertile as an adult, or die of old age and are autopsied. Some people live and die with intersex anatomy without anyone (including themselves) ever knowing. Which variations of sexual anatomy count as intersex? In practice, different people have different answers to that question. That's not surprising, because intersex isn't a discrete category. Nature presents us with a spectrum of sexual anatomy. Breasts, penises, clitorises, scrotums, labia, gonads—all of these vary in size and shape and morphology. However, in human cultures, sex categories generally get simplified into male or female, sometimes with "intersex", "third sex" or "two spirit" included as an "other" option.

So nature doesn't determine where the category of "male" ends and the category of "intersex" begins, or where the category of "intersex" ends and the category of "female" begins. Humans decide. Humans (typically doctors today, at least in Western cultures) decide how small a penis has to be, or how unusual a combination of parts has to be, before it counts as intersex. Humans decide whether a person with XXY chromosomes or XY chromosomes and androgen insensitivity will count as intersex. Doctors' opinions about what should count as intersex vary substantially. Some think you have to have "ambiguous genitalia" to count as intersex, even if your inside is mostly of one sex and your outside is mostly of another. Some think your brain has to be exposed to an unusual mix of hormones prenatally to count as intersex—so that even if you're born with atypical genitalia, you're not intersex unless your brain experienced atypical development. And some think you have to have both ovarian and testicular tissue (be a true hermaphrodite- very rare!) to count as intersex. Intersex and transgender (having a gender identity that does not match your assigned birth sex) are not interchangeable terms; many transgender people have no intersex traits, and many intersex people do not consider themselves transgender.

According to the Intersex Society of North America, it is up to the individual to decide if they "count" as intersex. However, since some forms of intersex signal underlying metabolic concerns, a person who thinks they might be intersex should seek a diagnosis and find out if they need professional healthcare. Everyone, regardless of their sexual anatomy, should be free of shame, secrecy, and unwanted genital surgeries, even if someone else believes they have non-standard sexual anatomy. It used to be standard practice to surgically alter any newborn who didn't quite fit to the typical "male" or "female" body pattern, which often meant attempting to make the child look more like a typical girl (Figure 13.3.7). Although that is much easier surgically, gender identity confusion, loss of sexual sensation, and possibly unsatisfactory future sexual relationships are unfortunate possible outcomes. Current recommendations are to simply leave the child alone and see where development takes them, being prepared for any possible combinations of gender expression and sexual orientation. If desired, the individual can choose genital reconstruction surgery for themselves when they are old enough to understand and consent to the procedure.

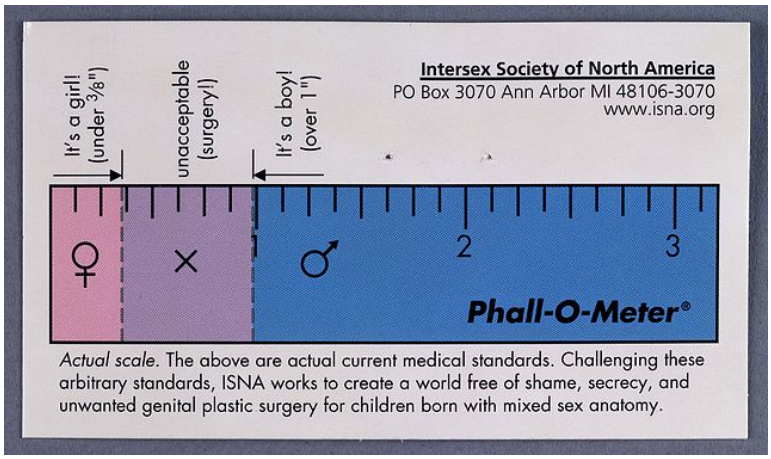


Figure **13.3.7**: Phall-O-Meter, a ruler used to determine when to perform genital surgery on a newborn baby.

13.3.3.2 Anomalous Female Sexual Differentiation

By studying individuals who do not neatly fall into the dichotic boxes of female or male and for whom the process of sexual differentiation is atypical, behavioral endocrinologists glean hints about the process of typical sexual differentiation. Prenatal exposure to androgens is the most common cause of anomalous sexual differentiation among females. The source of androgen may be internal (e.g., secreted by the adrenal glands- endocrine glands located on top of the kidneys) or external (e.g., exposure to environmental estrogens). One case of prenatal exposure to internal androgens is congenital adrenal hyperplasia (CAH). CAH develops when an enzyme needed to produce cortisol is defective or missing, resulting in abnormal hormonal feedback which leads to excessive production of androgens by the adrenal cortex. In a female (XX) fetus, the elevated androgen levels caused by CAH result in varying degrees of masculinization of the external genitalia. As a result, the baby's sex may appear ambiguous or may even be mistaken as male. CAH is the most common cause of intersex conditions.

Although the effects of CAH on anatomical development are more noticeable in biological females, males can also inherit CAH. If a male (XY) fetus has CAH, his prenatal body development is similar to other males. However, he is likely to show precocious (early) sexual development as a result of the excessive androgens, such as enlargement of the penis and facial hair as a toddler (Jones & Lopez, 2006).

13.3.3.3 Anomalous Male Sexual Differentiation

Female mammals are considered the “neutral” sex; additional physiological steps are required for male differentiation, and more steps bring more possibilities for errors in differentiation. One example of male anomalous sexual differentiation (seen in Dominican Republic populations) is Guevedoces syndrome ("eggs at twelve") or 5α-reductase deficiency. Individuals with 5α-reductase deficiency are genetic males (XY) with testes that produce testosterone and cells that respond to androgens. However, because they lack the enzyme 5α-reductase, which converts testosterone to dihydrotestosterone, they are born with ambiguous genitalia and are raised as females. As mentioned in Section 13.2, dihydrotestosterone is required for typical prenatal development of the penis and scrotum. The high levels of testosterone occurring during puberty result in recognizable masculinization of the body. The "girls" become boys in dress and behavior and generally have heterosexual orientations (Jones & Lopez, 2006). In other words, these individuals readily transition to a masculine gender identity and are attracted to women.

Another example of male anomalous sexual differentiation is androgen insensitivity syndrome (AIS). Individuals with AIS lack receptors for androgens and develop as females. Androgen insensitivity syndrome is an example of an endocrine disorder where an endocrine gland secretes a typical amount of hormone, but target cells do not respond normally to it. Individuals with AIS are genetically male and have an X and a Y chromosome, but they develop and are raised as females. This is due to a mutation in the androgen receptor gene which is located on the X chromosome. The androgen hormone testosterone normally causes the testes to descend and typical male characteristics to develop.

People with AIS have the external sex characteristics of females; they are typically raised as females and have a female gender identity. Affected individuals have male gonads (testes) that are undescended, which means they are located in the pelvis or abdomen (instead of inside a scrotum hanging between the legs). However, these individuals have neither female nor male internal reproductive structures. During prenatal development, the testes produced Mullerian inhibiting substance, which caused the duct

system that would form female internal sex organs to degenerate. Thus, these individuals do not have a uterus and do not menstruate. Without ovaries, Fallopian tubes, or a uterus, they are infertile and thus unable to conceive a child. Since the Wolffian ducts (that form male internal sex organs) require testosterone to develop, they also degenerate prenatally. The testes did (and do) produce testosterone, but since the body cells lack receptors for androgens, male development does not occur.

13.3.4 Conclusion: Variations in Biological Sex

"Although male and female are the most typical biologically ordained poles of sexual identity, a vast number of gradations can be produced by normally occurring variations in the underlying hormonal control mechanisms that guide gender differentiation" (Panksepp, 2004, page 234). In humans, intersex individuals make up about two percent—more than 150 million people—of the world's population (Blackless et al., 2000). There are dozens of conditions that can lead to intersex anatomical variations, such as Androgen Insensitivity Syndrome and Turner's Syndrome (Lee et al., 2006). The term "syndrome" can be misleading; although intersex individuals may have physical limitations (e.g., about a third of Turner's individuals have heart defects; Matura et al., 2007), they otherwise lead relatively normal intellectual, personal, and social lives. In any case, intersex individuals demonstrate the diverse variations of biological sex.

13.3.5 The Case of David Reimer

In August of 1965, Janet and Ronald Reimer of Winnipeg, Canada, welcomed the birth of their twin sons, Bruce and Brian. Within a few months, the twins were experiencing urinary problems; doctors recommended the problems could be alleviated by having the boys circumcised. A malfunction of the medical equipment used to perform the circumcision resulted in Bruce's penis being irreparably damaged. Distraught, Janet and Ronald looked to expert advice on what to do with their baby boy. By happenstance, the couple became aware of Dr. John Money at Johns Hopkins University and his theory of psychosexual neutrality (Colapinto, 2000).

Dr. Money had spent a considerable amount of time researching transgendered individuals and individuals born with ambiguous genitalia. As a result of this work, he developed a theory of psychosexual neutrality. His theory asserted that we are essentially neutral at birth with regard to our gender identity and that we don't assume a concrete gender identity until we begin to master language. Furthermore, Dr. Money believed that the way in which we are socialized in early life is ultimately much more important than our biology in determining our gender identity (Money, 1962).

Dr. Money encouraged Janet and Ronald to bring the twins to Johns Hopkins University, and he convinced them that they should raise Bruce as a girl. Left with few other options at the time, Janet and Ronald agreed to have Bruce's testicles removed and to raise him as a girl. When they returned home to Canada, they brought with them Brian and his "sister," Brenda, along with specific instructions to never reveal to Brenda that she had been born a boy (Colapinto, 2000).

Early on, Dr. Money shared with the scientific community the great success of this natural experiment that seemed to fully support his theory of psychosexual neutrality (Money, 1975). Indeed, in early interviews with the children it appeared that Brenda was a typical little girl who liked to play with "girly" toys and do "girly" things.

However, Dr. Money was less than forthcoming with information that seemed to argue against the success of the case. In reality, Brenda's parents were constantly concerned that their "daughter" wasn't really behaving as most girls did, and by the time Brenda was nearing adolescence, it was painfully obvious to the family that Brenda was really having a hard time identifying as a female. In addition, Brenda was becoming increasingly reluctant to continue the visits with Dr. Money to the point that Brenda threatened suicide if she ever saw him again.

At that point, Janet and Ronald disclosed the true nature of Brenda's early childhood to their children. While initially shocked, Brenda reported that things made more sense now, and ultimately, decided to identify as a male, choosing the name David Reimer (after the biblical story of David and Goliath, the small boy who defeated a giant). Brian did not take the news as well and distanced himself from his twin brother.

David was quite comfortable in his masculine role. He made new friends and began to think about his future. Although his castration had left him infertile, he still wanted to be a father. In 1990, David married a single mother and loved his new role as a husband and father. In 1997, David was made aware that Dr. Money was continuing to publicize his case as a success supporting his theory of psychosexual neutrality. This prompted David and his brother Brian to go public with their experiences in an attempt to discredit the doctor's publications. While this revelation created a firestorm in the scientific community for Dr. Money, it also triggered a series of unfortunate events that ultimately led to David committing suicide in 2004 (O'Connell, 2004).

This sad story speaks to the complexities involved in gender identity. While the Reimer case had earlier been paraded as a hallmark of how socialization trumped biology in terms of gender identity, the truth of the story made the scientific and medical communities more cautious in dealing with cases that involve intersex children and how to deal with their unique circumstances. In fact, stories like this one have prompted measures to prevent unnecessary harm and suffering to children who might have issues with gender identity. For example, in 2013, a law took effect in Germany allowing parents of intersex children to classify their children as indeterminate so that children can self-assign the appropriate gender once they have fully developed their own gender identities (Paramaguru, 2013). In 2017, California became the first US State to add a non-binary option to birth certificates and several other states have followed in recent years.

13.3.6 Summary

Although the scientific literature regarding sexual development that differs from typical XX/male and XY/female formats refers to these conditions as "Disorders of Sexual Development (DSD)", we use "Differences of Sexual Development" in this chapter. It is called a chromosomal abnormality (or aneuploidy) when an individual inherits too many or two few chromosomes, typically as a result of an error during gamete (egg or sperm) formation in one (or both) of the parents. The most common aneuploidies of the sex chromosomes (X and Y) include Triple X syndrome (XXX), Supermale syndrome (XYY), Klinefelter syndrome (XXY), and Turner syndrome (XO). (A zygote that inherits only one Y chromosome and no X chromosome is spontaneously aborted.) While individuals with Klinefelter syndrome and Turner syndrome usually have visible anatomical differences and are often infertile, individuals with Triple X syndrome and Supermale syndrome may have no noticeable symptoms and usually have normal fertility. Another chromosomal alteration that affects sexual development is the transfer of the SRY gene from the Y chromosome to the X chromosome. A zygote with a Y chromosome that has lost its SRY gene will develop in the "default" female pattern (since the gonads will not develop as testes without an SRY gene), and a zygote with an X chromosome that has the SRY gene attached to it will develop in a male pattern, but will not be fertile.

When an individual develops biological characteristics that are in between typical male and typical female structures during prenatal development, they are said to have an intersex condition. There are many anatomical and physiological variations which fit under this umbrella term, and there is no current consensus on what "counts" as intersex and what does not. The Intersex Society of North America states that it is up to the individual to decide if they "count" as intersex, and that everyone, regardless of their sexual anatomy, should be free of shame, secrecy, and unwanted genital surgeries. While immediate surgery to try to create a more typical (usually female) appearance used to be common practice, current recommendations are to simply leave the intersex child alone and see where development takes them, being prepared for any possible combinations of gender expression and sexual orientation. Although intersex individuals may have physical limitations, they otherwise lead relatively normal intellectual, personal, and social lives.

Prenatal exposure to androgens is the most common cause of anomalous sexual differentiation among females. Congenital adrenal hyperplasia (CAH) develops when excessive production of androgens by the adrenal cortex occurs due to a missing enzyme. In a female (XX) fetus, the elevated androgen levels result in varying degrees of masculinization of the external genitalia. As a result, the baby's sex may appear ambiguous or may even be mistaken as male. CAH is the most common cause of intersex conditions.

Androgen insensitivity syndrome (AIS) and 5α-reductase deficiency are the most common causes of anomalous male (XY) sexual differentiation. People with AIS have the external sex characteristics of females and they are typically raised as females and have a female gender identity. However, they are infertile because they do not have female internal structures and their gonads are testes rather than ovaries. Although their testes produce testosterone, their body cells lack testosterone receptors and thus do not respond accordingly. Individuals with 5α-reductase deficiency are born with ambiguous genitalia and are raised as females. During puberty, high levels of testosterone result in recognizable masculinization of the body, and these individuals usually transition to male gender identities and are generally attracted to women.

David Reimer was one of a pair of identical twin boys born in Canada in August 1965 (birth name Bruce). As a result of a tragic circumcision accident in infancy, his penis was destroyed, and he was raised as "Brenda" under the advice of Dr. John Money, who erroneously believed that gender identity was exclusively determined by experiences during childhood development. By early adolescence, it was clear that Brenda was having difficulties identifying as a girl, and the truth was revealed to Brenda and "her" twin brother Brian by their parents. Brenda returned to a masculine gender identity, choosing the name David, and initially his life was going well. In the early 2000s, after discovering that Dr. Money was still touting the "experiment" as a success, the Reimer

family went public with their story. Ultimately a number of unfortunate events (including the death of Brian due to a drug overdose) led to David's suicide in 2004. The truth of this story (which had originally been used to support decisions to reassign the gender of intersex children at birth) made the scientific and medical communities more cautious in dealing with cases that involve intersex children and how to deal with their unique circumstances.

13.3.7 References

Note: These references are specifically those added to the content of this page by Naomi Bahm and do not include citations from the outside sources used. Refer to the text attributions to locate citations for articles from other sources.

Jones, R.E. & Lopez, K.H. (2006) *Human Reproductive Biology, 3rd edition*. Academic Press.

Panksepp, J. (2004). *Affective neuroscience: The foundations of human and animal emotions*; *Chapter 12: The varieties of love and lust: Neural control of sexuality*. Oxford University Press.

13.3.8 Additional Resources

Intersex- Video link from Wakim & Grewal textbook: https://youtu.be/jhHGCvMlrb0

Websites for specific conditions:

- XO: https://rarediseases.org/rare-diseas...rner-syndrome/
- XXY: https://rarediseases.org/rare-diseas...lter-syndrome/
- XXX: https://rarediseases.org/rare-diseases/trisomy-x/
- XYY:
  - https://rarediseases.org/rare-diseases/xyy-syndrome/
  - http://www.pathguy.com/xyy.html
  - https://healthjade.net/xyy-syndrome/
  - https://www.rarechromo.org/media/inf...XYY%20FTNW.pdf
- AIS: https://rarediseases.org/gard-rare-d...vity-syndrome/
- CAH: https://rarediseases.org/rare-diseas...l-hyperplasia/
- Guevodoces:
  - https://rarediseases.org/gard-rare-d...se-deficiency/
  - https://www.bbc.com/news/magazine-34290981

Intersex Society of North America: https://isna.org/

13.3.9 Attributions

1. Figures:
   1. Chromosomes by Mariana Ruiz Villarreal (LadyofHats), CC BY-NC 3.0, for CK-12
   2. XXXSyndromeB By Nicole R Tartaglia, Susan Howell, Ashley Sutherland, Rebecca Wilson, and Lennie Wilson - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2883963/, CC BY 2.5, https://commons.wikimedia.org/w/inde...curid=5022965
   3. A girl with Turner's syndrome by Sküskü15, licensed CC BY-SA 4.0 via Wikimedia Commons
   4. Individuals of Asian descent with Turner's syndrome by Paul Kruszka, Yonit A Addissie, Cedrik Tekendo-Ngongang, Kelly L Jones, Sarah K Savage, Neerja Gupta, Nirmala D Sirisena, Vajira H W Dissanayake, C Sampath Paththinige, Teresa Aravena, Sheela Nampoothiri, Dhanya Yesodharan, Katta M Girisha, Siddaramappa Jagdish Patil, Saumya Shekhar Jamuar, Jasmine Chew-Yin Goh, Agustini Utari, Nydia Sihombing, Rupesh Mishra, Neer Shoba Chitrakar, Brenda C Iriele, Ezana Lulseged, Andre Megarbane, Annette Uwineza, Elizabeth Eberechi Oyenusi, Oluwarotimi Bolaji Olopade, Olufemi Adetola Fasanmade, Milagros M Duenas-Roque, Meow-Keong Thong, Joanna Y L Tung, Gary T K Mok, Nicole Fleischer, Godfrey M Rwegerera, María Beatriz de Herreros, Johnathan Watts, Karen Fieggen, Victoria Huckstadt, Angélica Moresco, María Gabriela Obregon, Dalia Farouk Hussen, Neveen A Ashaat, Engy A Ashaat, Brian H Y Chung, Eben Badoe, Sultana M H Faradz, Mona O El Ruby, Vorasuk Shotelersuk, Ambroise Wonkam, Ekanem Nsikak Ekure, Shubha R Phadke, Antonio Richieri-Costa, Maximilian Muenke, is in the public domain via Wikimedia Commons

5. [Left image] The symptoms of Klinefelter's syndrome in a human male By http://smithperiod6.wikispaces.com/K...ter's+Syndrome, CC BY-SA 3.0, https://commons.wikimedia.org/w/inde...curid=25557422 AND [right image] Klinefelter's syndrome By CDL69 - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/inde...curid=18683822
6. Wax human hermaphrodite genitals by Daniel Ullrich, Threedots, CC BY-SA 3.0, via Wikimedia Commons
7. Phall-O-Meter (Intersex Society of North America) by Wellcome Images, Creative Commons Attribution 4.0 International, via Wikimedia Commons

2. Text adapted from:
    1. "Genetics of Inheritance" by Suzanne Wakim & Mandeep Grewal, LibreTexts is licensed under CC BY-NC.
    2. " Chromosomal Abnormalities" by Martha Lally & Suzanne Valentine-French, LibreTexts is licensed under CC BY-NC-SA. From textbook Lifespan Development: A Psychological Perspective by Martha Lally and Suzanne Valentine-French is licensed under CC BY-NC-SA 3.0
    3. " Heredity" by Paris, Ricardo, Raymond, & Johnson, LibreTexts is licensed under CC BY./@go/page/10183
    4. Human Sexual Anatomy and Physiology by Don Lucas and Jennifer Fox, licensed CC BY-NC-SA 4.0 via Noba Project.
    5. The Psychology of Human Sexuality by Don Lucas and Jennifer Fox, licensed CC BY-NC-SA 4.0 via Noba Project.
    6. Intersex Conditions section (somewhat modified): "The Interplay of Sex and Gender" by LibreTexts is licensed under notset.
    7. CAH (partial): "Genetics Teacher Preparation Notes" by LibreTexts is licensed under CC BY-NC.
    8. Hormones & Behavior by Randy J. Nelson, licensed CC BY-NC-SA 4.0 via Noba Project.
    9. "Introduction to the Endocrine System" by Suzanne Wakim & Mandeep Grewal, LibreTexts is licensed under CC BY-NC.
    10. David Reimer story, a few phrases in Intersex Conditions: 10.3 Sexual Behavior (Psychology 2e) by Rose M. Spielman, William J. Jenkins, and Marilyn D. Lovett, licensed CC BY 4.0 via OpenStax.
3. Changes: Text (and some images) from above ten sources pieced together with some modifications, transitions and additional content (particularly much of the Aneuploidy and Other Chromosomal Alterations sections) added by Naomi I. Gribneau Bahm, PhD., Psychology Professor at Cosumnes River College, Sacramento, CA.

This page titled 13.3: Differences of Sexual Development is shared under a mixed license and was authored, remixed, and/or curated by Naomi Bahm (ASCCC Open Educational Resources Initiative (OERI)) .

- **2.1: Heredity** by Paris, Ricardo, Rymond, & Johnson is licensed CC BY 4.0. Original source: https://open.umn.edu/opentextbooks/textbooks/child-growth-and-development.
- **12.2: Introduction to the Endocrine System** by Suzanne Wakim & Mandeep Grewal is licensed CK-12. Original source: https://www.ck12.org/book/ck-12-human-biology/.

13.4: Gender

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Distinguish between sex, gender, and sexual orientation
2. Describe gender variations recognized in some cultures

13.4.1 Overview

This section begins by explaining the difference between sex, gender, and sexual orientation. The terms cisgender and transgender are introduced, and gender variations across cultures are explored. Finally, the role of hormones in transgender treatment and the scientific and societal understanding of gender variations is addressed.

13.4.2 Sex, Gender, and Sexual Orientation: Three Different Aspects of You

Applying for a credit card or filling out a job application requires your name, address, and often birth-date. Additionally, applications usually ask for your sex or gender. The terms “sex” and “gender” are commonly used interchangeably. However, these terms are distinct from one another. Sex describes means of biological reproduction. Sex includes sexual organs, such as ovaries- defining biological reproduction as a female- or testes- defining biological reproduction as a male. As detailed in the section on differences of sexual development, biological sex is not as easily defined or determined as you might expect.

By contrast, the term gender describes psychological (gender identity) and sociological (gender role) representations of biological sex. Sex and gender are important aspects of a person’s identity. However, they do not inform us about a person’s sexual orientation (Rule & Ambady, 2008). Sexual orientation refers to a person’s sexual attraction (or lack thereof) to others. Within the context of sexual orientation, sexual attraction refers to a person’s capacity to arouse the sexual interest of another, or, conversely, the sexual interest one person feels toward another.

We live in an era when sex, gender, and sexual orientation are controversial religious and political issues. Some nations have laws against homosexuality, while others have laws protecting same-sex marriages. At a time when there seems to be little agreement among religious and political groups, it makes sense to wonder, “What is normal?” and, “Who decides?”

The international scientific and medical communities (e.g., World Health Organization, World Medical Association, World Psychiatric Association, Association for Psychological Science) view variations of sex, gender, and sexual orientation as normal. Furthermore, variations of sex and the orientation of sexual behavior occur naturally throughout the animal kingdom. More than 65,000 animal species have intersex individuals (Figure **13.4.1**), born with either an absence or some combination of male and female reproductive organs, sex hormones, or sex chromosomes (Jarne & Auld, 2006). More than 500 animal species engage in homosexual or bisexual behaviors (Lehrer, 2006), such as the Mallard ducks shown in Figure **13.4.2**.



Figure **13.4.1**: A rare intersexed two-spotted bumble bee.



Figure **13.4.2**: A male Mallard duck couple.

13.4.3 Gender

At an early age, we begin learning cultural norms for what is considered masculine and feminine. For example, American children may associate long hair or dresses with femininity. Later in life, as adults, we often conform to these norms by behaving in gender-specific ways: as men, we build houses; as women, we bake cookies (Marshall, 1989; Money et al., 1955; Weinraub et al., 1984).

Because cultures change over time, so too do ideas about gender. For example, European and American cultures today associate pink with femininity and blue with masculinity. However, less than a century ago, these same cultures were swaddling baby boys in pink, because of its masculine associations with “blood and war,” and dressing little girls in blue, because of its feminine associations with the Virgin Mary (Kimmel, 1996).

13.4.3.1 Variations in Gender

Just as biological sex varies more widely than is commonly thought, so too does gender. Cisgender individuals’ gender identities correspond with their birth (biological) sexes- biological males with masculine gender identities and biological females with feminine gender identities. In contrast, transgender individuals’ gender identities do not correspond with their birth (biological) sexes. The Latin prefix "cis" means "on the same side"; the prefix "trans" means "across". Many cisgender people do not self-identify as such. As with transgender people, the term or usage of cisgender does not indicate a person's sexual orientation or gender role expression (TSER, 2022). As mentioned in the section on intersex conditions (people whose sex characteristics or anatomy differ from typical development), intersex and transgender are not interchangeable terms; many transgender people have no intersex traits, and many intersex people do not consider themselves transgender. Because gender is so deeply ingrained culturally, rates of transgender individuals vary widely around the world (Table **13.4.1**).

Nation	Transgender people per 100,000
Sweden	.17
Poland	.26
Ireland	1.4
Japan	1.4
India	167
Thailand	333
United States	476
Malaysia	1,333

Table **13.4.1**: Nations vary in the number of transgender people found in their populations (De Gascun et al., 2006; Dulko & Imielinskia, 2004; Landen et al., 1996; Okabe et al., 2008, Conron et al., 2012; Winter, 2009).

13.4.3.2 Cultural Recognition of Gender Variations

Some cultures formally recognize the presence of people who do not conform to an expectation of biological sex matching gender identity. For example, "[some Native American tribes] believed that in addition to the prevailing variants of *man within man* and *woman within woman*, nature sometimes created a *man's mind within the body of a woman* and a *woman's mind within the body of a man*" (Panksepp, 2004, page 232). Transgender women are referred to as *fa'afafine* in the Samoan population (prevalence of five percent, Tan, 2016); *muxes* in Oaxaca, Mexico (as many as six percent of biological males, Stephen, 2002); *kathoey* (Figure **13.4.3**) in Thailand; and *hijras* (Figure **13.4.3**) in Pakistan, India, Nepal, and Bangladesh. Hijiras are recognized by their governments as a third gender (Pasquesoone, 2014).



Figure **13.4.3**: Hjira dancer in Nepal and kathoey performers at a cabaret in Pattaya, Thailand.

Although incidence rates of transgender individuals differ significantly between cultures, transgender women (TGW)- whose birth sex was male- are by far the most frequent type of transgender individuals worldwide. Of the 18 countries studied by Meier and Labuski (2013), 16 of them had higher rates of TGW than transgender men (TGM)- whose birth sex was female- and the 18 country TGW to TGM ratio was 3 to 1.

However, the ratio of TGW to TGM individuals may be changing. A 2015 meta-analysis found that the overall incidence of transgender was 4.6 per 100,000 individuals, comprised of 6.8 TGW and 2.6 TGM, numbers that corroborate earlier reports of significantly higher rates of TGW than TGM individuals. A time analysis of the studies included revealed that the overall

prevalence of transgender individuals had increased over the past 50 years (Arcelus et al., 2015). Echoing this trend, Leinung and Joseph (2020) reported an increase in individuals seeking hormone therapy in the past 25 years, and also indicated that the number of people requesting hormone therapy specifically for TGM transitions had increased steadily since 1990, resulting in equivalent numbers of TGW and TGM individuals being treated by 2017. It is unclear whether or how much this change reflects greater acceptance, destigmatization, and decreasing barriers to care specifically for TGM individuals (as the authors suggest), or if it may also represent a true increase in the number of TGM individuals in contemporary populations.

13.4.3.3 Role of Sex Hormones in Transgender Treatment

Feminizing or masculinizing hormone therapy is the administration of exogenous endocrine agents to induce changes in physical appearance. Hormone therapy is inexpensive relative to surgery, and highly effective in the development of some secondary sex characteristics, such as facial and body hair in transgender men (TGM, sometimes called female-to-male [FTM] individuals) or breast development in transgender women (TGW, sometimes called male-to-female [MTF] individuals). Thus, hormone therapy is often the first (and sometimes only) medical gender affirmation intervention accessed by transgender individuals looking to develop masculine or feminine characteristics consistent with their gender identity. In some cases, hormone therapy may be required before surgical interventions can be conducted. Transgender women are prescribed estrogen and anti-testosterone medication (such as cyproterone acetate and spironolactone). Transgender men are prescribed testosterone.

13.4.3.4 Scientific and Societal Understanding of Gender Variations

Hormones that masculinize the brain in chromosomally male (XY) individuals are distinct from those that lead to the development of a typically male body form. "Due to this branching of control factors for brain and body organization, it is quite possible for a male-type body to contain a female-type brain, and for a female-type body to contain a male-type brain" (Panksepp, 2004, page 225). These developmental deviations could influence an individual's ultimate sexual orientation and/or gender identity. An example of conflicting brain and body organization comes from a study which found that the bed nucleus of the stria terminalis (BNST, an area located near the hypothalamus that is known to be important for sexual behavior) was "female-like" in size (smaller than cisgender males and similar to cisgender females) in the brains of six male-to-female transsexuals. This difference was not correlated with either adult hormone levels or sexual orientation (Zhou et al., 1995). Given that studies requiring post-mortem brain tissue to examine tiny brain areas are very difficult to conduct, this type of information is scarce.

Nonetheless, our scientific knowledge and general understanding of gender identity continue to evolve, and young people today have more opportunity to explore and openly express different ideas about what gender means than previous generations. Recent studies indicate that a majority of Millennials (birth years 1981-1996) regard gender as a spectrum instead of a strict male/female binary, and that 12% identify as transgender or gender non-conforming. Additionally, more people know others who use gender-neutral pronouns, such as they/them (Kennedy, 2017). This change in language may indicate that Millennials and Generation Z people (birth years 1997-2012) understand the experience of gender itself differently. As young people lead this change, other changes are emerging in a range of spheres, from public bathroom policies to retail organizations. For example, some retailers are starting to change traditional gender-based marketing of products, such as removing “pink and blue” clothing and toy aisles. Despite these changes, those who exist outside of traditional gender norms face difficult challenges. Even people who vary slightly from traditional norms can be the target of discrimination and sometimes even violence.

13.4.4 Summary

Although often used interchangeably, the terms sex and gender refer to distinct concepts, and sexual orientation is a third aspect of an individual's characteristics. Sex refers to the structures relevant for biological reproduction (such as ovaries in females and testes in males), whereas gender refers to an individual's sense of being masculine, feminine, neither or both. Gender identity is the psychological representation of biological sex, and gender role is the sociological representation. Sexual orientation describes a person's sexual attraction (or lack thereof) to others. Variations in each of these (sex, gender, and sexual orientation) is considered normal by international scientific and medical communities.

Children begin to learn cultural norms for masculine and feminine behaviors very early, but cultures vary widely and even the same culture changes over time- thus, ideas about gender also change. Cisgender individuals’ gender identities correspond with their birth (biological) sexes, while transgender individuals’ gender identities do not correspond with their birth (biological) sexes. Intersex (people whose sex characteristics or anatomy differ from typical development) and transgender are not interchangeable

terms; many transgender people have no intersex traits, and many intersex people do not consider themselves transgender. Rates of transgender individuals vary widely around the world. Some cultures formally recognize the presence of people who do not conform to an expectation of biological sex matching gender identity, such as some Native American tribes. Transgender women specifically are referred to as *fa'afafine* in Samoa; *muxes in* Oaxaca, Mexico; *kathoey* in Thailand; and *hijras* in Pakistan, India, Nepal, and Bangladesh.

Worldwide, transgender women (whose birth sex was male) have been by far the most frequent type of transgender individuals, with a ratio of about three transgender women to one transgender man. However, recent trends (at least in people requesting hormone therapy) have progressively changed since the 1990s, such that by 2017, an equivalent number of transgender women and transgender men were being treated. Hormone therapy is used to induce changes in physical appearance. It is inexpensive relative to surgery, and highly effective in the development of some secondary sex characteristics, such as facial and body hair in transgender men, and breast development in transgender women. Hormone therapy is often the first (and sometimes only) medical gender affirmation intervention accessed by transgender individuals.

During prenatal development, hormones that masculinize the brain are distinct from those that lead to the development of a typically male body form. Developmental deviations can occur (such that the body is masculinized but the brain is not, for example), and could in turn influence an individual's ultimate sexual orientation and/or gender identity. Our scientific knowledge and general understanding of gender identity continue to evolve. Recent studies indicate that a majority of Millennials (birth years 1981-1996) regard gender as a spectrum instead of a strict male/female binary, and that 12% identify as transgender or gender non-conforming. Nonetheless, individuals who exist outside of traditional gender norms face difficult challenges, and may be the target of discrimination and even violence.

13.4.5 References

Note: These references are specifically those added to the content of this page by Naomi Bahm and do not include citations from the outside sources used. Refer to the text attributions to locate citations for articles from other sources.

Arcelus, J., Bouman, W.P., Van Den Noortgate, W., Claes, L., Witcomb, G., & Fernández Aranda, F. (2015) Systematic review and meta-analysis of prevalence studies in transsexualism. *European Psychiatry, 30*, 807–815.

Dimock, M. (2019). Defining generations: Where Millennials end and Generation Z begins. Pew Research Center (January 17), https://pewrsr.ch/2szqtJz

Leinung, M.C. & Joseph, J. (2020) Changing demographics in transgender individuals seeking hormonal therapy: Are trans women more common than trans men? *Transgender Health, 5*(4): 241-245. doi: 10.1089/trgh.2019.0070

Panksepp, J. (2004). *Affective neuroscience: The foundations of human and animal emotions*; *Chapter 12: The varieties of love and lust: Neural control of sexuality*. Oxford University Press.

TSER- Trans Student Educational Resources (2022). Definitions. https://transstudent.org/about/definitions/ (accessed 5/12/2022)

Zhou, J.-N., Hofman, M., Gooren, L.J.G., & Swaab, D.F. (1995). A sex difference in the human brain and its relation to transsexuality. *Nature, 378*, 68-70.

13.4.6 Additional Resources

Trans Student Educational Resources: https://transstudent.org/

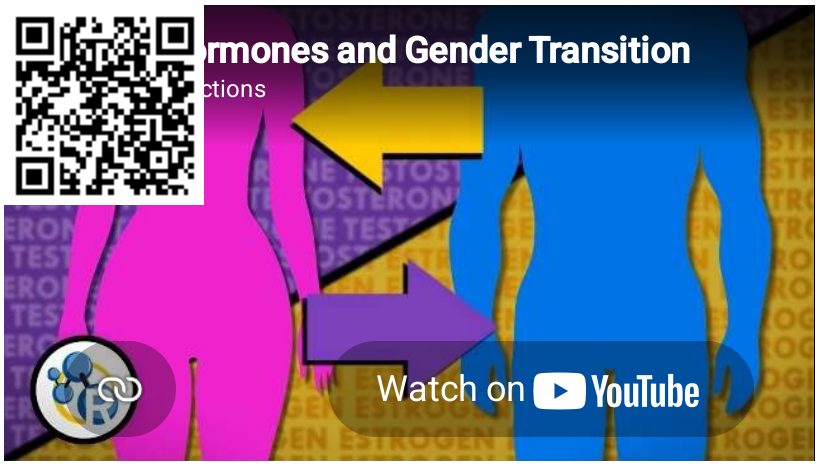
Guide to non-binary birth certificates and US State IDs: https://www.usbirthcertificates.com/articles/gender-neutral-birth-certificates-states

APA's non-binary (gender identity) fact sheet: https://www.apadivisions.org/divisio...fact-sheet.pdf

Educational graphics for understanding sex, gender, sexual orientation, and more: https://www.itspronouncedmetrosexual.com/2018/10/the-genderbread-person-v4/

Free e-book "A Guide to Gender: The Social Justice Advocate's Handbook" by Sam Killermann: https://impetus.gumroad.com/l/g2g2

People's sense of gender identity does not always match their anatomy. Some people do not identify as either male or female, and instead, they identify as non-binary, or genderqueer. Others may identify as a gender that is the opposite of what is typically associated with their chromosomes or reproductive organs. These people are called transgender, and they may choose to transition to the opposite gender, a process that may or may not involve physical modifications. Watch the video below to learn about the use of hormones in gender transitioning.



Muxes, a documentary about Mexican children identified as male at birth, but who choose at a young age to be raised as female.



13.4.7 Attributions

1. Figures:
    1. Intersex Bombus bimaculatus (two-spotted bumble bee) by USGS Bee Inventory and Monitoring Lab, public domain
    2. Couple of male Mallard ducks by Norbert Nagel, licensed CC BY-SA 3.0 via Wikimedia Commons
    3. Table 1 by Lucas & Fox, licensed CC BY-NC-SA 4.0 via NOBA module The Psychology of Human Sexuality
    4. Left image: Hjira dancer in Nepal by Adam Jones, licensed CC BY 2.0; Right image: Pattaya transwomen 2 by Ohconfucius, public domain via Wikimedia Commons
2. Text adapted from:
    1. The Psychology of Human Sexuality by Don Lucas and Jennifer Fox, licensed CC BY-NC-SA 4.0 via Noba Project.
    2. " Development of Sexual Identity" by Paris, Ricardo, Raymond, & Johnson, LibreTexts is licensed under CC BY.
    3. Role of Sex Hormones in Transgender Treatment section: "Introduction to the Reproductive System" by Suzanne Wakim & Mandeep Grewal, LibreTexts is licensed under CK-12.
    4. Part of Variations in Gender section: 10.3 Sexual Behavior (Psychology 2e) by Rose M. Spielman, William J. Jenkins, and Marilyn D. Lovett, licensed CC BY 4.0 via OpenStax.
3. Changes:

1. Text (and images) from above four sources pieced together with some modifications, transitions and additional content (particularly in the sections Cultural Recognition of Gender Variations, and Scientific and Societal Understanding of Gender Variations) added by Naomi I. Gribneau Bahm, PhD., Psychology Professor at Cosumnes River College, Sacramento, CA.
2. Birth year ranges for Millennials and Generation Z from: https://www.pewresearch.org/fact-tan...tion-z-begins/ (accessed 5/12/22; see Dimock, 2019 in references)

This page titled 13.4: Gender is shared under a mixed license and was authored, remixed, and/or curated by Naomi Bahm (ASCCC Open Educational Resources Initiative (OERI)) .

- **22.2: Introduction to the Reproductive System** by Suzanne Wakim & Mandeep Grewal is licensed CK-12. Original source: https://www.ck12.org/book/ck-12-human-biology/.
- **15.8: Development of Sexual Identity** by Paris, Ricardo, Rymond, & Johnson is licensed CC BY 4.0. Original source: https://open.umn.edu/opentextbooks/textbooks/child-growth-and-development.

13.5: Sexual Orientation

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Understand Kinsey's continuum of sexual orientation and the range of sexual orientation labels used now
2. Explain the evidence for at least two nonsocial causes of sexual orientation

13.5.1 Overview

This section discusses sexual orientation, a person's emotional and sexual attraction to other individuals of a particular sex or gender. Kinsey's continuum of sexual orientation (a range from exclusively heterosexual through equally bisexual to exclusively homosexual) is introduced, as well as some additional terms that are in contemporary use (polysexual, pansexual, and asexual). The development and origins of sexual orientation are addressed, including research exploring genetics, prenatal hormone exposure, the fraternal-birth-order effect, and childhood development. Regardless of the root cause(s) for a person's sexual orientation (and/or gender identity), it is not a conscious choice and cannot be easily changed.

13.5.2 Sexual Orientation

A person's **sexual orientation** is their emotional and sexual attraction to a particular sex or gender, including a continuing pattern of romantic or sexual attraction (or a combination of these) to persons of a given sex or gender. According to the American Psychological Association (APA) (2016), sexual orientation also refers to a person's sense of identity based on those attractions, related behaviors, and membership in a community of others who share those attractions. Although a person's intimate behavior may have sexual fluidity- changing due to circumstances (Diamond, 2009)- sexual orientations are relatively stable over one's lifespan, and are influenced by genetics (Frankowski, 2004). Sexual orientation is distinct and independent from both biological sex and gender, as discussed in Section 13.4.

While some argue that sexual attraction is primarily driven by reproduction (e.g., Geary, 1998), empirical studies point to pleasure as the primary force behind our sex drive. For example, in a survey of college students who were asked, "Why do people have sex?" respondents gave more than 230 unique responses, most of which were related to pleasure rather than reproduction (Meston & Buss, 2007). Here's a thought-experiment to further demonstrate how reproduction has relatively little to do with driving sexual attraction: Add the number of times you've had and/or hope to have sex during your lifetime. With this number in mind, consider how many times the goal was (or will be) for reproduction versus how many it was (or will be) for pleasure. Which number is greater?



Figure **13.5.1**: First photo: "Victory!" - This lesbian couple has been waiting 31 years to be able to marry (Sacramento, CA). Second photo: Two men holding hands, Gay Pride Project (Baltimore, MD).

13.5.2.1 A Continuum of Sexual Orientation

Instead of thinking of sexual orientation as being two dichotomous categories- homosexual (attracted to the same sex, Figure **13.5.1**) and heterosexual (attracted to the opposite sex)- sexuality researcher Alfred Kinsey and colleagues argued that it is a continuum (Kinsey, Pomeroy, & Martin, 1948). They measured sexual orientation using a seven-point scale called the Heterosexual-Homosexual Rating Scale (Figure **13.5.2**), in which zero is exclusively heterosexual, three is bisexual (with equal attractions to the same and opposite sexes), and six is exclusively homosexual. Research done over several decades has supported this idea that sexual orientation ranges along a continuum, from exclusive attraction to the opposite sex/gender to exclusive attraction to the same sex/gender (Carroll, 2016).

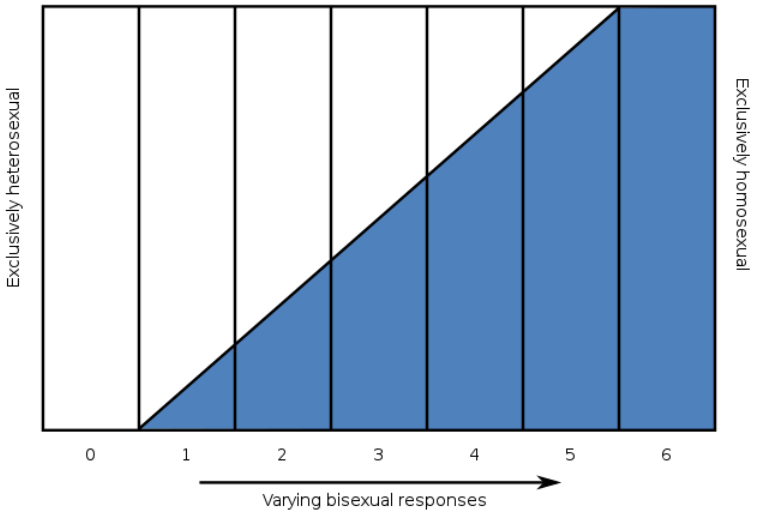


Figure **13.5.2**: The Kinsey scale, a tool which enables an individual to select a response on a scale between heterosexual and homosexual. 0= exclusively heterosexual, 1= predominantly heterosexual, only incidentally homosexual, 2= predominantly heterosexual, more than incidentally homosexual, 3= equally heterosexual and homosexual, 4= predominantly homosexual, more than incidentally heterosexual, 5= predominantly homosexual, only incidentally heterosexual, and 6= exclusively homosexual.

The commonly stated notion that "10% of people are homosexual" originates (erroneously) from Kinsey's 1948 study, which found that 10% of men reported being exclusively homosexual for at least three years during adulthood. In addition to the likelihood that Kinsey's sample overestimated the actual occurrence of homosexuality in the population, only 4% of the male respondents in his study reported being homosexual their entire lifetime. This figure is much closer to the 3.5% prevalence of non-heterosexual (gay/lesbian/bisexual) orientation found in more recent and representative studies of US and other Western populations (Bailey et al., 2016). However, it is important to note that how a respondent is asked to report their sexual orientation produces very different results. Researchers using the Kinsey Scale have found 18% to 39% of Europeans and Americans identifying as somewhere between heterosexual and homosexual (Lucas et al., 2017; YouGov.com, 2015). These percentages drop dramatically to only 0.5% to 1.9% non-heterosexual when researchers force individuals to respond using only two categories (Copen, Chandra, & Febo-Vazquez, 2016; Gates, 2011).

Although the percentage of US adults identifying as LGBT (lesbian/gay/bisexual/transgender) remained relatively stable through the early 2010s (and still remains stable in older-aged cohorts), there is evidence that younger people are more likely to identify as LGBT than previous generations, which is in turn increasing the overall prevalence in the population. Gates (2017, page 1221) reports that "the percentage of older age cohorts identifying as LGBT has remained stable or declined despite large increases among Millennials, who are now three times more likely than Baby Boomers to identify as LGBT (7.3% vs 2.4%)." (Millennials are people born during the years 1981 through 1996 and Baby Boomers are people born during the years 1946 through 1964 (Dimock, 2019).)

Indeed, even beyond a continuum from heterosexual to homosexual, sexual orientation is as diverse as gender identity. Some examples of sexual orientation include **heterosexuality** (attraction to the opposite sex/gender), **homosexuality** (attraction to the same sex/gender, also referred to as **same-sex attraction**- some people find the term homosexuality offensive since it was previously classified as a mental illness), **bisexuality** (attraction to two sexes/genders), **polysexuality** (attraction to multiple sexes/genders), **pansexuality** (attraction to all sexes/genders), and **asexuality** (no sexual attraction to any sex/gender).

13.5.2.2 Development and Origins of Sexual Orientation

According to current scientific understanding, individuals are usually aware of their sexual orientation between middle childhood and early adolescence. However, this is not always the case, and some do not become aware of their sexual orientation until much later in life. It is not necessary to participate in sexual activity to be aware of these emotional, romantic, and physical attractions; people can be celibate (not participating in any type of sexual activity) and still recognize their sexual orientation. Some researchers argue that sexual orientation is not static and inborn, but is instead fluid and changeable throughout the lifespan. Regardless of when sexual orientation is recognized or how it is expressed, it is clearly not a conscious "choice" for either heterosexual or non-heterosexual individuals.

There is no scientific consensus regarding the exact reasons why an individual holds a particular sexual orientation. Research has examined possible biological, developmental, social, and cultural influences on sexual orientation, but there is no conclusive evidence that links sexual orientation to one specific factor (APA, 2016). In an extensive review of evidence for the causal mechanisms of human sexual orientation, Bailey et al. (2016, page 46) state that "no causal theory of sexual orientation has yet gained widespread support", further pointing out that the most probable scientific hypotheses are difficult to test. The most commonly suggested social causes of homosexuality, "sexual recruitment by homosexual adults, patterns of disordered parenting, or the influence of homosexual parents", generally have only weak support and are compromised by many confounding factors (Bailey et al., 2016, page 46). They conclude that "there is considerably more evidence supporting nonsocial causes of sexual orientation than social causes". Some of the prominent nonsocial causes that have been examined include genetics, prenatal hormone exposure, the fraternal-birth-order effect, and behavior differences during childhood development.

13.5.2.2.1 Genetics

One method of measuring the genetic roots of sexual orientation is concordance rates, which is the probability that a pair of individuals has the same sexual orientation. If both twins have the same sexual orientation, they are "concordant" for this trait. Concordance rates are calculated and compared between people who share the same genetics (monozygotic twins, 99%), some of the same genetics (dizygotic twins, 50%, and siblings, 50%), and non-related people (randomly selected from the population). Researchers find that concordance for sexual orientation is highest for monozygotic twins; and concordance rates for dizygotic twins, siblings, and randomly-selected pairs are not significantly different (Bailey et al. 2016; Kendler et al., 2000). Since concordance rates are most similar for monozygotic twins, this indicates that "nature" (genetics) influences the expression of sexual orientation. However, monozygotic twins do not always have the same sexual orientation, so "nurture" (the environment and individual experiences) also plays a role in determining sexual orientation. Nonetheless, because sexual orientation is a hotly debated issue, an appreciation of the genetic aspects of attraction can be an important piece of this dialogue.

13.5.2.2.2 Prenatal Hormone Exposure

Excess or deficient exposure to hormones during prenatal development has also been theorized as an explanation for non-heterosexual orientation. One-third of females exposed to abnormal amounts of prenatal androgens due to congenital adrenal hyperplasia (CAH, see the section on differences of sexual development) identify as bisexual or lesbian (Cohen-Bendahan, van de Beek, & Berenbaum, 2005). In contrast, too little exposure to prenatal androgens may affect male sexual orientation (Carlson, 2011).

13.5.2.2.3 Fraternal-Birth-Order Effect

The "fraternal-birth-order effect" is the well-documented (and cross-culturally robust) finding that homosexual males tend to have more older brothers than heterosexual males (Bailey et al., 2016). Interestingly, this effect is particular to right-handed homosexual males (Blanchard et al., 2006; Bogaert, 2007). This effect is also specific to biological brothers (Figure **13.5.3**), meaning that the number of sons that have been gestated by the man's mother is the important factor, whether or not the man was raised with those brothers. "The effect is almost certainly causal, with each additional older brother causing an increase in the chances of a man's being homosexual. ... Assuming that a man without any older brothers has a 2% chance of being homosexual, a man with one older brother has a 2.6% chance; with two, three, and four older brothers, the chances are 3.5%, 4.6%, and 6.0%, respectively" (Bailey et al., 2016, page 79). One proposed mechanism for this is that the mother creates antibodies against proteins on her developing son's Y chromosome (a possibility which increases with each pregnancy of a son), and that those antibodies may prevent the normal functioning of subsequent sons' Y chromosomes (Bailey et al., 2016; O'Hanlan et al., 2018).



Figure 13.5.3: Four of the Marx Brothers in “I’ll Say She Is!”—caricatures by John Decker (1924). Every biological older brother increases the likelihood of a man being homosexual. Groucho (the brother with the mustache) was the fourth brother (of six- the eldest died in infancy, https://en.Wikipedia.org/wiki/Marx_Brothers) and called himself "straight but curved around the edges" (http://everydayheterosexism.blogspot.com/2012/11/the-marx-brothers.html).

13.5.2.2.4 Childhood Development

Two other forms of evidence supporting non-social causes of sexual orientation, particularly for males, are related to childhood development and behavior: gender nonconformity and outcomes from attempts to alter gender identity. Children who display gender nonconformity by not following the roles that society expects (Figure 13.5.4) are more likely to be homosexual as adults. This is a robust finding across cultures, but applies much more strongly to boys than girls. Additionally, "when infant boys are surgically and socially “changed” into girls, their eventual sexual orientation is unchanged (i.e., they remain sexually attracted to females)" (Bailey et al., 2016, page 46). This "natural experiment" is most commonly the result of irreparable damage to the penis, as was the case with David Reimer (see the section on differences of sexual development).



Figure 13.5.4: A two-year-old boy plays nursing his doll. Gender nonconformity in childhood is associated with homosexuality in adulthood.

13.5.3 In Closing: Gender and Sexual Orientation

Although gender (discussed in Section 13.4) and sexual orientation are distinct, some research addresses both concepts concurrently. A review of biological origins for both sexual orientation and gender identity concludes that "multidisciplinary evidence reveals that ... sexual orientation and gender identity are conferred ... during the first half of pregnancy" and that "multiple layers of evidence confirm that sexual orientation and gender identity are as biological, innate and immutable as the other traits conferred during that critical time in gestation" (O'Hanlan et al., 2018). In other words, you cannot change an individual's sexual

orientation and/or gender identity any more easily than you can change whether or not they developed a standard human body form with two arms and legs, two eyes and ears, a nose, mouth and typical internal organs.

13.5.4 Summary

Sexual orientation is an individual's emotional and sexual attraction to a particular sex or gender, including their sense of identity based on those attractions and related behaviors. A person may exhibit sexual fluidity (with attractions and behaviors changing due to circumstances), but sexual orientations tend to be relatively stable over the lifespan. Some argue that sexual attraction is primarily driven by reproduction, but empirical studies suggest that pleasure is the primary force behind our sex drive, and that reproduction has relatively little to do with it.

Arguing that dividing people into homosexual and heterosexual did not capture the breadth of human sexual orientation, Kinsey created a continuum with a seven-point scale called the Heterosexual-Homosexual Rating Scale. Zero is exclusively heterosexual, three is bisexual (with equal attractions to the same and opposite sexes), and six is exclusively homosexual. Subsequent research has supported this idea that sexual orientation ranges along a continuum. Representative studies of US and other Western populations report a 3.5% prevalence of non-heterosexual (gay/lesbian/bisexual) orientation. However, how a respondent is asked to report their sexual orientation produces very different results: using the Kinsey Scale, 18% to 39% of Europeans and Americans identify as somewhere between heterosexual and homosexual, whereas only 0.5% to 1.9% of respondents identify as non-heterosexual when researchers force individuals to respond using only two categories. While the percentage of US adults identifying as LGBT (lesbian/gay/bisexual/transgender) remains stable in older-aged cohorts, younger people are more likely to identify as LGBT than previous generations. Examples of terms currently used to identify sexual orientation include heterosexuality (attraction to the opposite sex/gender), homosexuality or same-sex attraction (attraction to the same sex/gender), bisexuality (attraction to two sexes/genders), polysexuality (attraction to multiple sexes/genders), pansexuality (attraction to all sexes/genders), and asexuality (no sexual attraction to any sex/gender).

Individuals are usually aware of their sexual orientation between middle childhood and early adolescence, and sexual activity is not necessary for this recognition. There is no scientific consensus regarding the root cause(s) of sexual orientation, but it is clearly not a conscious "choice" for either heterosexual or non-heterosexual individuals. Research has examined many possible biological, developmental, social, and cultural influences on sexual orientation. Social causes (such as having homosexual parents, recruitment by homosexual adults, and experiencing disordered parenting) have only weak support and many confounding factors. Nonsocial causes (such as genetics, prenatal hormone exposure, the fraternal-birth-order effect, and behavior differences during childhood development) have much more supporting evidence.

Genetic influence on sexual orientation is explored using twin studies. Concordance for sexual orientation is highest for monozygotic twins; and concordance rates for dizygotic twins, siblings, and randomly-selected pairs do not differ. Since concordance rates are most similar for monozygotic twins, this indicates that "nature" (genetics) influences the expression of sexual orientation. However, since monozygotic twins are not 100% concordant, "nurture" (the environment and individual experiences) also plays a role. During prenatal development, excess exposure to androgens for females and deficient exposure to androgens for males has also been theorized as explaining non-heterosexual orientation. Homosexual males tend to have more older brothers than heterosexual males, termed the "fraternal-birth-order effect. One proposed mechanism for this is that the mother creates antibodies against proteins on her developing son's Y chromosome, and that those antibodies may prevent the normal functioning of subsequent sons' Y chromosomes.

Childhood gender nonconformity and outcomes from attempts to alter gender identity also support non-social causes of sexual orientation (particularly for males). Boys who do not follow expected societal roles are more likely to be homosexual as adults. Additionally, when boys are “changed” into girls during childhood (often as a result of irreparable damage to the penis), they remain sexually attracted to women in adulthood. Both sexual orientation and gender identity are determined, at least in part, during prenatal development, and as such are not easily altered.

13.5.5 References

Note: These references are specifically those added to the content of this page by Naomi Bahm and do not include citations from the outside sources used. Refer to the text attributions to locate citations for articles from other sources.

Bailey, J.M., Vasey, P.L., Diamond, L.M., Breedlove, S.M., Vilain, E., & Epprecht, M. (2016). Sexual orientation, controversy, and science. *Psychological Science in the Public Interest, 17*(2), 45-101. doi: 10.1177/1529100616637616

Blanchard, R., Cantor, J., Bogaert, A., Breedlove, S.M., & Ellis, L. (2006). Interaction of fraternal birth order and handedness in the development of male homosexuality. *Hormones and Behavior, 49*, 405-414. doi: 10.1016/j.yhbeh.2005.09.002

Bogaert, A. (2007). Extreme right-handedness, older brothers, and sexual orientation in men. *Neuropsychology, 21*(1), 141-148. doi: 10.1037/0894-4105.21.1.141

Dimock, M. (2019). Defining generations: Where Millennials end and Generation Z begins. *Pew Research Center (January 17)*, https://pewrsr.ch/2szqtJz

Gates, G.J. (2017). LGBT data collection amid social and demographic shifts of the US LGBT community. *American Journal of Public Health, 107* (8), 1220-1222. https://doi.org/10.2105/AJPH.2017.303927

O'Hanlan, K.A., Gordon, J.C., & Sullivan, M.W. (2018). Biological origins of sexual orientation and gender identity: Impact on health. Gynecologic Oncology, 149, 33-42. https://doi.org/10.1016/j.ygyno.2017.11.014

13.5.6 Additional Resources

Educational graphics for understanding sex, gender, sexual orientation, and more: https://www.itspronouncedmetrosexual.com/2018/10/the-genderbread-person-v4/

How do queer couples have babies? Learn more here:



13.5.7 Attributions

1. Figures:
    1. First photo: Lesbian couple Sacramento by Bev Sykes from Davis, CA, USA, CC BY 2.0 via Wikimedia Commons; Second photo: Two Men Midtown Baltimore MD by Elvert Barnes, CC BY-SA 2.0, via Wikimedia Commons
    2. Own work (Kinsey Scale), licensed CC0, via Wikimedia Commons
    3. The four Marx Brothers by John Decker, Public domain, via Wikimedia Commons
    4. A boy playing, nursing his doll by Ms. Melissa, CC BY-SA 3.0, via Wikimedia Commons
2. Text:
    1. The Psychology of Human Sexuality by Don Lucas and Jennifer Fox, licensed CC BY-NC-SA 4.0 via Noba Project.
    2. " Development of Sexual Identity" by Paris, Ricardo, Raymond, & Johnson, LibreTexts is licensed under CC BY.
3. Changes:

1. Text from above two sources pieced together with some modifications, transitions and additional images and content (particularly in the sections A Continuum of Sexual Orientation, and Development and Origins of Sexual Orientation) added by Naomi I. Gribneau Bahm, PhD., Psychology Professor at Cosumnes River College, Sacramento, CA.
2. Birth year ranges for Millennials and Baby Boomers from: https://www.pewresearch.org/fact-tan...tion-z-begins/ (accessed 5/12/22; see Dimock, 2019 in references)

This page titled 13.5: Sexual Orientation is shared under a mixed license and was authored, remixed, and/or curated by Naomi Bahm (ASCCC Open Educational Resources Initiative (OERI)) .

- **15.8: Development of Sexual Identity** by Paris, Ricardo, Rymond, & Johnson is licensed CC BY 4.0. Original source: https://open.umn.edu/opentextbooks/textbooks/child-growth-and-development.

13.6: Sex Differences and Evolutionary Interpretations

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Explain sex differences that have been reported in the brain, sensory systems and cognitive functions, and the limitations of these findings
2. Describe how hormones may influence sex differences in behavior, such as childhood toy preferences and aggressive behaviors
3. Identify the core premises of sexual strategies theory

13.6.1 Overview

After noting that the dichotomy does not truly reflect the variations seen in both biological sex and gender identity, this section discusses sex differences that have been reported between human males and females. Sex differences in the brain, sex differences in sensory systems and cognition, and sex differences in behavior are covered, with the last topic including a review of hormonal influences on sex differences in behavior followed by discussions of childhood toy preferences and aggressive behaviors. Finally, evolutionary interpretations of reproductive behavior is addressed, including sexual strategies theory and error management theory (as it relates to reproductive behaviors).

13.6.2 Sex Differences

13.6.2.1 Preliminary Note

Earlier sections have emphasized the lack of a strict dichotomy between males and females, both in terms of biological sex and gender identity. Nonetheless, most research splits people into two categories- most commonly based on their biological sex assigned at birth- and often also assumes that the biological sex reported correlates with gender identity. While this is useful to a degree, hopefully future research will take a more nuanced approach to all three aspects of an individual's "sexual" self- sex, gender, and sexual orientation! You will note below that when we compare men and women on psychological traits and characteristics, there is actually only a slight difference (typically no more than 5%) between the two, and only at the group level. Think of it this way- adult men in the U.S. are about 9% taller than adult women, on average (McDowell et al., 2008), and about 15% larger in body mass (Larsen, 2003). Yet everyone can think of examples of men who are shorter and/or smaller than many women, and women who are taller and/or larger than many men- and this physical difference is at least twice the size of any of the differences found in psychological domains!

Also, as you will learn in the section on sex differences in the brain, while differences between male and female brains have been reported at a group level, there is no way to reliably classify an individual brain as either male or female, due to both the large overlap between males and females and the fact that each brain has its own unique mosaic of both "male" and "female" traits in differing (and unpredictable) ways.

13.6.2.2 Introduction to Sex Differences

Hens and roosters are different (Figure **13.6.1**). Cows and bulls are different. Men and women are different. Even girls and boys are different. Humans, like many animals, are sexually dimorphic (*di*, “two”; *morph*, “type”) in the size and shape of their bodies, their physiology, and for our purposes, their behavior.



Figure **13.6.1**: Sex differences in appearance are often more pronounced in nonhuman animals than in humans. Male birds particularly, for example roosters, tend to have physical features that differ from the females and also differ significantly in size.

Interestingly, the physical differences in size and appearance (for many individuals, we can readily determine biological sex from visible traits) between human males and females is related to some of the differences we see in human reproductive behaviors. In the animal kingdom, many animals whose mating strategy is to form monogamous pairs have virtually indistinguishable males and females, whereas polygamous species (where one male typically mates with many females, and some males completely miss out on mating) are conspicuously different, often in both physical features and in size. While apparent in some bird species, such as the obviously different polygamous rooster and hen mentioned above (versus monogamous swans, with nearly identical males and females), these differences are particularly evident in (non-human) primate species (and many other mammals). Examples in primate species include the similarly-appearing monogamous gibbons (Jaffe, 2019, page 210-211), versus the polygamous gorillas (Jaffe, 2019, page 209), where adult males are much larger than females and gain the characteristic "silverback" coloring as they mature (Figure **13.6.2**). A difference of around ten percent- precisely where the human species falls on this size-difference continuum- is the dividing line between "monogamous" and "polygamous"; in other words, species that have less than ten-percent difference in physical appearance and size between males and females are usually monogamous and species that have greater than ten-percent difference tend to be polygamous. For humans, "...the sexes differ more in human beings than in monogamous mammals, but much less than in extremely polygamous mammals..." (Daly & Wilson, 1996, page 13). Thus, it is not really surprising, from an evolutionary standpoint, that some humans adopt a monogamous strategy for mating and reproduction, and others follow a polygamous strategy.

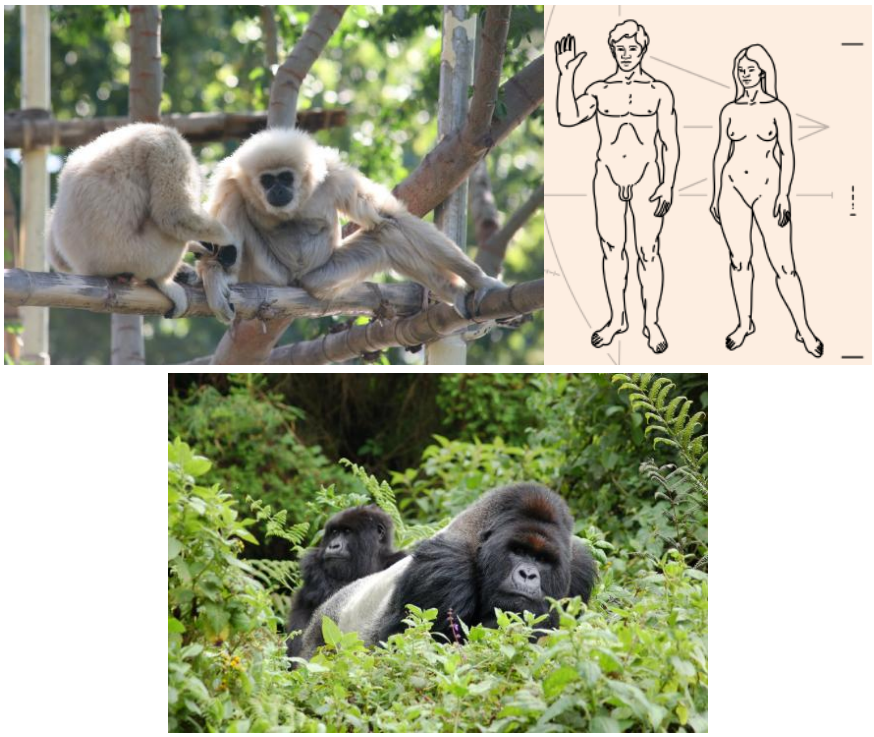


Figure 13.6.2: In humans (middle picture), males and females differ more than monogamous species, like White-handed Gibbons (left picture), but much less than polygamous species, like Mountain Gorillas (right picture).

Returning to human behavior, the behavior of boys and girls differs in many ways. Girls generally excel in verbal abilities relative to boys; boys are nearly twice as likely as girls to suffer from dyslexia (reading difficulties) and stuttering, and nearly four times more likely to suffer from autism. Boys are generally better than girls at tasks that require visuospatial abilities. Girls engage in nurturing behaviors more frequently than boys. More than 90% of all anorexia nervosa cases involve young women. Young men are twice as likely as young women to suffer from schizophrenia. Boys are much more physically aggressive and generally engage in more rough-and-tumble play than girls (Berenbaum, Martin, Hanish, Briggs, & Fabes, 2008). Many behavioral sex differences, such as the difference in physical aggressiveness, persist throughout adulthood. For example, there are many more men than women serving prison sentences for violent behavior. Below, this section addresses sex differences in the brain, in sensory systems and cognition, and in behavior, where sex differences in toy preferences and the previously mentioned aggressive behavior are discussed in greater detail.

13.6.2.3 Sex Differences in the Brain

Sex differences in overall human brain size have been reported for years (although the true significance of this finding is debatable). More recently, sex differences in specific brain structures have been discovered. For example, in rats, the sexually dimorphic nucleus (SDN-POA) of the preoptic area of the hypothalamus is much larger in males than females. This is a result of the organizing effects of gonadal steroid hormones upon brain and behavior, which is relatively constrained to the early stages of development. In rats, exposure to testosterone (which is converted to estradiol) or estradiol causes masculinization of the brain. (The same mechanism of a "female" hormone masculinizing the fetal brain is not present in primates, and thus presumably not in humans. Alpha-fetoprotein binds to estradiol in female rats to prevent masculinization of their brains.) Figure 13.6.3 depicts cross-sections through the brains of rats that show a male (A, left), a female (B, center), and a female treated with testosterone as a newborn (C, right). Note that the SDN-POA (the dark cell bodies) of the male are substantially larger than those of the untreated female but are equal in size to those of the testosterone-treated female. The extent that these sex differences in brain structure account for sex differences in behavior remain unspecified in mammals.

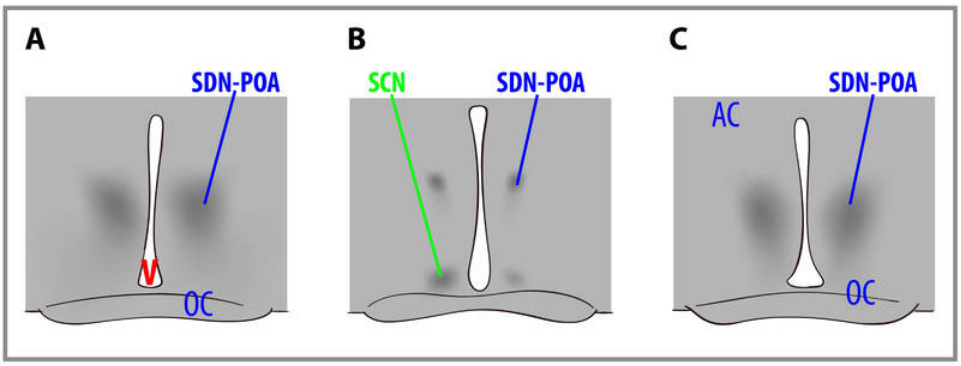


Figure **13.6.3**: The sexually dimorphic nucleus of the preoptic area (SDN-POA). Rat brain cross-sections from a A) male (left), B) female (center), and C) female treated with testosterone as a newborn (right). OC = optic chiasm; SCN = suprachiasmatic nucleus; V = third ventricle.

In an area similar to the rat's SDN-POA, Simon LeVay (1991) reported a comparable difference in the human hypothalamus-INAH-3, the third interstitial nucleus of the anterior hypothalamus. Specifically, heterosexual men had INAH-3 nuclei that were more than twice as large (by volume) as homosexual men, whose INAH-3 nuclei volumes were similar to those of women. Although LeVay was able to replicate the result, attempts by another researcher lead to only a trend-level difference between heterosexual and homosexual men in the volume of INAH-3 (Byne 2000 & 2001, as cited by Bailey et al., 2016). Regardless of the true status of the possible differences in the human male's INAH-3 nucleus (it could be intermediate between the two researchers' findings, for example), "it would be unlikely that the INAH-3 size would be a key factor regulating sexual orientation. This is because there would be too many exceptions—homosexual men with a large INAH-3 and heterosexual men with a small INAH-3—to believe that INAH-3 size is crucial" (Bailey et al., 2016, page 72). The INAH-3 is a tiny structure- about the size of a grain of sand- and studying it requires post-mortem tissue. As such, it is unlikely that it will be studied further in the near future (Bailey et al., 2016).

For larger brain features, which can be studied in living people using imaging techniques, many brain differences between adult men and women have been reported, both in cortical regions and subcortical structures. An example of this includes magnetic resonance imaging (MRI) data suggesting that women have larger volumes of frontal and medial paralimbic cortices and men have larger volumes in the frontomedial cortex, amygdala and hypothalamus (Goldstein et al., 2001). However, closer scrutiny sometimes renders these sex differences unimportant. For example, the corpus callosum was widely believed to differ between males and females- the purported increased size in women was touted to result in greater connectivity between the right and left hemispheres of the brain- but has been found to not differ at all when men and women were matched for brain size (Luders et al., 2014). Another study underlining the importance of taking brain size into consideration was a twin study that found that a weak correlation between sex differences in the brain and behavioral sex differences turned out to be mainly driven by brain size. The authors warned that caution should therefore be used when inferring causality (van Eijk et al., 2021).

Additionally, Joel et al. (2015) conducted an extensive study of over 1400 MRIs of human brains and found that when taking the entire brain into account (and not just looking at one specific area versus another, where sex differences have been reported at a group level), it is impossible to categorize brains into male and female forms. Instead, there was extensive overlap of “male” and “female” features in human brains and it was very rare to have internal consistency (only “male” or only “female” features) in a specific brain. Thus, while we can often categorize biological sex by looking at anatomical structures- there are often distinct differences between male and female bodies, and a given body usually has only male or only female structures- we cannot determine if a specific brain came from a male or female body, nor can we predict by knowing the sex/gender of a particular individual what form their brain has taken. The authors state that “each brain is a unique mosaic of features, some of which may be more common in females compared with males, others may be more common in males compared with females, and still others may be common in both males and females” (Joel et al., 2015, page 15472). Additionally, their findings support the notion that masculinization and feminization of brain areas are two separate processes that progress independently, allowing variations of differentiation in a “male” or “female” pattern to occur within the same brain.

13.6.2.4 Sex Differences in Sensory Systems and Cognition

Sex differences in a number of sensory and cognitive functions have also been reported. Females are generally more sensitive to auditory information, whereas males are more sensitive to visual information. Females are also typically more sensitive than males to taste and olfactory input. Women display less lateralization of cognitive functions than men. On average, females generally excel in verbal, perceptual, and fine motor skills, whereas males outperform females on quantitative and visuospatial tasks, including map reading and direction finding. Interestingly, women with Turner syndrome (whose ovaries are not functional) often have impaired spatial memory.

Although reliable sex differences can be documented, these differences in ability are slight. It is important to note that there is more variation *within* each sex than *between* the sexes for most cognitive abilities (Figure **13.6.4**).

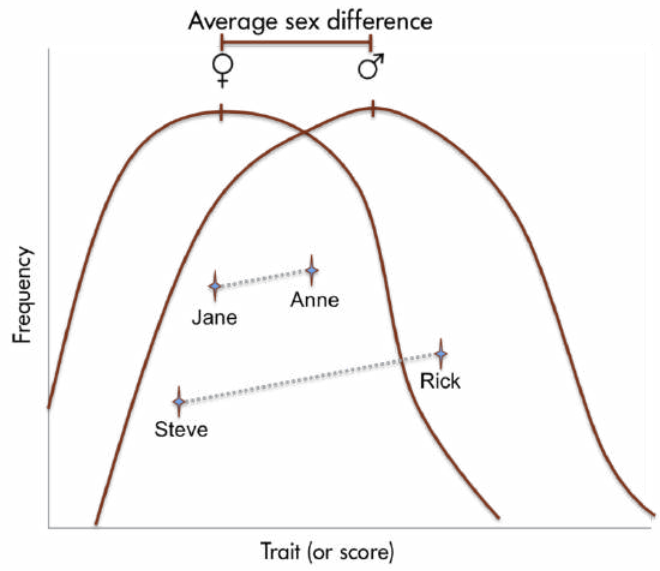


Figure **13.6.4**: The average sex differences in human performance often reflect significant overlap between the sexes There are often greater differences in performance between individuals of the same sex (for example, between Steve and Rick in the figure) than between individuals of the opposite sex (for example, between Steve and Jane in the figure).

To emphasize the importance of recognizing the vast variation of personality traits and behavioral expressions found within all humans, regardless of their sex/gender, consider the following study results. Joel et al. (2015) conducted an analysis of the personality traits, attitudes, interests, and behaviors of more than 5,500 individuals. Their findings mirrored their MRI findings (great overlap between males and females and little internal consistency within the same individual). They report: “In accordance with the brain data, our analyses of gender-related data revealed extensive overlap between females and males in personality traits, attitudes, interests, and behaviors. Moreover, we found that substantial variability of gender characteristics is highly prevalent, whereas internal consistency is extremely rare, even for highly gender-stereotyped activities... Thus, most humans possess a mosaic of personality traits, attitudes, interests and behaviors, some more common in males compared with females, others more common in females compared with males, and still others common in both males and females” (Joel et al., 2015, page 15472). It is reasonable to conclude that overall, human males and females are more similar than they are different, both with regards to brain features and psychological traits.

13.6.2.5 Sex Differences in Behavior

13.6.2.5.1 Review of Hormonal Influences on Sex Differences in Behavior

The hormonal differences between men and women may account for adult sex differences that develop during puberty, but what accounts for behavioral sex differences among children *prior* to puberty and activation of their gonads? As discussed in section 13.2, the presence (or absence) of androgens determine whether the fetus develops a male or female body form. Androgens from the fetal testes steer the development of the body, central nervous system, and subsequent behavior in a male direction. In contrast, fetal ovaries do not secrete high concentrations of hormones, so the body, central nervous system, and later, behavior, follow a female pathway.

As discussed in section 13.1, gonadal steroid hormones have organizational (or programming) effects upon brain and behavior (Phoenix, Goy, Gerall, & Young, 1959). These organizing effects are relatively constrained to the early stages of development. In rats, early steroid hormone treatment causes relatively irreversible and permanent masculinization of later adult behavior (mating and aggressive). In contrast, activational effects of hormones (provided in adulthood to influence behaviors) are temporary and reversible. Thus, typical male behavior requires exposure to androgens during gestation (in humans) or immediately after birth (in rats) to somewhat masculinize the brain and also requires androgens during or after puberty to activate these neural circuits. Typical female behavior requires a lack of exposure to androgens early in life which leads to feminization of the brain and also requires estrogens to activate these neural circuits in adulthood. But this simple dichotomy, which works well with animals with very distinct sexual dimorphism in behavior, has many caveats when applied to people.

13.6.2.5.2 Childhood Toy Preferences

If you walk through any major toy store, then you will likely observe a couple of aisles filled with pink boxes and the complete absence of pink packaging of toys in adjacent aisles. Remarkably, you will also see a strong self-segregation of boys and girls in these aisles. It is rare to see boys in the “pink” aisles and vice versa. The toy manufacturers are often accused of making toys that are gender biased, but it seems more likely that boys and girls enjoy playing with specific types and colors of toys. Indeed, toy manufacturers would immediately double their sales if they could sell toys to both sexes. Boys generally prefer toys such as trucks and balls and girls generally prefer toys such as dolls. Although it is doubtful that there are genes that encode preferences for toy cars and trucks on the Y chromosome, it is possible that hormones might shape the development of a child’s brain to prefer certain types of toys or styles of play behavior. It is also reasonable to believe that children learn which types of toys and which styles of play are appropriate to their gender (Figure **13.6.5**). How can we separate the contribution of learning from underlying physiological mechanisms to understand sex differences in human behaviors? To untangle these issues, animal models are often used. Unlike the situation in humans, where sex differences are usually only a matter of degree (often slight), in some animals, members of only one sex may display a particular behavior. For example, often only male songbirds sing. Studies of such strongly sex-biased behaviors are particularly valuable for understanding the interaction among behavior, hormones, and the nervous system.



Figure **13.6.5**: If you think back to the toys you played with and clothing you wore in your youth, do you think they were more a result of your hormonal activity or the choices that society and your parents made for you?

A study of vervet monkeys calls into question the primacy of learning in the establishment of toy preferences (Alexander & Hines, 2002). Female vervet monkeys preferred girl-typical toys, such as dolls or cooking pots, whereas male vervet monkeys preferred boy-typical toys, such as cars or balls. There were no sex differences in preference for gender-neutral toys, such as picture books or stuffed animals. Presumably, monkeys have no prior concept of “boy” or “girl” toys. Young rhesus monkeys also show similar toy preferences.

What then underlies the sex difference in toy preference? It is possible that certain attributes of toys (or objects) appeal to either boys or girls. Toys that appeal to boys or male vervet or rhesus monkeys, in this case, a ball or toy car, are objects that can be

moved actively through space, toys that can be incorporated into active, rough and tumble play. The appeal of toys that girls or female vervet monkeys prefer appears to be based on color. Pink and red (the colors of the doll and pot) may provoke attention to infants.

Society may reinforce such stereotypical responses to gender-typical toys. The sex differences in toy preferences emerge by 12 or 24 months of age and seem fixed by 36 months of age, but are sex differences in toy preference present during the first year of life? It is difficult to ask pre-verbal infants what they prefer, but in studies where the investigators examined the amount of time that babies looked at different toys, eye-tracking data indicate that infants as young as three months showed sex differences in toy preferences; girls preferred dolls, whereas boys preferred trucks. Another result that suggests, but does not prove, that hormones are involved in toy preferences is the observation that girls diagnosed with congenital adrenal hyperplasia (CAH), whose adrenal glands produce varying amounts of androgens early in life, played with masculine toys more often than girls without CAH. Further, a dose-response relationship between the extent of the disorder (i.e., degree of fetal androgen exposure) and degree of masculinization of play behavior was observed. Are the sex differences in toy preferences or play activity, for example, the inevitable consequences of the differential hormonal environments of boys and girls, or are these differences imposed by cultural practices and beliefs? Are these differences the result of receiving gender-specific toys from an early age, or are these differences some combination of hormonal and cultural factors? Again, these are difficult questions to unravel in people.

13.6.2.5.3 Aggressive Behaviors

Note that in this section "aggressive behavior" is referring specifically to the expression of physical aggression (arguably the only type of aggression that can be studied in animals as well as humans). When aggression is defined more broadly to include relational aggression (such as non-physical bullying, gossiping, and humiliation), human males and females show much more similar overall levels of aggression (Swearer Napolitano, 2008).

The possibility for aggressive behavior exists whenever the interests of two or more individuals are in conflict (Nelson, 2006). Conflicts are most likely to arise over limited resources such as territories, food, and mates. A social interaction decides which animal gains access to the contested resource. In many cases, a submissive posture or gesture on the part of one animal avoids the necessity of actual combat over a resource. Animals may also participate in threat displays or ritualized combat in which dominance is determined but no physical damage is inflicted.

There is overwhelming circumstantial evidence that androgens mediate aggressive behavior across many species. First, seasonal variations in blood plasma concentrations of testosterone and seasonal variations in aggression coincide. For instance, the incidence of aggressive behavior peaks for male deer in autumn, when they are secreting high levels of testosterone. Second, aggressive behaviors increase at the time of puberty, when the testes become active and blood concentrations of androgens rise. Juvenile deer do not participate in the fighting during the mating season. Third, in any given species, males are generally more aggressive than females. This is certainly true of deer; relative to stags, female deer rarely display aggressive behavior, and their rare aggressive acts are qualitatively different from the aggressive behavior of males. Finally, castration (removal of the testes) typically reduces aggression in males, and testosterone replacement therapy restores aggression to pre-castration levels. There are some interesting exceptions to these general observations that are outside the scope of this section.

As mentioned, males are generally more aggressive than females. Certainly, human males are much more physically aggressive than females. Many more men than women are convicted of violent crimes in North America. The sex differences in human aggressiveness appear very early. At every age throughout the school years, many more boys than girls initiate physical assaults. Almost everyone will acknowledge the existence of this sex difference, but assigning a cause to behavioral sex differences in humans always elicits much debate. It is possible that boys are more aggressive than girls because androgens promote aggressive behavior and boys have higher blood concentrations of androgens than girls. It is possible that boys and girls differ in their aggressiveness because the brains of boys are exposed to androgens prenatally and the “wiring” of their brains is thus organized in a way that facilitates the expression of aggression (Figure **13.6.6**). It is also possible that boys are encouraged and girls are discouraged by family, peers, or others from acting in an aggressive manner. These three hypotheses are not mutually exclusive, but it is extremely difficult to discriminate among them to account for sex differences in human aggressiveness.

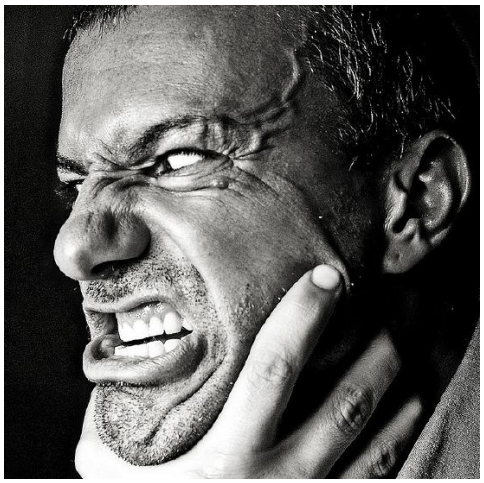


Figure **13.6.6**: Researchers have electrically stimulated particular regions in people's brains, and these individuals have burst into aggressive, violent behavior, helping demonstrate that such responses are hardwired into us.

What kinds of studies would be necessary to assess these hypotheses? It is usually difficult to separate out the influences of environment and physiology on the development of behavior in humans. For example, boys and girls differ in their rough-and-tumble play at a very young age, which suggests an early physiological influence on aggression. This pattern of more physical play behavior is seen in a number of other species including nonhuman primates, rats, and dogs. Is the difference in the frequency of rough-and-tumble play between boys and girls due to biological factors associated with being male or female, or is it due to cultural expectations and learning? Parents interact with their male and female offspring differently; they usually play more roughly with male infants than with females, which suggests that the sex difference in aggressiveness is partially learned. This difference in parental interaction style is evident by the first week of life. If there is a combination of biological and cultural influences mediating the frequency of rough-and-tumble play, then what proportion of the variation between the sexes is due to biological factors and what proportion is due to social influences? Importantly, is it appropriate to talk about "normal" sex differences when these traits virtually always arrange themselves along a continuum rather than in discrete categories? Because of these complexities in the factors influencing human behavior, the study of hormonal effects on sex-differentiated behavior has been pursued in nonhuman animals, where it is easier to hold environmental influences relatively constant.

With the appropriate animal model, we can address the questions posed above: Is the sex difference in aggression due to higher adult blood concentrations of androgens in males than in females, or are males more aggressive than females because their brains are organized differently by perinatal hormones? Are males usually more aggressive than females because of an interaction of early and current blood androgen concentrations? If male mice are castrated prior to their sixth day of life, then treated with testosterone in adulthood, they show low levels of aggression. Similarly, female mice ovariectomized prior to their sixth day but given androgens in adulthood do not express male-like levels of aggression. Treatment of perinatally gonadectomized males or females with testosterone prior to their sixth day life and also in adulthood results in a level of aggression similar to that observed in typical male mice. Thus, in mice, the proclivity for males to act more aggressively than females is organized perinatally by androgens but also requires the presence of androgens after puberty in order to be fully expressed. In other words, aggression in male mice is both organized and activated by androgens. Testosterone exposure in adulthood without prior organization of the brain by steroid hormones does not evoke typical male levels of aggression. The hormonal control of aggressive behavior in mice is thus similar to the hormonal mediation of heterosexual male mating behavior in other rodent species. Aggressive behavior is similarly both organized and activated by androgens in many species, including rats, hamsters, voles, dogs, and possibly some primate species.

13.6.3 Evolutionary Interpretations of Reproductive Behavior

Evolutionary psychology connects evolutionary principles with modern psychology and focuses primarily on psychological adaptations: changes in the way we think in order to improve our survival. (See Chapter 3 for more information on evolution.) This section describes one of the major evolutionary psychological theories that is relevant to reproductive behaviors- sexual strategies theory- which describes the psychology of human mating strategies and the ways in which women and men differ in those

strategies. Additionally, an example from another evolutionary psychological theory- error management theory- that is relevant to human mating is provided.

13.6.3.1 Sexual Strategies Theory

Sexual strategies theory is based on sexual selection theory- the evolution of characteristics due to mating advantage rather than survival advantage, such as traits that make an individual more attractive to the opposite sex, and competition between members of the same sex. It proposes that humans have evolved a list of different mating strategies, both short-term and long-term, that vary depending on culture, social context, parental influence, and personal mate value (desirability in the “mating market”). Initially the focus was on the differences between men and women in mating preferences and strategies (Buss & Schmitt, 1993), starting by looking at the minimum parental investment needed to produce a child. For women, even the minimum investment is significant: after becoming pregnant, they have to carry that child for nine months inside of them (Figure **13.6.7**). For men, on the other hand, the minimum investment to produce the same child is considerably smaller—simply the act of sex.



Figure **13.6.7**: Because women bear responsibility for pregnancy, they may use different mating strategies than men do.

These differences in parental investment have an enormous impact on sexual strategies. For a woman, the risks associated with making a poor mating choice is high. She might get pregnant by a man who will not help to support her and her children, or who might have poor-quality genes. And because the stakes are higher for a woman, *wise* mating decisions for her are much more valuable. While this discussion is not intended in any way to endorse "bad behavior", for men, biologically speaking, the need to focus on making wise mating decisions isn’t as important. That is, unlike women, men 1) don’t biologically have the child growing inside of them for nine months, and 2) do not have as high a cultural expectation to raise the child. This logic leads to a powerful set of predictions: In short-term mating, women will likely be choosier than men (because the costs of getting pregnant are so high), while men, on average, will likely engage in more casual sexual activities (because this cost is greatly lessened). Due to this, men will sometimes deceive women about their long-term intentions for the benefit of short-term sex, and men are more likely than women to lower their mating standards for short-term mating situations.

An extensive body of empirical evidence supports these and related predictions (Buss & Schmitt, 2011). Men express a desire for a larger number of sex partners than women do. They let less time elapse before seeking sex. They are more willing to consent to sex with strangers and are less likely to require emotional involvement with their sex partners. They have more frequent sexual fantasies and fantasize about a larger variety of sex partners. They are more likely to regret missed sexual opportunities. And they lower their standards in short-term mating, showing a willingness to mate with a larger variety of women as long as the costs and risks are low.

However, it is also important to note that in situations where both the man and woman are interested in long-term mating, both sexes tend to invest substantially in the relationship and in their children. In these cases, the theory predicts that both sexes will be extremely choosy when pursuing a long-term mating strategy. Much empirical research supports this prediction, as well. In fact, the qualities women and men generally look for when choosing long-term mates are very similar: both want mates who are intelligent, kind, understanding, healthy, dependable, honest, loyal, loving, and adaptable.

Nonetheless, women and men do differ in their preferences for a few key qualities in long-term mating, because of somewhat distinct adaptive problems. Modern women have inherited the evolutionary trait to desire mates who possess resources, have qualities linked with acquiring resources (e.g., ambition, wealth, industriousness), and are willing to share those resources with them. On the other hand, men more strongly desire youth and health in women, as both are cues to fertility. These male and female differences are universal in humans. They were first documented in 37 different cultures, from Australia to Zambia (Buss, 1989), and have been replicated by dozens of researchers in dozens of additional cultures (for summaries, see Buss, 2012).

As we know, though, just because we have these mating preferences (e.g., men with resources; fertile women), people don't always get what they want. There are countless other factors which influence who people ultimately select as their mate. For example, the sex ratio (the percentage of men to women in the mating pool), cultural practices (such as arranged marriages, which inhibit individuals' freedom to act on their preferred mating strategies), the strategies of others (e.g., if everyone else is pursuing short-term sex, it's more difficult to pursue a long-term mating strategy), and many others all influence who we select as our mates.

Sexual strategies theory—anchored in sexual selection theory— predicts specific similarities and differences in men and women's mating preferences and strategies. Whether we seek short-term or long-term relationships, many personality, social, cultural, and ecological factors will all influence who our partners will be.

13.6.3.2 Error Management Theory

How we think, make decisions, and evaluate uncertain situations (error management theory) has also been used to predict adaptive biases in the domain of mating. Consider something as simple as a smile. In one case, a smile from a potential mate could be a sign of sexual or romantic interest. On the other hand, it may just signal friendliness. Because of the costs to men of missing out on chances for reproduction, error management theory predicts that men have a *sexual overperception bias*: they often misread sexual interest from a woman, when really it's just a friendly smile or touch. In the mating domain, the sexual overperception bias is one of the best-documented phenomena. It's been shown in studies in which men and women rated the sexual interest between people in photographs and videotaped interactions. As well, it's been shown in the laboratory with participants engaging in actual "speed dating," where the men interpret sexual interest from the women more often than the women actually intended it (Perilloux, Easton, & Buss, 2012). In short, error management theory predicts that men, more than women, will over-infer sexual interest based on minimal cues, and empirical research confirms this adaptive mating bias.

13.6.4 Summary

Humans, like many animals, are sexually dimorphic in the size and shape of their bodies, their physiology, and for our purposes, their behavior. Girls generally excel in verbal abilities, whereas boys generally have better visuospatial abilities. Boys are more likely to suffer from dyslexia, stuttering, autism, and schizophrenia, while most anorexia nervosa cases involve young women. Girls engage in nurturing behaviors more frequently, and boys generally engage in more rough-and-tumble play. One behavioral sex difference that persists into adulthood is the higher levels of physical aggression exhibited by boys and men.

In rats, the sexually dimorphic nucleus of the preoptic area in the hypothalamus (SDN-POA) is much larger in males than females. A female rat treated with testosterone as a newborn develops an SDN-POA similar in size to a typical male rat. In humans, a similar brain area, the third interstitial nucleus of the anterior hypothalamus (INAH-3), was found to be larger in heterosexual men than homosexual men (whose INAH-3 was similar in size to those of women). However, replication outside of the initial lab did not reach statistical significance. Brain imaging studies have reported various differences in cortical regions and subcortical structures, but closer scrutiny sometimes renders these sex differences unimportant, and caution should be used when inferring causality. A study of over 1400 MRIs of human brains found that when taking the entire brain into account, it is impossible to categorize brains into male and female forms. There was extensive overlap of "male" and "female" features in human brains and it was very rare to have internal consistency in a specific brain. These findings also support the notion that masculinization and feminization of brain areas are two separate processes that progress independently, allowing for considerable variation across individuals.

Sex differences in a number of sensory and cognitive functions have also been reported, but these differences in ability are slight. Furthermore, there is more variation *within* each sex than *between* the sexes for most cognitive abilities. An analysis of personality traits, attitudes, interests, and behaviors of more than 5,500 individuals mirrored the MRI findings (great overlap between males and females and little internal consistency within the same individual). It is reasonable to conclude that overall, human males and females are more similar than they are different, both with regards to brain features and psychological traits.

Boys generally prefer toys such as trucks and balls and girls generally prefer toys such as dolls. It is possible that hormones shape the development of a child's brain to prefer certain types of toys or styles of play behavior, but children also learn which types of toys and which styles of play are appropriate to their gender. Studies of vervet monkeys and rhesus monkeys find that females prefer girl-typical toys, and males prefer boy-typical toys, with no sex differences in preference for gender-neutral toys. While society likely reinforces stereotypical responses to gender-typical toys in human children, sex differences in toy preferences emerge by 12 or 24 months of age and seem fixed by 36 months of age. Studies using eye-tracking indicate that infants as young as three months show stereotypical sex differences in toy preferences. Additionally, in a dose-response relationship, girls diagnosed with congenital adrenal hyperplasia (CAH) played with masculine toys more often than girls without CAH, suggesting that prenatal hormones may affect later toy preferences.

Aggressive behavior may occur when two (or more) individuals are in conflict, typically over limited resources such as territories, food, and mates. Evidence strongly suggests that androgens mediate aggressive behavior across many species. Seasonal variations in testosterone levels and variations in aggression coincide; aggressive behaviors increase at puberty (when the testes become active and androgens rise); in any given species, males are generally more aggressive than females; and castration (removal of the testes) typically reduces aggression in males, while testosterone replacement therapy restores aggression. Human males are much more physically aggressive than females, and these sex differences appear very early. It is possible that boys are more aggressive than girls because androgens promote aggressive behavior, it could be due to exposure to androgens prenatally, or boys may be encouraged to act in aggressive ways. These three hypotheses are not mutually exclusive, but it is extremely difficult to discriminate among them to account for sex differences in human aggressiveness. In many animal species (mice, rats, hamsters, voles, dogs, and possibly some primates), aggressive behavior is both organized and activated by androgens.

Sexual strategies theory describes the psychology of human mating strategies and the ways in which women and men differ in those strategies. Because differences in parental investment are so large between males and females- men have little necessary investment and consequently don't need to make wise choices, whereas mating decisions for women are much more important (since a minimum of none months carrying the child is required)- women are predicted to be choosier than men in short-term mating. Thus, men will sometimes deceive women for the benefit of short-term sex, and men are more likely than women to lower their mating standards for short-term mating situations. However, when choosing long-term mates, the qualities women and men seek are very similar: mates who are intelligent, kind, understanding, healthy, dependable, honest, loyal, loving, and adaptable. There are nonetheless some universal differences in mate preference- women desire mates who possess resources (and are willing to share them), and men more strongly desire youth and health in women (as both are cues to fertility).

13.6.5 References

Note: These references are specifically those added to the content of this page by Naomi Bahm and do not include citations from the outside sources used. Refer to the text attributions to locate citations for articles from other sources.

Bailey, J.M., Vasey, P.L., Diamond, L.M., Breedlove, S.M., Vilain, E., & Epprecht, M. (2016). Sexual orientation, controversy, and science. *Psychological Science in the Public Interest, 17*(2), 45-101. doi: 10.1177/1529100616637616

Daly. M., & Wilson, M. (1996). Evolutionary psychology and marital conflict. In Buss, D.M., & Malamuth, N.M. (editors), *Sex, Power, Conflict: Evolutionary and Feminist Perspectives*. Oxford University Press.

Goldstein, J., Seidman, L.J., Horton, N.J., Makris, N., Kennedy, D.N., Caviness, V.S. Jr., Faraone, S.V., & Tsuang M.T. (2001). Normal sexual dimorphism of the adult human brain assessed by in vivo magnetic resonance imaging. *Cerebral Cortex, 11*, 490-497.

Jaffe, K.E. (2019). Chapter 6: Primate Behavior and Ecology. In Shook, B., Nelson, K., Aguilera, K.& Braff, L. (editors), *Explorations: An Open Invitation to Biological Anthropology*. American Anthropological Association, licensed CC BY-NC 4.0 International, ISBN 978-1-931303-63-7

Joel, D., Berman, Z., Tavor, I., Wexler, N., Gaber, O., Stein, Y., Shefi, N., Pool, J., Urchs, S., Margulies, D.S., Liem, F., Hänggi, J., Jäncke, L. & Assaf, Y. (2015). Sex beyond the genitalia: The human brain mosaic. *PNAS, 112*(50), 15468-15473. https://doi.org/10.1073/pnas.1509654112

Larsen, C.S. (2003). Equality for the sexes in human evolution? Early hominid sexual dimorphism and implications for mating systems and social behavior. *Proceedings of the National Academy of Sciences of the United States of America, 100* (16), 9103–

9104. https://www.pnas.org/doi/10.1073/pnas.1633678100

LeVay, S. (1991). A difference in hypothalamic structure between heterosexual and homosexual men. *Science*, *253*(5023), 1034-1037. doi: 10.1126/science.1887219

Luders, E., Toga, A.W., & Thompson, P.M. (2014) Why size matters: Differences in brain volume account for apparent sex differences in callosal anatomy. The sexual dimorphism of the corpus callosum. *NeuroImage*, *84*, 820-824. http://dx.doi.org/10.1016/j.neuroimage.2013.09.040

McDowell, M.A., Fryar, C.D., Ogden, C.L.,& Flegal, K.M. (2008). Anthropometric reference data for children and adults: United States, 2003–2006. *National Health Statistics Reports*, *10* (October 22, 2008), 1-48. https://www.cdc.gov/nchs/data/nhsr/nhsr010.pdf

Swearer Napolitano, S.M. (2008) Relational aggression: Not just a female issue. *Journal of School Psychology*, *46*, 611-616. doi: 10.1016/j.jsp.2008.08.001 [available at https://digitalcommons.unl.edu/edpsychpapers/142/]

van Eijk, L., Zhu, D., Couvy-Duchesne B., Strike, L.T., Lee, A.J., Hansell, N.K., Thompson, P.M., de Zubicaray, G.I., McMahon, K.L., Wright, M.J., & Zietsch, B.P. (2021) Are sex differences in human brain structure associated with sex differences in behavior? *Psychological Science*, *32*(8), 1183-1197. doi: 10.1177/0956797621996664

13.6.6 Additional Resources

Web: Main international scientific organization for the study of evolution and human behavior, HBES

http://www.hbes.com/

Web: Books and Interviews with David Buss

https://labs.la.utexas.edu/buss/books/

Web: Publications by Buss and colleagues

https://labs.la.utexas.edu/buss/publications/

Journal article: Founders of Evolutionary Psychology

13.6.7 Attributions

1. Figures:
    1. Rooster and hen by John Cudworth, licensed CC BY-NC 2.0, found in NOBA Hormones and Behavior by Randy Nelsen
    2. Left photo: White-handed Gibbons by Cliff from Arlington, Virginia, USA, CC BY 2.0 via Wikimedia Commons; Middle image: Humans extracted from the Pioneer plaque by NASA; vectors by Mysid, Public domain, via Wikimedia Commons; Right photo: Mountain gorillas by Joachim Huber, CC BY-SA 2.0, via Wikimedia Commons
    3. SDN-POA, no attribution or license information, found in NOBA Hormones and Behavior by Randy Nelsen
    4. Graph of sex differences overlap, no attribution or license information, found in NOBA Hormones and Behavior by Randy Nelsen
    5. Left: Brother and sister, by Amanda Westmont, https://goo.gl/ntS5qx, licensed CC BY-NC-SA 2.0, found in NOBA Social and Personality Development in Childhood by Ross Thompson; Right: Seated girl, licensed CC0 Public Domain, found in NOBA Hormones and Behavior by Randy Nelsen
    6. Aggressive expression by Riccardo Cuppini, CC BY-NC-ND 2.0, found in NOBA Hormones and Behavior by Randy Nelsen
    7. Pregnant woman, licensed CC0 Public Domain, found in NOBA Evolutionary Theories in Psychology by David Buss
2. Text adapted from:
    1. Hormones & Behavior by Randy J. Nelson, licensed CC BY-NC-SA 4.0 via Noba Project.
    2. Evolutionary Theories in Psychology by David M. Buss, licensed CC BY-NC-SA 4.0 via Noba Project.
3. Changes: Text (and images) from above two sources pieced together with some modifications, transitions and additional content (particularly in the Preliminary Note, Introduction to Sex Differences, Sex Differences in the Brain, and Sex Differences in

Sensory Systems and Cognition sections) added by Naomi I. Gribneau Bahm, PhD., Psychology Professor at Cosumnes River College, Sacramento, CA.

This page titled 13.6: Sex Differences and Evolutionary Interpretations is shared under a mixed license and was authored, remixed, and/or curated by Naomi Bahm (ASCCC Open Educational Resources Initiative (OERI)) .

CHAPTER OVERVIEW

14: Intelligence and Cognition

> This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

14.1: Overview of Brain Mechanisms in Intelligence, Language, and Cognition
14.2: Cognition, Intelligence, and Language as Psychological Adaptations
14.3: High Intelligence-Nature or Nurture; Learning Disabilities
14.4: Cognition- Categories, Concepts, Schemas and the Brain
14.5: Brain Mechanisms and Intelligence

This page titled 14: Intelligence and Cognition is shared under a mixed license and was authored, remixed, and/or curated by Keneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

14.1: Overview of Brain Mechanisms in Intelligence, Language, and Cognition

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Discuss the general relationship across species between brain size, body size, and intelligence, including trends in corticalization and cortical folding
2. Describe the functions of glial cells
3. Describe the locations and functions of each of the lobes of the cerebral cortex
4. Describe contralateral control
5. Discuss general principles that determine the amounts of somatosensory and motor cortex devoted to each part of the body
6. Describe visual agnosia and the location of the brain damage that causes it
7. Describe neuroplasticity and briefly discuss examples of it
8. Discuss what the split-brain experiments tell us about cerebral lateralization of function

14.1.1 Overview

We begin this chapter with a general review of the brain areas involved in complex psychological functions, focusing on the cerebral cortex, including studies of the split brain. Although an introductory psychology course is a prerequisite for this course in Biological Psychology, a refresher will be useful for you, the student, in preparation for the detail to be discussed in subsequent modules of this chapter.

14.1.2 **The Cerebral Cortex Creates Intelligence, Language, and Thinking: Overview of the Basics**

All animals have adapted to their environments by developing abilities that help them survive. Some animals have hard shells, others run extremely fast, and some have acute hearing. Human beings do not have any of these particular characteristics, but we do have one big advantage over other animals—we are very, very smart.

You might think that we should be able to determine the intelligence of an animal by looking at the ratio of the animal's brain weight to the weight of its entire body. But this does not really work. The elephant's brain is one thousandth of its weight, but the whale's brain is only one ten-thousandth of its body weight. On the other hand, although the human brain is one 60th of its body weight, the mouse's brain represents one fortieth of its body weight. Despite these comparisons, elephants do not seem 10 times smarter than whales, and humans definitely seem smarter than mice.

The key to the advanced intelligence of humans is not found in the size of our brains. What sets humans apart from other animals is our larger cerebral cortex—*the outer bark-like layer of our brain that allows us to so successfully use language, acquire complex skills, create tools, and live in social groups* (Gibson, 2002). In humans, the cerebral cortex is wrinkled and folded, rather than smooth as it is in most other animals. This creates a much greater surface area and size, and allows increased capacities for learning, remembering, and thinking. The increasing amount of cortex in mammals involving folding of the cerebral cortex is referred to as *corticalization*.

Although the cortex is only about one tenth of an inch thick, it makes up more than 80% of the brain's weight. The human cortex contains about 20 billion nerve cells (the entire human brain has somewhere between 100 and 200 billion neurons) and at least 300 trillion synaptic connections (de Courten-Myers, 1999). Supporting all these neurons are billions more glial cells (glia), *cells that surround and link to the neurons, protecting them, providing them with nutrients, and absorbing unused neurotransmitters*. The glia come in different forms and have different functions. For instance, the myelin sheath surrounding the axon of many neurons is a type of glial cell. The glia are essential partners of neurons, without which the neurons could not survive or function (Miller, 2005).

As you recall from the chapter on anatomy of the nervous system, the cerebral cortex is divided into two *hemispheres*, and each hemisphere is divided into four *lobes*, each separated by little "valleys" known as *fissures* (the little hills between fissures are called

gyri, singular, *gyrus)*.

If we look at the cortex starting at the front of the brain and moving over the top (see Figure 14.1.1), we see first the **frontal lobe** (behind the forehead), which is *responsible primarily for thinking, planning, memory, and judgment*. In addition, in most people the *left frontal lobe contains Broca's area* essential for speech and other language functions (in some left handers, Broca's area may be in the right frontal lobe or language may be less lateralized and instead is spread more equally between the two hemispheres; thus, left handers tend to lose less language function than right handers if they suffer damage to left frontal lobe).

Following the frontal lobe is the **parietal lobe**, *which extends from the middle to the back of the skull and which is responsible primarily for processing information about touch and spatial perception, and parts of which appear to be involved in visualization* (discussed later in this chapter). Spatial perception is one component of intelligence measured on IQ tests. Damage to parts of the *right parietal association cortex* produces *unilateral neglect*, the inability to understand that the left side of your body belongs to you--thus it is "neglected" to the point where a patient may not be able "to find" their left arm, or may dress and groom only the right side of their body.

Then comes the **occipital lobe**, *at the very back of the skull, which processes visual information (the central region of the occipital lobe is called primary visual cortex and has a point-for-point mapping of the retina on its surface; surrounding the primary visual cortex is visual association cortex which is involved in higher order, more complex visual processing and perception, and along with primary visual cortex appears to store long-term visual memories)*.

Finally, in front of the occipital lobe (near the temples and ears) is the **temporal lobe**, *responsible primarily for hearing and language (Wernicke's Area is here in the left temporal lobe in most people and interacts with Broca's area in the left Frontal lobe in language processing)*. Also located in the temporal lobe is the *Inferotemporal (IT) cortex* which is involved in visual recognition. Damage in IT cortex produces *visual agnosia*--people can see and describe visual detail but they can't put the details together to recognize what it is that they are looking at. A small area of the right inferior temporal cortex known as the *fusiform face area (FFA)* appears to be specialized for face processing (Kanwisher & Yovel, 2006). Damage to the FFA produces a related, but more specific disorder, prosopagnosia or face blindness, the inability to recognize familiar faces, even those of close family members, and even one's own face in photographs (see Barton et al., 2002).

Another area involved in facial recognition in humans is the *occipital face area (OFA)* located in the lateral occipital lobe near the occipital gyrus. Brain damage in the OFA results in impaired face recognition. Research suggests that there may be a topographic face map in the OFA where adjacent areas of the human face are represented by adjacent areas of cortex in the OFA (Henriksson, et al., 2015). The FFA and OFA are interconnected and are part of a network for face processing and face recognition in the human inferior temporal and lateral occipital cortices.

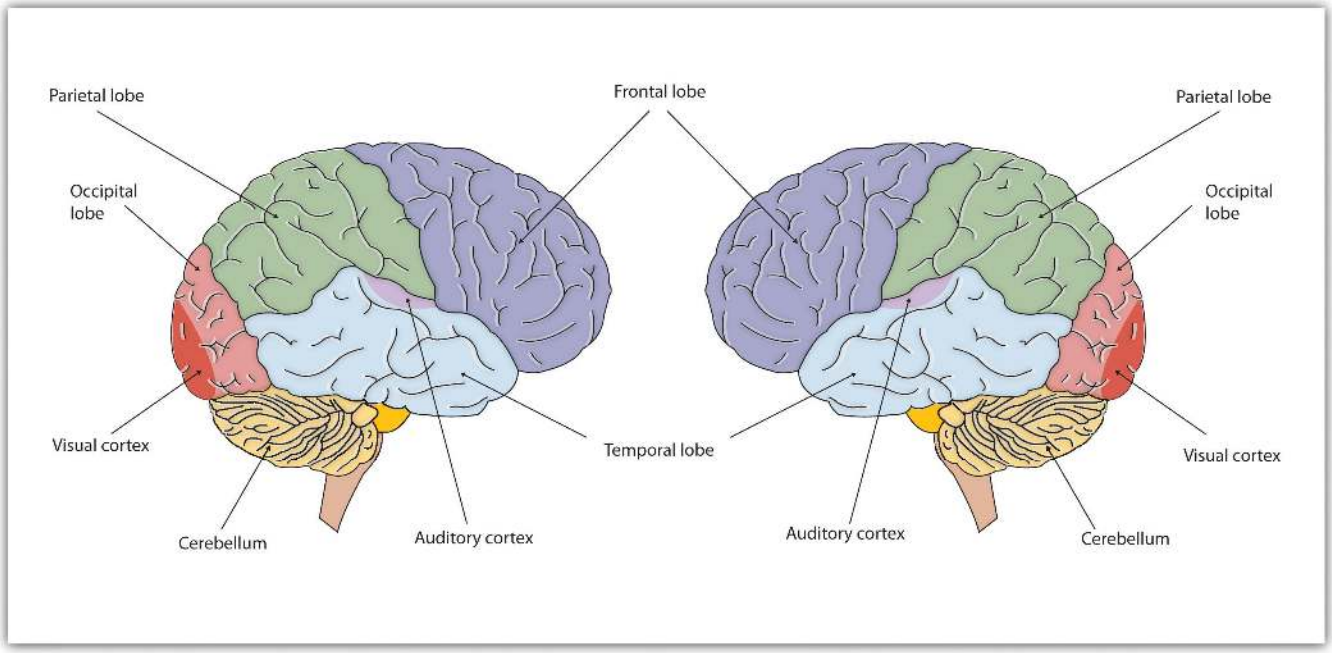


Figure **14.1.1**: The brain is divided into two hemispheres (right and left), each of which has four lobes (temporal, frontal, occipital, and parietal). Furthermore, there are specific cortical areas within the lobes that control different processes. The FFA in interior temporal lobe and the OFA in the occipital lobe are specialized for facial processing. Specialized areas of cortex for specific functions illustrate the principle of localization of function within the cerebral cortex.

14.1.3 Functions of the Cortex

When the German physicists Gustav Fritsch and Eduard Hitzig (1870/2009) applied mild electric stimulation to different parts of a dog's cortex, they discovered that they could make different parts of the dog's body move. Furthermore, they discovered an important and unexpected principle of brain activity. They found that stimulating the right side of the brain produced movement in the left side of the dog's body, and vice versa. This finding follows from a general principle about how the brain is structured, called *contralateral control*. The brain is wired such that in most cases the left hemisphere receives sensations from and controls the right side of the body, and vice versa.

Fritsch and Hitzig also found that the movement that followed the brain stimulation only occurred when they stimulated a specific arch-shaped region that runs across the top of the brain from ear to ear, just at the front of the parietal lobe (see Figure 14.1.2 below). Fritsch and Hitzig had discovered the **motor cortex**, *the part of the cortex that controls and executes movements of the body by sending signals to the cerebellum and the primary motor neurons of the spinal cord*. More recent research has mapped the motor cortex even more fully, by applying mild electronic stimulation to different areas of the motor cortex in conscious patients while observing their bodily responses (because the brain has no sensory receptors, these patients feel no pain). As you can see in Figure 14.1.2, this research has revealed that the motor cortex is specialized for providing control over the body; there is a topographic representation of the body surface on the surface of the motor cortex. In fact, the parts of the body that require more precise and finer movements, such as the face and the hands, are allotted the largest areas of motor cortical space. This makes sense since finer movements require more processing and thus more cortical tissue.

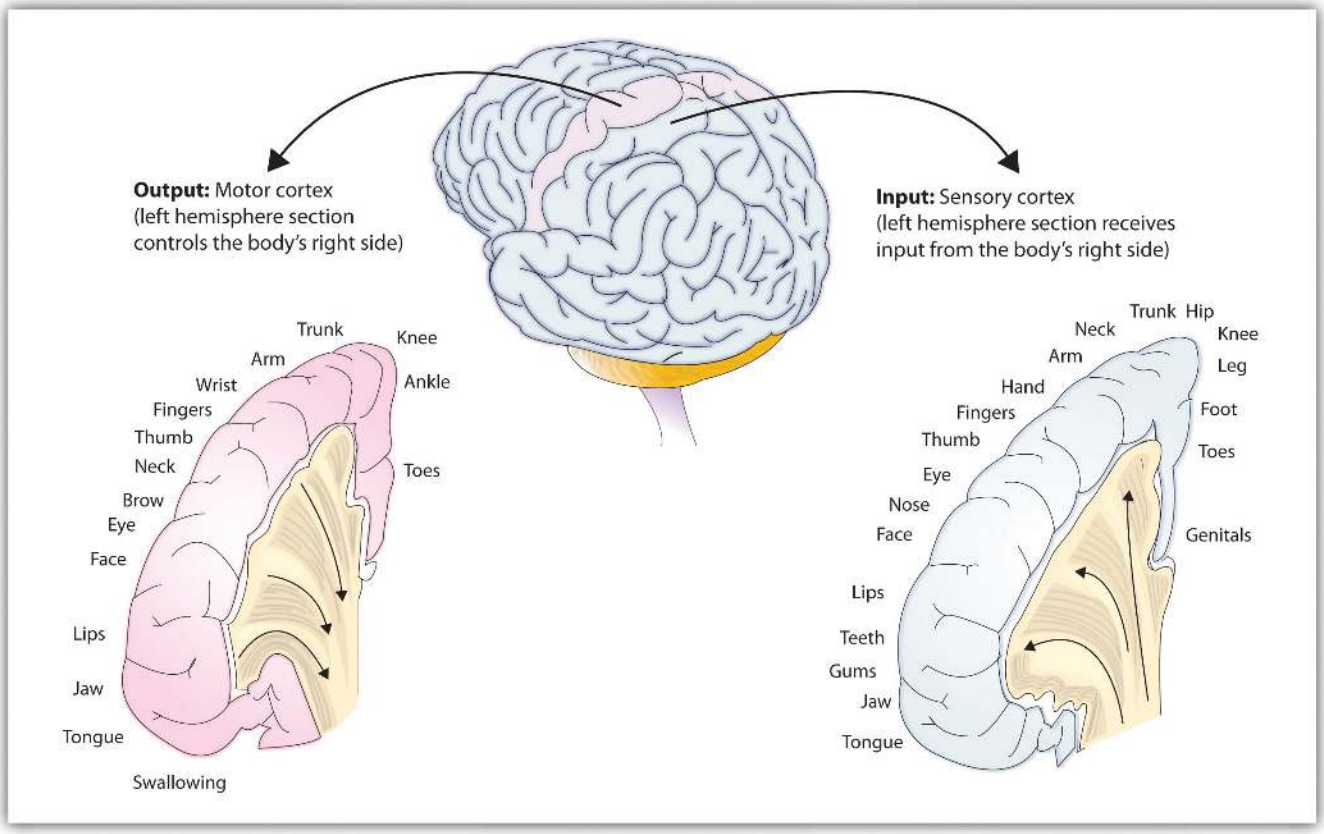


Figure **14.1.2**: The Somatosensory Cortex and the Motor Cortex.

The portion of the sensory and motor cortex devoted to receiving messages that control specific regions of the body is determined by the amount of fine movement that area is capable of performing. Thus the hand and fingers have as much area in the cerebral cortex as does the entire trunk of the body.

Just as the *primary motor cortex (the precentral gyrus, at the back of the frontal lobe)* sends out messages to the specific parts of the body, the ***primary somatosensory cortex*** *(the post-central gyrus, just behind the central fissure), an area just behind and parallel to the motor cortex, receives information from the skin's sensory receptors and the movements of different body parts*. Again, the more sensitive the body region, the more area is dedicated to it in the sensory cortex. Our sensitive lips, for example, occupy a large area in the sensory cortex, as do our fingers and genitals, whereas the trunk of the body has relatively less area in the sensory cortex devoted to it.

Other areas of the cortex process other types of sensory information. The ***visual cortex*** is *the area located in the* ***occipital lobe*** *(at the very back of the brain) that processes visual information*. If your brain were stimulated, while conscious on the operating table, by an electrode in the visual cortex, you would see flashes of light or color. Perhaps you remember having had the experience of "seeing stars" when you were hit or fell on the back of your head (this finding suggests that the luminosity and color that we attribute to light are really produced in the brain, when occipital cortex is activated, and that these sensations don't exist in the exterior world outside our heads--luminosity and color are not properties of the world, but are just brain/psychological codes for dark and colorless electromagnetic energies of various wavelengths in the visible spectrum which only become luminous and colored after they are transduced into neuron potentials which activate visual system neurons). The ***temporal lobe***, located on the lower side of each hemisphere, near your temples, *which contains* ***auditory cortex as well as Wernicke's area*** *(on the left temporal lobe of most people) responsible for hearing and language comprehension, respectively*. The inferior temporal lobe (IT cortex) also processes some visual information, providing us, as discussed above, with the ability to recognize and name the objects around us (Martin, 2007).

As you can see in Figure 14.1.2, the motor and sensory areas of the cortex account for a relatively small part of the total cortex. The remainder of the cortex is made up of ***association areas*** of cortex *in which sensory and motor information is combined and associated with our stored knowledge*. These *association areas of cortex* (to be discussed in more detail later in this chapter) are the places in the brain that are responsible for most of the things that make human beings seem human. *The association areas are involved in higher mental functions, such as learning, thinking, planning, judging, moral reflecting, figuring, and spatial reasoning.*

14.1.4 The Brain Is Flexible: Neuroplasticity

The control of some specific bodily functions, such as movement, vision, and hearing, is performed in specified areas of the cortex, and if these areas are damaged, the individual will likely lose the ability to perform the corresponding function. For instance, if an infant suffers damage to facial recognition areas in the temporal lobe (FFA), it is likely that he or she will never be able to recognize faces (Farah, Rabinowitz, Quinn, & Liu, 2000). On the other hand, the brain is not divided up in an entirely rigid way. The brain's neurons have a remarkable capacity to reorganize and extend themselves to carry out particular functions in response to the needs of the organism, and to repair damage. As a result, the brain constantly creates new neural communication routes and rewires existing ones. Neuroplasticity refers to *the brain's ability to change its structure and function in response to experience or damage*. Neuroplasticity enables us to learn and remember new things and adjust to new experiences as discussed in chapter 10 on learning and memory.

Our brains are the most "plastic" when we are young children, as it is during this time that we learn the most about our environment. On the other hand, neuroplasticity continues to be observed even in adults (Kolb & Fantie, 1989). The principles of neuroplasticity help us understand how our brains develop to reflect our experiences. For instance, accomplished musicians have a larger auditory cortex compared with the general population (Bengtsson et al., 2005) and also require less neural activity to move their fingers over the keys than do novices (Münte, Altenmüller, & Jäncke, 2002). This is because accomplished musicians through practice have eliminated synaptic connections associated with incorrect movements, making neural control more efficient (see chapter 10 on long-term depression. LTD).

Plasticity is also observed when there is damage to the brain or to parts of the body that are represented in the motor and sensory cortexes. When a tumor in the left hemisphere of the brain impairs language, the right hemisphere will begin to compensate to help the person recover the ability to speak (Thiel et al., 2006). And if a person loses a finger, the area of the sensory cortex that previously received information from the missing finger will begin to receive input from adjacent fingers, causing the remaining digits to become more sensitive to touch (Fox, 1984).

Although neurons cannot repair or regenerate themselves as skin or blood vessels can, new evidence suggests that the brain can engage in neurogenesis, *the forming of new neurons* (Van Praag, Zhao, Gage, & Gazzaniga, 2004). These new neurons originate deep in the brain and may then migrate to other brain areas where they form new connections with other neurons (Gould, 2007). This leaves open the possibility that someday scientists might be able to "rebuild" damaged brains by creating drugs that help grow neurons.

14.1.4.1 Research Focus: Identifying the Unique Functions of the Left and Right Hemispheres Using Split-Brain Patients

We have seen that the left hemisphere of the brain primarily senses and controls the motor movements on the right side of the body, and vice versa. This fact provides an interesting way to study brain lateralization—*the idea that the left and the right hemispheres of the brain are specialized to perform different functions*. Gazzaniga, Bogen, and Sperry (1965) studied a patient, known as W. J., who had undergone an operation to relieve severe seizures. In this surgery *the region that normally connects the two halves of the brain and supports communication between the hemispheres*, known as the corpus callosum, is severed. As a result, the patient essentially becomes a person with two separate brains. Because the left and right hemispheres are separated, each hemisphere develops a mind of its own, with its own sensations, concepts, and motivations (Gazzaniga, 2005).

In their research, Gazzaniga and his colleagues tested the ability of W. J. to recognize and respond to objects and written passages that were presented to only the left or to only the right brain hemispheres (see Figure 14.1.3). The researchers had W. J. look straight ahead and then flashed, for a fraction of a second, a picture of a geometrical shape to the left of where he was looking. By doing so, they assured that—because the two hemispheres had been separated—the image of the shape was experienced only in the right brain hemisphere (remember that sensory input from the left side of the body is sent to the right side of the brain). Gazzaniga

and his colleagues found that W. J. was able to identify what he had been shown when he was asked to pick the object from a series of shapes, using his left hand, but that he could not do this when the object was shown in the right visual field. On the other hand, W. J. could easily read written material presented in the right visual field (and thus experienced in the left hemisphere) but not when it was presented in the left visual field.

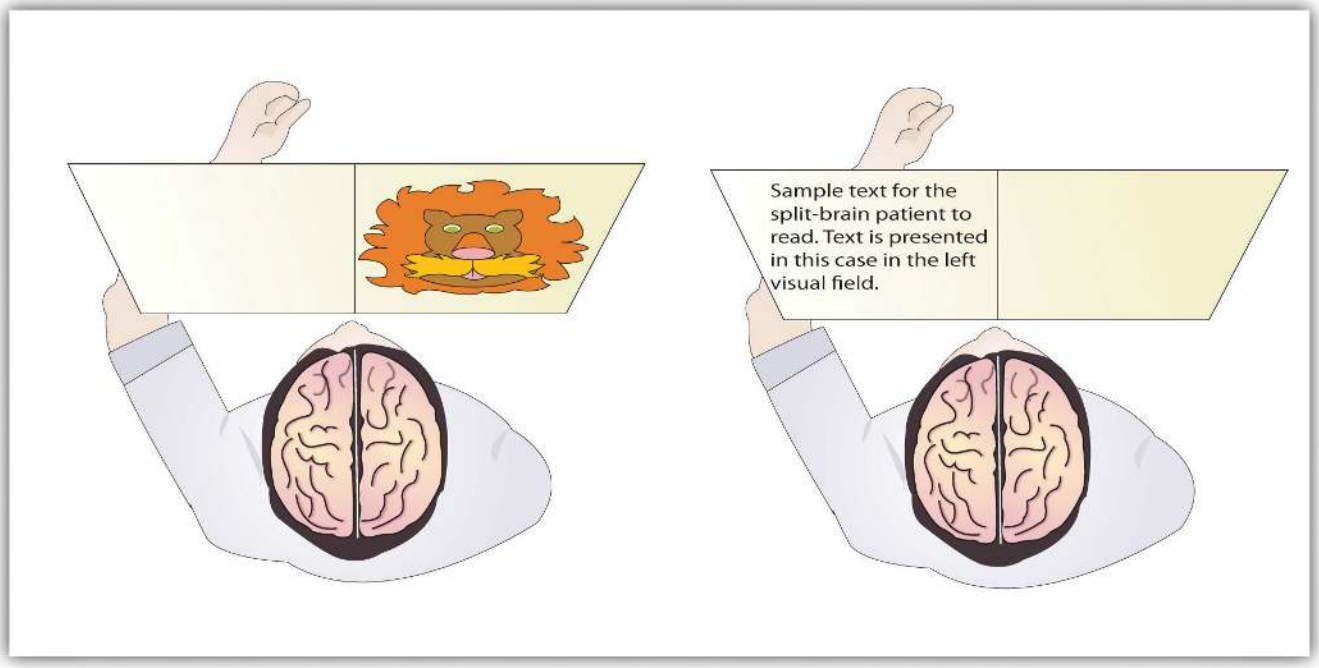


Figure **14.1.3**: Visual and Verbal Processing in the Split-Brain Patient

The information that is presented on the left side of our field of vision is transmitted to the right brain hemisphere, and vice versa. In split-brain patients, the severed corpus callosum does not permit information to be transferred between hemispheres (as it is in normal persons without brain damage), which allows researchers to learn about the functions of each hemisphere. In the perceptual test on the left, the split-brain patient could not choose which image had been previously presented because the left hemisphere cannot process visual information. In the test on the right side of Figure 14.1.3, the patient could not read the passage because the right brain hemisphere cannot process language.

This research, and many other studies following it, has demonstrated that the two brain hemispheres specialize in different abilities. In most people the ability to speak, write, and understand language is located in the left hemisphere. This is why W. J. could read passages that were presented on the right side and thus transmitted to the left hemisphere, but could not read passages that were only experienced in the right brain hemisphere. The left hemisphere is also better at math and at judging time and rhythm. It is also superior in coordinating the order of complex movements—for example, lip movements needed for speech. The right hemisphere, on the other hand, has only very limited verbal abilities, and yet it excels in perceptual skills. The right hemisphere is able to recognize objects, including faces, patterns, and melodies, and it can put a puzzle together or draw a picture. This is why W. J. could pick out the image when he saw it on the left, but not the right, visual field.

Although Gazzaniga's research demonstrated that the brain is in fact lateralized, such that the two hemispheres specialize in different activities, this does not mean that when people behave in a certain way or perform a certain activity they are only using one hemisphere of their brains at a time. That would be drastically oversimplifying the concept of brain differences. We normally use both hemispheres at the same time, and the difference between the abilities of the two hemispheres is not absolute (Soroker et al., 2005).

14.1.4.2 Why Are Some People Left-Handed?

Across cultures and ethnic groups, about 90% of people are mainly right-handed, whereas only 10% are primarily left-handed (Peters, Reimers, & Manning, 2006). This fact is puzzling, in part because the number of left-handers is so low, and in part because

other animals, including our closest primate relatives, do not show any type of handedness. The existence of right-handers and left-handers provides an interesting example of the relationship among evolution, biology, and social factors and how the same phenomenon can be understood at different levels of analysis (Harris, 1990; McManus, 2002).

At least some handedness is determined by genetics. Ultrasound scans show that 9 out of 10 fetuses suck the thumb of their right hand, suggesting that the preference is determined before birth (Hepper, Wells, & Lynch, 2005), and the mechanism of transmission has been linked to a gene on the X chromosome (Jones & Martin, 2000). It has also been observed that left-handed people are likely to have fewer children, and this may be in part because the mothers of left-handers are more prone to miscarriages and other prenatal problems (McKeever, Cerone, Suter, & Wu, 2000).

But culture also plays a role. In the past, left-handed children were forced to write with their right hands in many countries, and this practice continues, particularly in collectivistic cultures, such as India and Japan, where left-handedness is viewed negatively as compared with individualistic societies, such as the United States. For example, India has about half as many left-handers as the United States (Ida & Mandal, 2003).

There are both advantages and disadvantages to being left-handed in a world where most people are right-handed. One problem for lefties is that the world is designed for right-handers. Automatic teller machines (ATMs), classroom desks, scissors, microscopes, drill presses, and table saws are just some examples of everyday machinery that is designed with the most important controls on the right side. This may explain in part why left-handers suffer somewhat more accidents than do right-handers (Dutta & Mandal, 2006).

Despite the potential difficulty living and working in a world designed for right-handers, there seem to be some advantages to being left-handed. Throughout history, a number of prominent artists have been left-handed, including Leonardo da Vinci, Michelangelo, Pablo Picasso, and Max Escher. Because the right hemisphere is superior in imaging and visual abilities, there may be some advantage to using the left hand for drawing or painting (Springer & Deutsch, 1998). Left-handed people are also better at envisioning three-dimensional objects, which may explain why there is such a high number of left-handed architects, artists, and chess players in proportion to their numbers (Coren, 1992). However, there are also more left-handers among those with reading disabilities, allergies, and migraine headaches (Geschwind & Behan, 2007), perhaps due to the fact that a small minority of left-handers owe their handedness to a birth trauma, such as being born prematurely (Betancur, Vélez, Cabanieu, & le Moal, 1990). Interestingly, there have been a disproportionally large number of U.S. Presidents who have been left-handed, including Gerald Ford, George H.W. Bush, Bill Clinton, and Barack Obama.

14.1.5 Summary

The evolutionarily old brain—including the brain stem, medulla, pons, reticular formation, thalamus, cerebellum, amygdala, hypothalamus, and hippocampus—regulates basic survival functions, such as breathing, moving, resting, feeding, emotions, and memory.

The cerebral cortex, made up of billions of neurons and glial cells, is divided into the right and left hemispheres and into four lobes. The frontal lobe is primarily responsible for thinking, planning, memory, and judgment. The parietal lobe is primarily responsible for bodily sensations and touch. The temporal lobe is primarily responsible for hearing and language. The occipital lobe is primarily responsible for vision. Other areas of the cortex act as association areas, responsible for integrating information. The motor cortex controls voluntary movements. Body parts requiring the most control and dexterity take up the most space in the motor cortex. The sensory cortex receives and processes bodily sensations. Body parts that are the most sensitive occupy the greatest amount of space in the sensory cortex.

The brain changes as a function of experience and potential damage in a process known as plasticity. Neuroplasticity allows the brain to adapt and change as a function of experience or damage. The brain can generate new neurons through neurogenesis.

The severing of the corpus callosum, which connects the two hemispheres, creates a “split-brain patient,” with the effect of creating two separate minds operating in one person. Studies with split-brain patients as research participants have been used to study brain lateralization. The left cerebral hemisphere is primarily responsible for language and speech in most people, whereas the right hemisphere specializes in spatial and perceptual skills, visualization, and the recognition of patterns, faces, and melodies.

14.1.6 References

Barton, J. J., Press, D. Z., Keenan, J. P., & O'Connor, M. (2002). Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. *Neurology*, *58* (1), 71-78.

Bengtsson, S. L., Nagy, Z., Skare, S., Forsman, L., Forssberg, H., & Ullén, F. (2005). Extensive piano practicing has regionally specific effects on white matter development. *Nature Neuroscience*, *8* (9), 1148–1150.

Betancur, C., Vélez, A., Cabanieu, G., & le Moal, M. (1990). Association between left-handedness and allergy: A reappraisal. *Neuropsychologia*, *28*(2), 223–227.

Bodmer, W., & McKie, R. (1994). *The book of man: The quest to discover our genetic heritage*. London, England: Little, Brown and Company.

Coren, S. (1992). *The left-hander syndrome: The causes and consequences of left-handedness*. New York, NY: Free Press.

de Courten-Myers, G. M. (1999). The human cerebral cortex: Gender differences in structure and function. *Journal of Neuropathology and Experimental Neurology*, *58*, 217–226.

Dutta, T., & Mandal, M. K. (2006). Hand preference and accidents in India. *Laterality: Asymmetries of Body, Brain, and Cognition*, *11*, 368–372.

Farah, M. J., Rabinowitz, C., Quinn, G. E., & Liu, G. T. (2000). Early commitment of neural substrates for face recognition. *Cognitive Neuropsychology*, *17*(1–3), 117–123.

Fox, J. L. (1984). The brain's dynamic way of keeping in touch. *Science*, *225*(4664), 820–821.

Fritsch, G., & Hitzig, E. (2009). Electric excitability of the cerebrum (Über die Elektrische erregbarkeit des Grosshirns). *Epilepsy & Behavior*, *15*(2), 123–130. (Original work published 1870)

Gazzaniga, M. S., Bogen, J. E., & Sperry, R. W. (1965). Observations on visual perception after disconnexion of the cerebral hemispheres in man. *Brain*, *88*(2), 221–236.

Gazzaniga, M. S. (2005). Forty-five years of split-brain research and still going strong. *Nature Reviews Neuroscience*, *6*(8), 653–659.

Geschwind, N., & Behan, P. (2007). *Left-handedness: Association with immune disease, migraine, and developmental learning disorder*. Cambridge, MA: MIT Press.

Gibson, K. R. (2002). Evolution of human intelligence: The roles of brain size and mental construction. *Brain Behavior and Evolution 59*, 10–20.

Gould, E. (2007). How widespread is adult neurogenesis in mammals? *Nature Reviews Neuroscience 8*, 481–488. doi:10.1038/nrn2147

Harris, L. J. (1990). Cultural influences on handedness: Historical and contemporary theory and evidence. In S. Coren (Ed.), *Left-handedness: Behavioral implications and anomalies*. New York, NY: Elsevier.

Henriksson, L., Mur, M., & Kriegeskorte, N. (2015). Faciotopy—A face-feature map with face-like topology in the human occipital face area. *Cortex*, *72*, 156-167.

Hepper, P. G., Wells, D. L., & Lynch, C. (2005). Prenatal thumb sucking is related to postnatal handedness. *Neuropsychologia*, *43*, 313–315.

Ida, Y., & Mandal, M. K. (2003). Cultural differences in side bias: Evidence from Japan and India. *Laterality: Asymmetries of Body, Brain, and Cognition*, *8*(2), 121–133.

Jones, G. V., & Martin, M. (2000). A note on Corballis (1997) and the genetics and evolution of handedness: Developing a unified distributional model from the sex-chromosomes gene hypothesis. *Psychological Review, 107*(1), 213–218.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1476), 2109-2128.

Kolb, B., & Fantie, B. (1989). Development of the child's brain and behavior. In C. R. Reynolds & E. Fletcher-Janzen (Eds.), *Handbook of clinical child neuropsychology* (pp. 17–39). New York, NY: Plenum Press.

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology, 58*, 25–45.

McKeever, W. F., Cerone, L. J., Suter, P. J., & Wu, S. M. (2000). Family size, miscarriage-proneness, and handedness: Tests of hypotheses of the developmental instability theory of handedness. *Laterality: Asymmetries of Body, Brain, and Cognition, 5*(2), 111–120.

McManus, I. C. (2002). *Right hand, left hand: The origins of asymmetry in brains, bodies, atoms, and cultures*. Cambridge, MA: Harvard University Press.

Miller, G. (2005). Neuroscience: The dark side of glia. *Science, 308* (5723), 778–781.

Münte, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience, 3*(6), 473–478.

Peters, M., Reimers, S., & Manning, J. T. (2006). Hand preference for writing and associations with selected demographic and behavioral variables in 255,100 subjects: The BBC Internet study. *Brain and Cognition, 62*(2), 177–189.

Soroker, N., Kasher, A., Giora, R., Batori, G., Corn, C., Gil, M., & Zaidel, E. (2005). Processing of basic speech acts following localized brain damage: A new light on the neuroanatomy of language. *Brain and Cognition, 57*(2), 214–217.

Springer, S. P., & Deutsch, G. (1998). *Left brain, right brain: Perspectives from cognitive neuroscience* (5th ed.). A series of books in psychology. New York, NY: W. H. Freeman/Times Books/Henry Holt & Co.

Thiel, A., Habedank, B., Herholz, K., Kessler, J., Winhuisen, L., Haupt, W. F., & Heiss, W. D. (2006). From the left to the right: How the brain compensates progressive loss of language function. *Brain and Language, 98*(1), 57–65.

Van Praag, H., Zhao, X., Gage, F. H., & Gazzaniga, M. S. (2004). Neurogenesis in the adult mammalian brain. In *The cognitive neurosciences* (3rd ed., pp. 127–137). Cambridge, MA: MIT Press.

14.1.7 Attributions

Content, including figures, adapted by Kenneth A. Koenigshofer, Ph.D., from Introduction to Psychology (Saylor Foundation), Chapter 3. https://saylordotorg.github.io/text_...r-thought.html; license **(CC BY-NC-SA 3.0) https://creativecommons.org/licenses/by-nc-sa/3.0/.** Some material on functions of cortical areas was added by Kenneth A. Koenigshofer.

This page titled 14.1: Overview of Brain Mechanisms in Intelligence, Language, and Cognition is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

14.2: Cognition, Intelligence, and Language as Psychological Adaptations

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Explain what is meant by the claim that psychological processes, including intelligence, serve adaptive movement
2. Explain why intelligence and cognition are examples of psychological adaptations
3. Explain why natural selection requires recurrent, across-generation conditions to generate adaptations
4. Describe some of the environmental regularities, including several universal recurrent relational features of the world, that may have been genetically incorporated into brain organization generating intelligence, including general "fluid" intelligence
5. Discuss Spearman's definition of "g"
6. Describe the brain areas that are involved in intelligence, broadly defined as brain systems that guide behavior toward successful adaptation
7. Discuss the role of Shepard's concept of "genetic internalization" of biologically important environmental regularities in the evolution of intelligence
8. Discuss three abstract relational regularities of the world which may have been genetically internalized into the brain operations of general intelligence
9. Describe vector coding
10. Discuss mental models, prospection, and intelligence, and list brain structures most directly involved in the formation of mental models/cognitive maps
11. Describe the neural correlates of differences in intelligence across species and the P-FIT theory of differences in intelligence among humans
12. Describe the Broad Visual Perception Factor and the role it plays in the intelligent control of movement/behavior

14.2.1 Intelligence, Cognition, Language, and Biological Adaptation

by Kenneth A. Koenigshofer, PhD.

14.2.1.1 Overview

After a general review of cortical functions in the previous section, we now examine higher cognitive functions as modes of achieving biological adaptation to the environment. In a previous chapter, you learned about the neural mechanisms of movement. Let's start this module with the premise that, to be useful, movement must be effectively guided in order for it to be responsive to environmental demands and adaptive opportunities. The senses are the first tier in the systems that guide movement (even some single-celled organisms can sense harmful chemicals in their fluid environment and then use their cilia to swim away from them, showing how even a simple guidance system for movement improves adaptation). Imagine you were one of our evolutionary ancestors; if you couldn't sense an approaching saber-toothed tiger and make accurate judgments about its distance and speed, it is obvious that you would be unable to effectively run or hide from it. If you couldn't feel fear and understand danger, you would not be *motivated* to run or hide. Without senses and perception, it would be impossible for you to organize movements to find food or secure a mate.

However, in addition to the senses, perception, and emotion as guidance systems for movement, higher order functions of the brain, some of which were discussed in the chapter on the mechanisms of movement (e.g. premotor cortex, parietal cortex), are also critically involved in the more advanced control of action. In this chapter, we consider intelligence and cognition as part of the sophisticated systems that guide movement in animals and humans to maximize their adaptation to the environment.

We will also consider other research on intelligence with a very different focus. Early work by psychologists interested in human intelligence took a more practical approach. These psychologists in the first part of the twentieth century focused on the development of methods of measuring intelligence in humans. This approach to the study of psychological attributes is called

psychometrics. Though early research emphasized the creation and refinement of intelligence tests, rather than the biological origins and functions of intelligence, psychometric analysis of the performance of large groups of people on intelligence tests (Spearman, 1904, 1925) ultimately led to theories of the structure of human intelligence (Carroll, 1993; Cattell, 1987). These theories are important to biological psychology because they have influenced much of the modern thinking and research on the genetic and brain mechanisms involved in human thinking and intelligence. Additional details about psychometrics and theories of intelligence can be found in Supplement 1, Traditional Models of Human Intelligence, in this chapter.

More recently, intelligence research by psychologists and neuroscientists has expanded to include the study of brain mechanisms underlying intelligence in humans and in other species. This approach has led to conceptions of intelligence and cognition in a broader biological and evolutionary context. On this view, thinking, intelligence, and language are products of evolution, just like other genetic traits of organisms. Thinking, intelligence, and language exist because they helped ancestral humans meet the challenges presented by the environment, challenges which drove natural selection. In short, thinking, intelligence, and language are biological in origin, biological in function, and, of course, they are generated in a biological organ, the brain.

One of the key evolutionary trends responsible for the evolution of intelligence and cognition was genetic incorporation of information about biologically significant regularities of the world into the circuitry and operations of the brain. As a consequence, the brain is equipped with many "cognitive instincts" and innate implicit knowledge about biologically important, enduring regularities of the terrestrial environment, forming the groundwork for much of our genetically evolved intelligence as a species.

This view contrasts with and rejects the "blank slate" view of the mind/brain assumed in the Standard Social Science Model (SSSM), the set of assumptions that much of psychology and the social sciences were founded upon--the view that humans lacked any innate psychological nature and that the mind and brain were essentially blank at birth, leaving it to learning and culture to form human behavior free of genetics and biological evolution. In this chapter, we take a different point of view--that biology, genes and evolution, are the primary determinants of human cognition and behavior, in interaction with learning and cultural influences, which themselves are ultimately biological in nature.

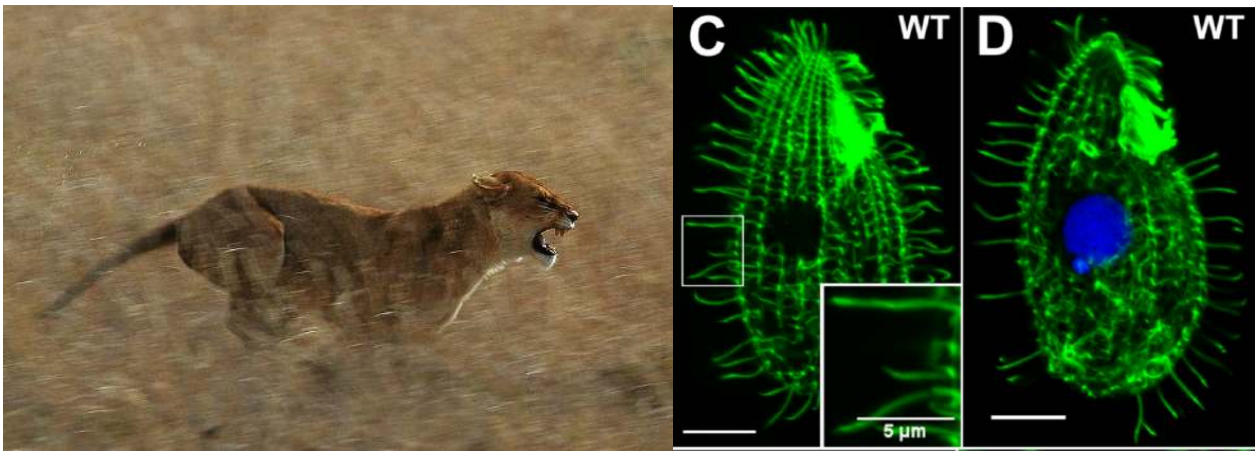


Figure **14.2.1**: (Left) Lioness hunting in the Serengeti region of Tanzania. Its mechanisms of sensation and perception, emotion, and intelligence are pitted against these same guidance systems for movement found in its prey. Predator-prey interactions may have been an escalating evolutionary impetus for the development of intelligence in both predator and prey. (Right) A single-celled organism featuring its cilia which it uses to move toward favorable parts of its fluid environment and away from harmful regions, thereby facilitating its adaptation. Approach to beneficial elements of the environment and avoidance of and withdrawal from harmful elements is a primary rule governing guidance systems for movement in all motile species. (Image on left is from Wikimedia Commons, https://commons.wikimedia.org/wiki/F..._saturated.jpg, by Schuyler Shepherd; licensed under the Creative Commons Attribution-Share Alike 2.5 Generic license. Caption by Kenneth A. Koenigshofer, PhD. Image on right is from Wikimedia Commons, https://commons.wikimedia.org/wiki/F...C3%BAcleos.png, by Seixas C, Cruto T, Tavares A, Gaertig J, Soares H; licensed under the Creative Commons Attribution 4.0 International license. Caption by Kenneth A. Koenigshofer, PhD).

14.2.1.2 Thinking and intelligence serve the adaptive organization of movement

As argued in the chapter on evolution and genetics, nothing makes sense in psychology except in the light of evolution. This claim is especially relevant for a course in biological psychology. When we consider thinking and intelligence from a biological and evolutionary perspective, it is important to ask what functions they perform for the organism. The brain does all kinds of complex processing, but it is important to understand that for that processing to have any effect on the environment, brain activity must ultimately converge onto motor neurons in the spinal cord ("the final common path") that stimulate the muscles to produce movement--behavior (see the chapter on movement). Like every aspect of our psychology and its corresponding brain activity, the function of thinking and intelligence is to generate adaptive behavior, movement, to successfully meet environmental challenges to survival and reproduction and to exploit opportunities, thereby increasing biological fitness.

Consider plants, for a moment. They don't have intelligence or thinking--they don't need to, because they depend very little upon movement for survival and reproduction--instead, water and their source of energy, sunlight, come to them (plants can move slowly in a limited way; they grow toward sunlight and roots grow toward water sources, but contrast this with all the complex social behaviors modern humans engage in to get food and water--e.g. agriculture, supply chains, plumbing, dams, water companies, etc.). Neither do plants have to flee or hide from predators; they protect themselves with thorns or poisons or simply grow back if partially eaten. Neither do they need to move for reproduction--wind and insects and other animals carry the reproductive cells for them (imagine if this were true for humans! What would happen to human courtship, dating, and romantic love? The purpose of which, from a biological perspective, is to get egg and sperm cells together).

But none of this is true for animals. For animals, including the human animal, movement is key to survival and reproduction, and, as noted above, the movement must be guided to form patterns of action that solve adaptive problems and exploit adaptive opportunities. According to Darwin's theory of evolution, which he explained as "descent with modification," the roots of human cognition and intelligence lie deep in our species' evolutionary ancestry. Evidence for this can be found in Darwin's principle of the "continuity of species" applied to psychology, as shown by the great degrees of



Figure **14.2.2**: Movement is essential for survival and reproduction in animals. The great Wildebeest migration with over 2 million animals follows the rains to lush new feeding grounds. (Image from Wikimedia Commons; https://commons.wikimedia.org/wiki/F...n_crossing.jpg by Naturaltracksafaris; licensed under the Creative Commons Attribution-Share Alike 4.0 International license. Caption by Kenneth A. Koenigshofer, PhD).

similarity in the guidance and control systems for movement and their underlying brain mechanisms across mammalian species--for example, all mammals have similar brain structures, including limbic structures and cerebral cortex. All species respond to potentially harmful stimuli with threat, attack, or escape. All mammals respond to sexual signals by moving toward and making contact with their source. Each suggests common behavioral control mechanisms evolutionarily conserved across species

But intelligence also exists in non-mammalian species. Intelligence in distantly related species including some birds, such as Corvids (e.g. ravens, jays, crows) and parrots, and in invertebrates such as the octopus and cuttlefish (Adams & Burbeck, 2012; Mather, 2019; Mather & Dickel, 2017), illustrate convergent evolution, suggesting the high adaptive utility of intelligence in diverse ecological niches. Intelligence and thinking, like all psychological processes, evolved to serve the organization of movement to make behavior adaptive, successful, in the Darwinian struggle for survival and reproduction (Koenigshofer, 2011, 2016).

14.2.1.3 Cognition and Intelligence are psychological adaptations

Intelligence and thinking (i.e. cognition) are psychological adaptations evolved by natural selection over millions of years. A psychological adaptation is a psychological or behavioral trait that has developed through evolutionary processes such as natural and sexual selection and which is encoded into a species' DNA (see Ellis & Ketelaar, 2002). Dicke and Roth (2016) offer a comprehensive definition of intelligence recognizing its role in adaptation to the environment in a wide range of animals and humans:

> According to the majority of behaviorists and animal psychologists, ‘intelligence’ can be understood as mental or behavioral flexibility or the ability
> of an organism to solve problems occurring in its natural and social environment, culminating in the appearance of novel solutions that are not
> part of the animal’s normal repertoire. This includes forms of associative learning and memory formation, behavioral flexibility and innovation rate, as well as abilities requiring abstract thinking, concept formation and insight.

By contrast, Colom et al. (2010) define intelligence simply as "a general mental ability for reasoning, problem solving, and learning."



Figure **14.2.3**: Because animals cannot photosynthesize, they must move in order to obtain their sources of energy found in plants and other animals--animals which can hide, run, and fight back--unlike the plants' source of energy, the sun. Movement is organized by brain circuitry which has been configured by natural selection over evolutionary time to solve adaptive problems. Here a hyena on the plains of Africa solves the problem of getting sufficient energy to survive and ultimately reproduce its genes. (Image from Wikimedia Commons; https://commons.wikimedia.org/wiki/F...il_breghys.jpg; by Brwynog; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

A typical definition of general intelligence (discussed below) shows how intelligence and general intelligence are terms sometimes used interchangeably, even by psychologists. General intelligence is often defined as "ability to reason deductively and inductively, to think abstractly, use analogies, to synthesize information, and to apply that information to new domains" (Gottfredson, 1997; Neisser et al., 1996).

In spite of some differences in definition among researchers, there is unanimous agreement that intelligence and cognition are properties of brain processes which have been shaped and refined over millions of years of evolution. Once again, Darwin's principle of the "continuity of species" leads us to suspect that there are many similarities between intelligence and cognition in humans and these same processes in animals, especially mammals. To help us understand human intelligence and cognition and their evolutionary origins, many biological psychologists study these processes in other animals, including apes, monkeys, rats,

dogs, parrots, and Corvids, such as ravens, jays, and crows. Animals often surprise us with the sophistication of their emotional sensitivity and their intelligence even though they lack language (Huber & Gajdon, 2006; Wasserman, et al., 2006). An interesting anatomical feature of the cerebral cortex is what we might call microprocessors or "chips" in the human cortex. These are columnar structures, known as **cortical columns**, containing approximately 3,000 neurons each. There are approximately 150,000 cortical columns in the human cerebral cortex (Hawkins, et al., 2017) . These structures are very similar from one part of the cortex to another and from one mammalian species to the next and suggest highly generalized computational functions in cerebral cortex, whereas specialized modules or genetically dedicated circuits for processing of emotional and motivational information are localized to subcortical regions of the brain, shared by all mammals (Panksepp & Panksepp, 2000).

Intelligence and cognition (i.e. thinking) are exceedingly complex processes. As a consequence, the brain mechanisms involved in these processes are not well understood. As you can see, even definitions of these processes vary widely in the scientific community. However, all of the definitions above include terms such as reasoning, abstract thinking, insight, or problem solving, processes which themselves are not well defined or clearly understood by psychologists and neuroscientists. This highlights the fact that psychologists and neuroscientists are still in the early stages of gaining a real understanding of how intelligence and thinking work and what neural processes in the brain produce them. Perhaps this is why textbooks in biological psychology have tended to give little or no attention to these processes even though they are perhaps the most wondrous and important products of brain activity.

14.2.1.4 Psychometrics, the Measurement of Human Intelligence, and "g"

Early in the history of psychology, research about intelligence was not focused upon its evolutionary origins or its adaptive functions. Instead, psychologists concentrated their research efforts on more practical problems such as the measurement of intelligence for classifying students and military recruits. Major study of human intelligence by psychologists began in the early 1900's when psychologists focused on the measurement of intellectual abilities, an approach known as psychometrics. This research approach resulted in the development of intelligence tests and the concept of IQ (intelligence quotient). Additional research in psychometrics, focused on mathematical analysis of performance scores of large populations on IQ tests using a method called factor analysis, which analyzes patterns of correlations between measures. This research led to the finding that people who did well on any particular measure of intellectual performance tended to do well on all such measures. This led psychologist, Charles Spearman (1904), to hypothesize the existence of a single, unitary "g" factor (g for general) in human intelligence. This g factor referred to a general mental capacity which was hypothesized to be involved to varying degrees in multiple mental abilities including verbal, spatial, mechanical, computational, and other performances. "This *g* factor provides an index of the level of difficulty that an individual can handle in performing induction, reasoning, visualization, or language comprehension tests" and the g factor accounts for "more than half" of the variation among individuals "in a cognitive test or task" (Jung & Haier, 2007, p. 492). The g factor came to be known among psychologists as "general intelligence." Spearman believed that "g" was most closely related to what he called the "eduction [from the Latin root "educere" which means to "draw out"] of relations and correlates," important in inductive and deductive logic, grasping relationships, inferring rules, and recognizing differences and similarities. In the sections below, we examine some key components of intelligence, relevant brain mechanisms, and how intelligence might have evolved.

14.2.1.5 Conditions required for evolution of psychological adaptations

Understanding the evolutionary origins of intelligence and thinking may offer insights about how they work and about their underlying brain mechanisms. First of all, remember the claim made above that intelligence and cognition are psychological adaptations, or more correctly, a collection of psychological adaptations, each evolved by natural selection to process certain types of information in specific ways. A key principle of how evolution works is relevant here. Evolution by natural selection can only capture information about features of the world that are present generation after generation. This means that any innate brain mechanisms and the innate "instincts" produced by these mechanisms must have been formed as adaptations to regularly recurrent properties of the environment. This is because natural selection requires many generations to work a genetic change in a population (see chapter on evolution and genetics). Without consistent (statistically regular) selection criteria present over many generations, natural selection cannot fashion complex adaptations. For example, if the consistent force of gravity had not been regularly present over generations, land-living vertebrates could not have evolved bones with tensile strength sufficient to adapt to the downward pull of gravity. In this case, these animals would be incapable of movement or even standing. Another example is color vision. The color of many fruits indicates their state of ripeness. Without the ripeness of fruit being regularly signaled by its color, generation after generation, color vision would not have evolved in fruit-eating primates. Stated another way, information about the strength of

gravity or about the correlation between wavelength of light and ripeness of fruit could not have been incorporated by natural selection into the evolved adaptations of organisms unless these features of the world were consistently present, and affecting rates of survival and reproduction, generation after generation.

By this reasoning, long-term, across-generation regularities of the world that have adaptive significance should be expected to play a special role in the evolution of the mind/brain (just like they do in the evolution of anatomical and physiological adaptations). As evolutionary psychologists, Tooby and Cosmides (1992, p. 69), state: “Long-term, across-generation recurrence of conditions ... is central to the evolution of adaptations." Kaufman et al. (2011, p.213) express a similar idea. They state, "Evolutionary psychologists sometimes argue that a class of situations must be relatively narrow to exert consistent selection pressure, but this claim is insufficiently justified. Any regularity in the environment can exert selection pressure if it poses a challenge or an opportunity to the organism . . ." And as former Stanford psychologist, Roger Shepard, put it, there has evolved "a mesh between the principles of the mind and the regularities of the world" (Shepard, 1987a). This means that the organization of the mind reflects regularities of the world, because of the fact that natural selection operates on conditions that repeat, that are regular, over long periods of time. So, when thinking about the evolution of brain and mind, we should be on the lookout for "instincts" that innately predispose the brain to process environmental inputs using innate principles derived from adaptively significant regularities of the world. In other words, the brain is filled with circuitry that operates by genetically programmed rules derived from information about enduring environmental regularities.

Examples of such long-term regularities or enduring properties of the world are: the widespread presence of environmental stimuli that can damage body tissue; that some things in the environment contain sources of nutrition while others are poisonous; that some potential mates are more likely than others to be healthy and fertile; that snakes and spiders are frequently dangerous; that sugars and fats are concentrated sources of energy; that high status in the social group in social animals gives greater access to resources including mates; and so on. In response to biologically important statistical regularities in the world such as these, neural circuitry has evolved in animal and human brains that causes: withdrawal from harmful stimuli in response to pain; approach toward and consumption of sources of nourishment and avoidance of potential foods that cause illness; powerful innate drives and emotions to mate with sexual partners possessing features indicative of health and reproductive potential (Ellis & Ketelaar, 2002, p. 162; Gangestad & Simpson, 2000); a genetic readiness in humans and other primates to learn fear of snakes and spiders (DeLoache & LoBue, 2009); a startle response to sudden loud noises; status seeking and sensitivity to social standing in social primates; a preference in humans for sugars, salt, and fats; and so on (see Tooby & Cosmides, 2015).





Figure 14.2.4: (Left) For most people, spiders like this one induce a strong aversive reaction motivating avoidance and withdrawal behavior, an example of a psychological adaptation evolved to protect us from possible poisoning from this class of stimulus. (Right) Humans have evolved a preference for fats, sweets, and salt, a psychological adaptation from the Pleistocene which can be harmful to health if followed too frequently in the modern urban environment full of fast food restaurants--a very different environment from that of our hunter-gatherer ancestors. (Image on left from Wikimedia Commons, https://commons.wikimedia.org/wiki/F...ia,_Brazil.jpg; by Alex Popovkin, Bahia, Brazil; licensed under the Creative Commons Attribution 2.0 Generic license. Image on right from Wikimedia Commons, https://commons.wikimedia.org/wiki/F..._Bigntasty.jpg; by مانفی in Persian; licensed under the Creative Commons Attribution-Share Alike 4.0 International license. Captions by Kenneth A. Koenigshofer, PhD).

Each of these innate behaviors evolved in response to specific regularities of the world, "recurrence of conditions," that consistently had important adaptive consequences--consequences for survival and reproduction--over countless generations of evolution.



Figure 14.2.5: According to evolutionary psychologists, Cosmides and Tooby (2003), forming and maintaining friendships is important to us today because this was one of many adaptive problems that our human ancestors encountered and had to solve in our Pleistocene past. Friendships led to alliances that were important in securing resources within Pleistocene bands of individuals that depended upon one another for survival. For this reason, we are motivated to find friends and having them feels good, reinforcing the behavior. Information processing by the brain necessary for formation and maintenance of relationships is one of the topics studied by biological psychologists interested in intelligence and social cognitive neuroscience. (Image from Wikimedia Commons, https://commons.wikimedia.org/wiki/F...923360765).jpg; by Rod Waddington; licensed under the Creative Commons Attribution-Share Alike 2.0 Generic license. Caption by Kenneth A. Koenigshofer, PhD.).

Evolutionary psychologists have identified some additional recurrent problems that had to be regularly solved by our human ancestors: " . . . winning social support from [group] members, remembering the locations of edible plants, hitting game animals with projectiles, …, recognizing emotional expressions, protecting family members, maintaining mating relationships, …, assessing the character [and social valuations] of self and others, causing impregnation, acquiring language, maintaining friendships, thwarting antagonists, and so on" (Cosmides and Tooby 2003, p. 59).

In each case, evolutionary psychologists suspect that brain mechanisms have evolved to contribute to solutions of each of these adaptive problems regularly present generation after generation. On this view, cognition is believed to consist of “many mental rules that are specialized for reasoning about various evolutionarily important domains, such as cooperation, aggressive threat, parenting, disease avoidance, predator avoidance, object permanence, and object movement” (Cosmides & Tooby, 1992, p. 179). But these are not the only kinds of recurrent, across-generation conditions that have been genetically represented by natural selection in brain circuitry. Such regularities can be much more abstract and widely distributed throughout the environment.

14.2.1.6 Innate (instinctual) Knowledge of Universal Regularities of the World: The Foundations of Intelligence

One approach to understanding the nature and origins of intelligence is the idea that natural selection incorporated information about “general—perhaps even universal—properties” of the world (Shepard, 1992, p. 500) into brain organization over evolutionary time (Shepard, 1992, 1994, 2001). The core idea is that natural selection has selected genes that equip human and animal brains with fundamental information about how the physical world is organized, information that is essential for intelligent adaptation to the environment. This inborn information provides innate "knowledge" about biologically significant, universal properties of the world which have existed continuously for countless generations. This approach also may explain fundamental principles of how our thinking processes are organized and how they evolved.

According to Shepard (1987a, 1992, 1994, 2001), natural selection has favored the genetic incorporation or “internalization” of biologically significant regularities of the world, “whether . . . within a particular species' local niche or throughout all habitable environments. . . . Genes that have internalized these pervasive and enduring facts about the world should ultimately prevail over genes that leave it to each individual to acquire such facts by trial and possibly fatal error” (Shepard, 1994, p.2).

What Shepard means by "genetic or evolutionary internalization" is that natural selection has favored genes that incorporate information into the brain about specific, biologically important, universal regularities of the terrestrial environment. Consequently, information about these enduring environmental regularities, such as the fact that the sun rises and sets approximately every 24 hours, becomes genetically encoded into genes which organize specific brain circuitry or other properties of the brain. In the case of the roughly 24 hour cycle of light and dark, we see daily rhythms of sleep and wakefulness controlled by inborn brain mechanisms and by cycles of melatonin release from the **pineal gland**, a small gland near the center of the brain, which acts on the hypothalamus, involved in the control of these daily cycles. We also see daily cycles of a number of other hormones as well. The release of these hormones is regulated by the **pituitary gland** which itself is under the control of genetically organized circuitry within the **suprachiasmatic nucleus of the hypothalamus**. Animals and humans have many so-called circadian (about a day) rhythms. These rhythms have originated from "genetic or evolutionary internalization" of information about the repeating daily cycle of light and dark into brain organization as a result of natural selection (see chapter on sleep).



Figure **14.2.6**: A woman sleeping. She does not know that daily cycles of sleep and wakefulness and other circadian rhythms are the result of genetic internalization over the course of evolution of an enduring regularity of the physical environment, the rotation of Earth on its axis (see text below and the chapter on sleep). (Image from Wikimedia Commons; https://commons.wikimedia.org/wiki/F...g_in_Sepia.jpg; by Massimo Danieli; licensed under the Creative Commons Attribution-Share Alike 2.0 Generic license. Caption by Kenneth A. Koenigshofer, PhD.).

Some of these enduring properties of the environment which have been "genetically internalized" by natural selection (Shepard, 1992, 1994, 2001) into the brain's functioning are relatively concrete--such as the approximately 24 hour cycle of light and dark (mentioned above). Others are “three-dimensional, locally Euclidian space” with a "gravitationally conferred unique upward direction," and one-dimensional time with a "thermodynamically conferred unique forward direction" (Shepard, 1992, p. 500). As noted above, circadian rhythms built into us and other species are the result of this "genetic internalization" of the daily light/dark cycle which characterizes the terrestrial environment in all but the polar regions of Earth. The genetic internalization of information about three-dimensional space is evidenced by the presence of "**place cells**" and spatial cognitive maps in the **hippocampus**, as we discussed earlier in Chapter 10 on learning and memory, and involvement of the **parietal cortex** in spatial processing. And one-dimensional time has been internalized into the functional organization of the **cerebellum** where cells which monitor time have been identified (Irvy, et al., 1989; Irvy & Spencer, 2004; Hayashi, et al., 2014). Some additional regularities of the world that have been genetically internalized by evolution may help solidify the point. Objects in the universe tend to have certain shapes, reflectance, and paths of motion and these statistical regularities of the visual world have been "genetically internalized" (Shepard, 1992) by natural selection into the human **visual system** and the visual systems of many other species (Anderson, 2009). Other biologically significant regularities of the world, including regularities of the social environment, leading to evolution of other quite specific components of intelligence (including social intelligence; see sections 18.11 and 18.14) are the recurrent problems of survival and reproduction mentioned above. All of these are examples of how the principles of the mind reflect some of the more concrete regularities of the world (recall the quote above from Shepard).

14.2.1.7 Natural Selection Favored Representations of Abstract Relational Regularities: Evolution of General Intelligence

In addition to the above relatively concrete regularities of the physical world, there are other enduring and pervasive regularities of the physical environment that are quite abstract and relational and information about these have also been "genetically internalized" into brain operations by natural selection. These relational regularities of the world are **cause-effect**, **similarity, and predictive relations** between objects and events. These relations are so ubiquitous and so familiar to us that often we barely notice them. Yet they are essential to our intelligence. Natural selection has also "genetically internalized" information about these adaptively important relational regularities of the world into our brain organization. As a consequence, our brains are innately predisposed to readily understand cause-effect, to recognize similarity relations which we use to form categories and concepts, and to look for and to find predictive relations among events in the world (Koenigshofer, 2017). This means that we are born with many genetically encoded "instincts" about the abstract, yet fundamental properties of the physical environment. In effect, we inherit some general, universal principles about how the world works. These "cognitive instincts" comprise much of our genetically inherited "general" intelligence ("g," see discussion of "g" above, and also see section 18.11) as a species--properties of the brain derived from eons of evolution by natural selection (Koenigshofer, 2017). These instincts are other ways in which "there has evolved a mesh between the principles of the mind and the regularities of the world" (Shepard, 1987a). Intelligent action depends upon comprehension of causality, similarity, and predictive relations in one's environment.

14.2.1.7.1 Causality

Because the world is governed by natural causal laws, **cause-effect** relations between things regularly occur everywhere in the environment and have consistently existed since the beginning of evolutionary time. Such regularities, though abstract and relational, can drive natural selection (Koenigshofer, 2017; see Cosmides & Tooby, 1992, p. 48; Kaufman et al., 2011, p. 213). As a result, this relational property of the world has been genetically internalized, just like information about three dimensional space and the cycles of light and dark, so that our brains (and the brains of many other species) are innately "tuned" to look for and to understand cause-effect in the environment. Penn and Povinelli (2007, p. 98) state: "Animals of all taxa have evolved cognitive mechanisms for taking advantage of causal regularities in the physical world." Evidence for an innate understanding of causal relations in humans comes from experiments showing that children, from a very early age, readily understand cause-effect and use statistical regularities to tease out causes of specific events (see Gopnik, 2010, 2012; Penn & Povinelli, 2007). The inborn predisposition to understand cause-effect as a general property of the world is a central component of human and animal general intelligence and cognition (Koenigshofer, 2017). The famous 18th century British philosopher, David Hume (1748/1988), recognized this fundamental property of thought when he wrote: "All reasonings concerning matter of fact seem to be founded on the relation of Cause and Effect." A modern definition expresses the same idea: "causal reasoning [is] the ability to identify the functional relationship between a cause and its effect (Amodio, 2019).

From the standpoint of biological adaptation, implicit knowledge about causality allows intelligent creatures to navigate and exploit an enormous variety of complex causal relations, to make causal inferences and predictions, which benefit adaptation to the problems and opportunities presented by the environment. A chimpanzee who understands that putting a long stick into a termite hill will cause termites to bite onto the stick gets a lot more juicy termites to eat than a chimpanzee who does not understand this causal relationship (Goodall, 2000). Birds have been observed to toss small pieces of food that float on the surface of a pond to lure fish to the surface where they can be captured more easily. Human understanding of complex causes led to the control of fire, the invention of cooking and agriculture, the construction of shelters and tools, and other innovations making it possible for our species to successfully occupy every region of the planet.

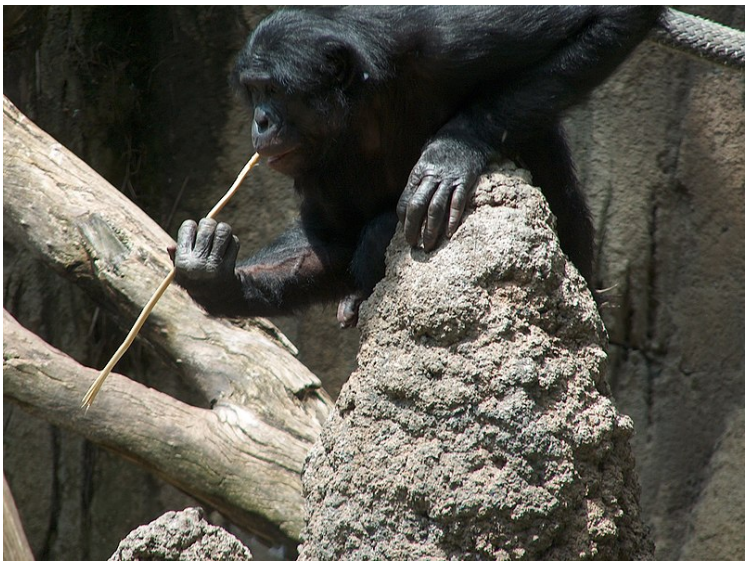


Figure **14.2.7**: Bonobo chimpanzee termite fishing. (Image from Wikimedia Commons; https://commons.wikimedia.org/wiki/F...r_termites.jpg; by Mike Richey; licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license. Caption by Kenneth A. Koenigshofer, PhD.).

If humans and at least some animals are born with an innate disposition to understand cause and effect, we should expect to find brain structures that seem to be especially involved in perception of and reasoning about cause-effect. Brain imaging studies show that "specific brain networks are involved in the extraction of causal structure from the world" (Fugelsang, et al., 2005, p.45). Cues humans use to determine if two events are causally related include covariation (things occur together), temporal order, contiguity (closeness) in time and space, information about possible causal mechanism in the specific instance of causality being examined, and similarity between cause and effect. Analysis of immediate *visual perception* of causality in humans using fMRI implicates several brain structures including **right middle frontal gyrus**, **right inferior parietal lobule**, **right prefrontal, right parietal, and right temporal lobes** (Fugelsang, et al., 2005). Fugelsang, et al. (2005) also reported that **right parietal cortex** was involved in *spatial cues of causality* (spatial contiguity) and **right temporal cortex** was involved in processing *temporal cues* (temporal contiguity) *of causation*. **Right prefrontal cortex** showed increased activation for both or either type of cue for causation. **Bilateral prefrontal** cortical activation was seen with tasks requiring **logical inference** to make a judgement of causation. These results add to understanding of how the visual system extracts causal information from spatial and temporal cues and further suggests that *causality arising from inference and real world knowledge*, rather than from direct perceptual experience, may correspond to **bilateral activation of** the **prefrontal cortex**.

Fugelsang et al. (2005) state that their findings confirm theories that *extraction of causality is inherent in the visual system*, akin to the innate Gestalt principles of completion of partial contours and perceptual grouping based on similarity or spatial proximity. In addition, "perception of causality appears very early in human life and is culturally invariant" supporting the view that the "visual system may be specially tuned to recover causal structure from the environment" (Fugelsang, et al., 2005, p. 41). Several other studies also support the hypothesis that **extraction of causal structure** is an **innate** property of the brain structures of the visual system (Blakemore, et al., 2001; Fonlupt, 2003; Roser et al., 2005). Perception of causality is immediate like perception of motion or perception of faces suggesting innate mechanisms for perception of causation. Some authors have suggested that the detection of causality may even be served by a specialized brain module for recognizing and understanding causality (Leslie & Keeble, 1987; Scholl & Nakayama, 2002). Furthermore, infants are sensitive to physical causality at only 6 months of age (Leslie & Keeble, 1987; Oakes, 1994).

The exact brain areas activated during detection of causation depends on a number of factors. Judgments of causality that require integration of a working hypothesis with relevant data activate neural tissue in **prefrontal and occipital cortices**. In addition, evaluation of data that is consistent with a *plausible* causal theory, was reported to recruit neural tissue in the **parahippocampal gyrus.** By contrast, **e**valuation of data inconsistent with a plausible causal theory recruited neural tissue in **anterior cingulate, left dorsolateral prefrontal cortex, and precuneus,** suggesting a neural mechanism by which working hypotheses and evidence are

integrated in the brain (Fugelsang & Dunbar, 2005). These empirical studies support the theory that, *over the course of evolution, information about cause-effect relations has been genetically internalized into the brain.* Innate understanding of causality is *a highly adaptive component of intelligence* in humans and animals. For example, without an understanding of causality, the planning and execution of goal-directed action on the environment would not be possible.

14.2.1.7.2 Similarity

It is a fundamental property of the world that things show similarity to other things in various features and at varying levels of abstraction (i.e. "There is nothing new under the sun."). An apple is similar to other apples and apples are similar to many other fruits. A wolf and a panther are both dangerous predators. All birds have beaks and wings. Prey animals frequent water holes. Both an obsidian knife and an obsidian arrowhead can penetrate a prey animal's flesh. Fast flowing rivers are dangerous to cross. Similarities among things in the world hold important adaptive information. Similarity allows categorization, prediction, and inferences based on generalization. If in the past I have witnessed fast flowing rivers wash people away, then when I encounter another fast flowing river, I can use the similarity to make the prediction that crossing this river could be dangerous, so I adjust my behavior accordingly and perhaps save my life.

This seems so obvious that we can't imagine not behaving so. It would be very stupid, not to be able to take advantage of similarities between past experiences and a current situation, especially when it might mean life or death. And that is exactly the point. It would be stupid, but we and other animals are not stupid, but the reverse, we are intelligent. But where did this ability for generalization come from? It is in our brain so firmly that we readily make such judgments of similarity, and generalize from them to make predictions, and we do this with such ease that we don't even think about it. But in neuroscience, it is those things that we do with greatest ease that most require explanation.

Extraction and exploitation of information such as this, contained in similarities, is an extremely powerful adaptive opportunity which natural selection could not have missed during the course of brain evolution. The fact that generalization and even categorization are found in a wide range of animal species (Soto & Wasserman, 2015) shows that these species are capable of using similarities in the environment to improve adaptation and biological fitness. This suggests that the adaptive information present in similarities drove the evolution of brain mechanisms, from early on, that could efficiently extract this information and put it to adaptive use, at least in birds and mammals, and perhaps in other animal groups as well (e.g. generalization has been found even in some insect species; see below).

Thus, “evolutionary/genetic internalization” (Shepard, 1994, p. 26) by natural selection of similarity, as an abstract, ubiquitous property of the world, provides another innate principle of how intelligence and cognition are organized. Just like we and many other animals have a "cognitive instinct" to understand and to experience the world in terms of cause-effect, we have a cognitive instinct to detect and readily utilize similarities to organize adaptive behavior and improve biological fitness.

Both causality and similarity are properties of how the physical world is put together and both have been incorporated into evolved mechanisms of intelligence in the brain. This makes sense. The brain must organize movement, behavior, in the physical world on our planet and therefore brain mechanisms must be effective in guiding behavior in that physical world. To do so, intelligence must have evolved to reflect and be organized around the general principles of how this physical world works--it operates by these universal principles, cause-effect and the fact that things in the world are naturally similar to other things in ways that provide organisms with information that is important for adaptation.

Genetic internalization of similarity as a general feature of the world created a disposition in human and animals brains to “expect” and to find similarities in the world, to group things by similarity into categories, to readily match new instances to the appropriate category based on similarity, and to infer properties of new instances of a category based on knowledge of properties of the category as a whole. The cognitive processes of generalization, concept formation, categorization, inductive reasoning, categorically based inference, and categorical logic, all emerged as a family of related functional properties of the brain as natural selection fine-tuned the **genetic/evolutionary internalization of similarity** as a fundamental relational regularity in the world (on this basis, it is expected that these cognitive abilities should be correlated). These observations are reflected in “what William James (1890/1950) called ‘the very keel and backbone of our thinking’: sameness. The ability to evaluate . . . similarity . . . is clearly the sine qua non of biological cognition, subserving nearly every cognitive process from stimulus generalization and Pavlovian conditioning to object recognition, categorization, and inductive reasoning” (Penn et al., 2008, p. 111). Note that many

of these psychological capabilities are the same ones identified earlier in the definition of general intelligence from Gottfredson (1997).

As noted above, **similarities hold important adaptive information**. Similarities bring order out of chaos. They allow inference and prediction. If you know about one leopard as a member of a category of dangerous things, then you know about many leopards, perhaps all leopards, even those you have yet to encounter. If and when you encounter a leopard again, you already know much about it that can be put to use to organize your response, your behavior, and that knowledge that you gain from an inference about this new leopard, based on the category to which it belongs, may save your life. Evolution has "tuned" our brains to find similarities among things in the world--along with genetic internalization of causation, another exceedingly powerful property of general intelligence as described earlier. As noted above, the inborn disposition of the brain to find and record similarities in the environment guides us (our brains) to form categories and concepts of varying degrees of abstraction which embed adaptive information within them (e.g. all saber-toothed cats are dangerous; pieces of flint are sharp making them a good source of cutting and scraping tools; all Datura flowers are poisonous; things with feathers that fly are all birds; situations where someone is treated unfairly are all examples of injustice; all atomic particles with a negative charge are electrons, and so on). Categories and concepts are one way in which the brain captures the information contained in similarities in the world and forms knowledge structures that reflect and exploit one form of order in the environment, that order that arises from similarity, thereby improving the guidance of movement, increasing successful adaptation by intelligent action.

Recall that Shepard argued that “the evolutionary internalization of universal regularities” (1994, p. 26) included “three-dimensional, locally Euclidian space,” one-dimensional time with a "thermodynamically conferred unique forward direction," and cycles of light and dark. The genetic internalization of these regular features of the world underlies important components of our ability to understand the world. However, in addition, Shepard has also argued that there has been "evolutionary internalization" of the recurrent fact of the world that "objects having an important consequence are of a particular natural kind . . . however much those objects may vary in their sensible properties . . . " (Shepard, 1992, p. 500). In other words, things can be categorized on the basis of their consequences, regardless of their superficial physical properties. This more abstract form of similarity leads to grouping things which otherwise differ in surface ways but which share the same category of effects, for example, all large predators including bears, wolves, sharks, etc. are dangerous. Other things with very different superficial features may all be good to eat, or poisonous, or have other important adaptive consequences by which they can be grouped. This kind of more abstract categorization, based on similarities in effects, allows a higher level of behavioral guidance by the brain. Some of these categories and the behavioral responses to them may even be innate. For example, feces or spider-like things may innately cause feelings of disgust or fear and lead to withdrawal and avoidance behavior, thereby protecting the organism from potential harm from pathogens or a poisonous bite. The claim that "objects having an important consequence are of a particular natural kind" is basis for Shepard's “universal law of generalization” (Shepard, 1987b), a law which has been verified in diverse species including insects (e.g., see Cheng, 2000).

A quick review is important here. The psychological process of generalization exemplifies a basic property of human and animal minds--that they tend to form groupings, categories, based on similar instances (Broschard, et al., 2021), and then, exploiting these groupings, the mind/brain uses generalization to make inferences or projections about new instances of a category. For example, if you have encountered a number of cats in your past, you will form a grouping in your mind, a category, "cat," and when you see another creature in the future similar to previous cats you have encountered, you can already make a number of inferences about this new instance of the category--e.g. it has claws, it purrs, it eats mice, etc. You can infer these properties even though you have just encountered this particular cat only a few moments before. Because of categorization and generalization, you can make inferences and predictions about this new creature which you have recognized as a member of the category, "cat," based on similarities among all cats. This is an exceedingly important component of our thinking processes and our intelligence. We use this type of process much of the time when we are engaged in "thinking" and "reasoning."

Recognizing similarities among things, and using those similarities to make useful inferences and predictions has exceedingly beneficial effects on survival and sometimes directly upon reproduction. For example, chimpanzee males of lower status have been observed to steal a mating opportunity usually not available to them, when a more dominant male is sufficiently distracted. When a lower status male observes more dominant males fight these males are usually distracted from other matters. If in most situations where dominant males are fighting they are distracted from their watch over their favorite females in the group, and if a lower status male chimp recognizes a similarity between a current fight between dominant males and past such battles, and if, based on

the similarity, he infers that during a current fight between dominant males, he likely has a mating opportunity (especially if he is quick), he may successfully mate with a desirable female that he would not otherwise have access to. Males who can make inferences like these based on similarity have a higher reproductive rate than males who cannot make such inferences.

Ancestral humans who recognized similarities between a new animal, never encountered before, and predators encountered in the past, can put this new animal in the category "dangerous predator" and then infer that this new animal is probably also dangerous and should be avoided. The adaptive value of this inference based on similarity is immediate. Imagine an ancestral human who had brain organization incapable of recognizing similarity at all and consequently could not form categories and concepts, and therefore could not make any inferences based on similarity. In this hypothetical case, where similarity to previous predators holds important survival information, clearly survival is at stake, and if one had a brain incapable of exploiting the survival information in similarity then one's chances of seeing another day are vastly reduced and the genes responsible will probably soon be eliminated (an instance of natural selection at work).

Research suggests that category formation depends on the **prefrontal cortex** in humans, and on the pre-limbic cortex in rats (Broschard, et al., 2021). Studies of brain damaged humans also implicate a subcortical involvement of the **striatum**. Specifically, Lagarde, et al. (2015) report involvement of *"a prefrontal-striatal loop in abstract categorization*, and more particularly, the involvement of the head of the **caudate** nuclei and **left frontal lobe** in access to abstract representations in verbal concept formation." To emphasize the point once more, the ability to form groupings, or categories, based on similarity, and to then make inferences about newly encountered instances of the category, is a powerfully adaptive property of the brain's operations. Categories and concepts based on similarity permit our brains to make inferences about newly encountered instances of the category or concept, like those in the examples above. Categorization, formation of concepts, and inferences based on similarity are key features of general intelligence. All of these cognitive abilities can be understood as arising from the genetic internalization of similarity relations into brain organization by natural selection (Koenigshofer, 2017).

14.2.1.7.3 Predictive Relations (Predictive Co-occurrence of Events)

In addition to causality and similarity, it is also an enduring property of the world that some things regularly predict the occurrence of other things (a growling bear predicts an impending attack, sudden heavy rainfall predicts possible flash flooding, an approaching range fire predicts potential danger, a water hole on the plains of Africa predicts that prey animals will be near, and so on). Clearly, **predictive relations** are present everywhere in the environment. It is also clear that predictive relations among things in the world hold exceedingly important adaptive information. This source of adaptive information is far too important for natural selection to have missed it. Thus, we should expect that natural selection must have organized brain systems that capture, not only causality and similarity, but also the *predictive relations* between events in the world. Using Shepard's (1992) terminology, we should expect that over the course of brain evolution, there occurred the "genetic internalization" (Shepard, 1992) of mechanisms that capture predictive relations between events. These mechanisms underlie another central component of general intelligence in humans and many animals.

The brain evolved for prediction (Clark, 2013, 2016; Nave, et al., 2020). The brain can be characterized as "a probabilistic prediction engine" (Nave, et al., 2020, p. e1542). It is not only evolved for prediction, but to implicitly make probability assessments about how likely its predictions about future are to be true. Natural selection for brain organization that captures the adaptive information in predictive relations has predisposed us (i.e. tuned our brains, and the brains of other animals) to find and to exploit the adaptive information held in predictive relations, an exceedingly adaptive property of general intelligence. Predictive relations exist whenever specific events, objects, or situations consistently co-vary or correlate--whenever one thing consistently follows another; for example, think of the CS and US in classical conditioning. After conditioning has occurred, the CS predicts the US. This implicit prediction then controls behavior--the dog salivates immediately after the CS is presented in anticipation of coming food, a response that prepares the dog for the expected arrival of the US (e.g. the dog drools, and this saliva prepares the dog to swallow and digest efficiently if the expected food follows). As discussed in the chapter on learning and memory *classical conditioning* specifically involves the **cerebellum**, whereas associations, in general, between co-occurring events appears to depend on **long-term potentiation (LTP)**, and corresponding changes at synapses in **hippocampus** (Oishi, et al., 2019), **cerebral cortex** (Daw, et al., 2009; De Pasquale, et al., 2014) and other brain regions.

The adaptive significance of this property of general intelligence is hard to overestimate. An innate predisposition built into the brain to search for predictive relations in the environment and to exploit the adaptive information in those relations is an

exceedingly powerful cognitive tool for understanding the world. Using this information to make biologically important predictions (a low growl in the nearby brush may be soon followed by an attack from a large predator), to organize behavior serves biological fitness (survival and reproduction of one's genes). To understand the powerful effect on adaptation of an innate disposition to recognize and exploit predictive relations between events in the world, just imagine the crippling effects on survival if this ability were absent in animals and humans. If you had no knowledge of *what follows what* in the environment, the world would seem a chaotic jumble of unrelated events. In such a world, it would be very difficult if not impossible to organize adaptive action to survive the demands of a complex environment. One might speculate, and this is only speculation, that something like this might be how some persons with autism experience the world; perhaps the inability to understand cause-effect or to recognize similarities in order to form categories and make sensible inferences, or inability to recognize predictive relations in the world, or perhaps loss of all three of these abilities might be responsible for some of the symptomology of people who are very severely autistic. The world would seem to lack any order at all without these cognitive abilities. Ordering movement into adaptive patterns would be exceedingly difficult if not impossible.



Figure **14.2.8**: Contrasting behaviors to satisfy water needs in two species. The woman uses invention and technology, two products of human intelligence, to provide a more consistent and higher quality water source. (Images from Wikimedia Commons; Elephants drinking; https://commons.wikimedia.org/wiki/F...k,_Kenya_4.jpg ; by CT Cooper; licensed under the Creative Commons Attribution 3.0 Unported license. Woman pumping well; https://commons.wikimedia.org/wiki/F...an_Working.jpg; by Osaba Gerald; licensed under the Creative Commons Attribution-Share Alike 4.0 International license; caption by Kenneth A. Koenigshofer, PhD).

14.2.1.8 Intelligence: Cognitive Instincts and "the Infinite Use of Finite Means"

Recall this quote from above: "Evolutionary psychologists sometimes argue that a class of situations must be relatively narrow to exert consistent selection pressure, but this claim is insufficiently justified. Any regularity in the environment can exert selection pressure if it poses a challenge or an opportunity to the organism . . . (Kaufman, et al., 2011, p. 213)

In the case of the abstract relational regularities which drove the evolution of general intelligence, the adaptive opportunity is difficult to overstate. The genetic internalization by natural selection of fundamental information about these abstract relational regularities in the environment has equipped the mind with an exceedingly powerful set of cognitive structures--a set of inborn, general, rule-like principles of how the world works. Deployment of this inborn, implicit “understanding” about causality, similarity, and predictive relations gives general intelligence (g) its adaptive punch because it permits innovative and improvisational solutions to an enormous variety of adaptive problems. This is an example of what Steven Pinker (2007) calls the “infinite use of finite means,” applying a finite set of rules to understand and to solve an enormous variety of adaptive problems and to exploit the diverse opportunities in the environment. We see the power of human general intelligence in the stunning progress in human cultural and technological development from prehistoric times to the present. The existence of **general intelligence** (sometimes called "**fluid intelligence**") contradicts the traditional view sometimes ascribed to evolutionary psychologists that there are specialized modules in the brain for every problem type that ancestral humans had to solve (this is the theoretical assumption of the "massive modularity" of the human mind which leaves no room for more general cognitive processes including general intelligence--see Lloyd & Feldman, 2002; but for an alternative view held by most evolutionary psychologists,

"soft modularity," which does allow for some general cognitive processes, see Ellis & Ketelaar, 2002). However, arguably there is no real contradiction if one understands that general intelligence is actually composed of several adaptive specializations that have genetically internalized the abstract properties of causality, similarity, and predictive relations. It is the internalization of these relations, because they are universally present in the world, that gives the mistaken impression of "general process" mechanisms which in fact are actually evolved adaptive specializations to these three specific abstract relations found in the structure of the world (Koenigshofer, 2017).

Genetically internalized information about causality, similarity, and predictive relations as general features of the structure of the world creates a scaffolding of **instinctual knowledge** which **guides** the **learning** of specific details of these relations in **one's own particular environment**, an environment full of specific, perhaps even unique, instances of these pervasive relational regularities of the world (Koenigshofer, 2017). For example, understanding causality as a general principle of the world allowed our ancient ancestors to learn that fire could be used to cook meat making it easier to digest; that setting well-placed fires on grasslands could be used to heard prey animals toward an ambush to facilitate the hunt; that some stones could be shaped into spear heads and tips of arrows; and that skins of animals could be taken and used for clothing. This same instinctual knowledge about causality permitted modern humans to learn how to make and use concrete to lay building foundations; to learn that fertilizers increase crop yields; that rising interest rates cause stock prices to fall; and that sunlight could be converted to useable electricity.

Innate knowledge about cause-effect as a general property of the world allows humans to analyze complex causal relations to solve a wide range of adaptive problems with creativity and innovation unprecedented in the animal kingdom. For example, while most animals seek shelter in vegetation, or in caves, or other features of the natural habitat, humans build castles, houses and skyscrapers. While other animals walk and run or fly to move from one place to another, humans invented the wheel, engines, and steel and now ride in cars, trains, planes, or even a Blue Origin space capsule. While most animals hunt for water and food, humans have invented plumbing, water companies, agriculture and supermarkets. While animals fight with teeth and claws, humans have invented and used spears, guns, and nuclear weapons. These intellectual achievements involve a general understanding of causality, similarity, and predictive relations as well sophisticated understanding of three dimensional spatial relationships, ability for visualization, and culturally transmitted knowledge accumulated and transmitted over generations, primarily and most efficiently by language.

This view that we are born with genes which give our brains innate knowledge about many things in the world is a radical departure from the view that the mind is a **blank slate** at birth. This conception of the mind was promoted by a group of influential 18th century philosophers known as the British Empiricists (such as John Locke). This view of the human mind has been carried forward all the way into 21st century social science and philosophy. Evolutionary psychologists, Cosmides and Tooby (2007), call this old empiricist view of the mind the **Standard Social Science Model (SSSM)**. We now know that this blank slate view of the mind/brain (SSSM) is wrong. For example, as mentioned earlier, Alison Gopnik and her colleagues at Berkeley (Gopnik and Sobel, 2000; Gopnik, et al., 2001; Gopnik, et al., 2004) have found that children, at a very early age, readily discover cause-effect and predictive relations between events in the laboratory, supporting the view that the human brain is genetically predisposed to understand abstract relations such as causality and predictive relations among things in the world (Koenigshofer, 2017). Evolutionary psychologists, Cosmides and Tooby, call the blank slate view of the human mind (the SSSM) "**biologically naive**," "radically defective," and an outdated, theoretical leftover from the days of radical behaviorism.

14.2.2 Cognitive Instincts and Intelligence

As was described in the theoretical account above, the brain has *genetically internalized by natural selection* crucial information about regularities in the world of varying degrees of concreteness and abstraction. This equips us and other animals with inborn, instinctual knowledge about many enduring facts about the environment. Snakes, spiders, and heights have been consistently dangerous to humans generation after generation so natural selection has built this across-generation fact into our brain circuitry in the form of fear circuits readily activated by these stimuli. Certain facial and body forms in the opposite sex are consistently associated over generations with health and high reproductive potential so we have brain circuitry that innately attracts us to these body and facial forms. Three-dimensional space, the forward progression of time, cycles of light and dark, and Shepard's universal law of generalization have been "genetically internalized" into the brain over eons of evolution. And, as we have just discussed, natural selection has "genetically internalized" information about some quite abstract and universal regularities of the world: causality, similarity, and predictive relations. As a consequence, we are innately predisposed to understand cause-effect, the three-dimensionality of space, the progression of time, similarity relations used to form categories and concepts, to fear snakes and

spiders, to seek particular traits in sexual partners, to look for and find predictive relations among events in the world, to understand the minds of others, to protect our young, to form and maintain friendships, and so on. This means that we are born with many genetically encoded "**cognitive instincts**" which make up much of our genetically **inherited intelligence** as a species, the consequence of eons of natural selection.

Details of how natural selection could bring about the genetic internalization of abstract relational regularities of the world (causality, similarity, and predictive relations) are described elsewhere (see Koenigshofer, 2017). Suffice it to say, that adaptive problems and opportunities, though variable or even “novel” in details, are nevertheless invariant in common abstract, relational structure. That recurrent structure provides stable selection criteria over generations for genetic internalization. Since natural selection can only operate over many generations, the specific details of individual cases of each of the abstract relations (cause-effect, similarity, predictive relations) drop out over generations. This leaves only enduring, across-generation, "distilled" abstract relational regularities (independent of specific concrete contents) to act as selection criteria for the genetic internalization of these relational regularities of the world into brain operations. This process has apparently taken place not only in humans, but also in some number of non-human animals as well, since **general intelligence** has also been identified in a variety of non-human **animal species** (Bird & Emery, 2009a, 2009b; Reader, et al., 2011), as well as in humans.

14.2.3 What makes humans so smart?

According to the theoretical views explained above, genetic internalization of these enduring, across-generation, abstract, relational regularities of how the world works drove the evolution of flexible general intelligence in humans and many other animal species. But what makes humans so smart compared to other species on the planet? One key difference between human intelligence and intelligence in non-human animal species is the much higher degrees of **abstraction** that humans are capable of representing, compared to other animals. The ability of the human brain for high levels of abstraction, setting it apart from the brains of other species, is likely due to the greater complexity of circuits in human cerebral cortex compared to other mammals (recall that only mammals have six layered cerebral cortex and approximately 150,000 cortical columns--discussed above--while non-human mammals have significantly fewer). Humans easily detect highly abstract similarities in the environment compared to non-human animals and are capable of forming highly abstract categories and concepts based on these abstract similarities in properties or functions (Koenigshofer, 2017; Penn et al., 2008). For example, Penn et al. (2008) found that the use of causal principles and similarity assessments of high degrees of abstraction is a distinguishing feature of human cognition and problem solving. As they state: “Even preschool-age children understand that the relation between a bird and its nest is similar to the relation between a dog and its doghouse despite the fact that there is little “surface” or “object” similarity between the relations’ constituents” (Penn et al., 2008, p. 111).

Research in molecular genetics suggests one possible explanation for how human ability for high levels of abstraction may have come about. Pollard (2009), comparing human and chimpanzee genomes, found “massive mutations” in humans in the “DNA switches” controlling size and complexity of cerebral cortex, extending the period of prenatal cell division in human cerebral cortex by several days compared to our closest primate relatives. Research using **artificial neural networks** suggests that increasing cortical complexity leads to sudden leaps in ability for abstraction and rule-like understanding of general principles (Clark, 1993), lending further support to the hypothesis that **superior ability for abstraction due to cortical complexity** may be the key component explaining differences in general intelligence between humans and nonhuman animals (see module on artificial neural networks in this chapter).

The frontal cortex is involved in concrete rule learning. **More anterior regions** along the rostro-caudal axis **of frontal cortex support rule learning at higher levels of abstraction.** These abilities for high levels of abstraction may involve **anterior dorsolateral prefrontal cortex** in humans (Kroger et al., 2002; Reber, Stark, and Squire, 1998). Along with human language, the exceptional capacity of the human brain for abstraction may explain the unusual achievements of the human species ranging from agriculture, technology and science, to the invention of complex economies and governments (Koenigshofer, 2016, 2017).

Furthermore, **cultural transmission** of learned information in humans, ranging from books and educational institutions to film and the internet, allows each generation to profit from the accumulated knowledge of prior generations (it might even be said that cultural transmission is a key specialty of the human species). Cultural transmission, along with cooperative problem solving and development of experts in different fields, creates a kind of **group intelligence** not possible in other species. Of course, human language, spoken and written, plays an enormous role in these shared cognitive processes among humans, perhaps explaining at

least in part why language evolved. These considerations bring to the forefront the role of culture in human genetic evolution. **Gene-culture coevolution theory** proposes a multidirectional coevolution between genes and culture in which culture and human created artifacts such as tools, weapons, clothing, and pottery have driven the last 100,000 years of human evolution (Lloyd & Feldman, 2002; Durham, 1991). This theory also assumes an important role for **language in cultural transmission of learned knowledge** and behavior. It also postulates existence of hominin biological changes that facilitated development of sophisticated language in humans (Lloyd & Feldman, 2002).

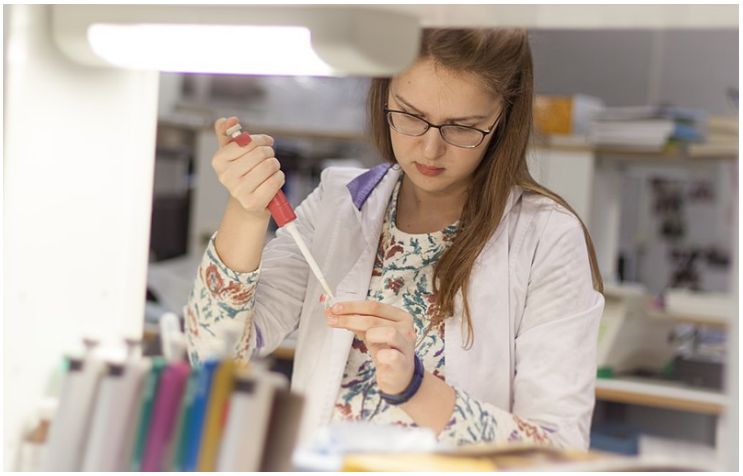


Figure **14.2.9**: Scientists seek cause-effect, similarity, and predictive relations between variables. Scientific methods are an example of human general intelligence at its best. Alison Gopnik (2010, 2012) at UC Berkeley has found that young children engage in logical analysis of causation, teasing out causes from non-causal factors, in much the same way that scientists do. (Image from Wikipedia, https://commons.wikimedia.org/wiki/F..._reaction).jpg; by Alenakopytova; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

Another factor that may contribute to the superior intellectual capacities of humans is the sophisticated control systems in the human prefrontal cortex. Miller and Cohen (2001) proposed that “cognitive control stems from the active maintenance of patterns of activity in the **prefrontal cortex** that represents goals and means to achieve them. They provide bias signals to other brain structures whose net effect is to guide the flow of activity along neural pathways that establish the proper mappings between inputs, internal states, and outputs needed to perform a given task.” On their view, the prefrontal cortex (PFC) guides the flow of neural activity in relevant brain areas, which allows for cognitive control of behavior. As they state: "depending on their target of influence, representations in the PFC can function variously as attentional templates, rules, or goals by providing top-down bias signals to other parts of the brain that guide the flow of activity along the pathways needed to perform a task" (Miller & Cohen, 2001).

Goldman-Rakic (1996) proposed that the **prefrontal cortex** represents information not currently in the environment, creating a "**mental sketch pad**," to hold visual images in working memory, to represent plans, and to focus attention to intelligently guide thought, action, and emotion. This includes the inhibition of distracting thoughts, actions, and feelings. Although such control does exist to some degree in some non-human animals it is less developed and less influential over behavior than in our own species. The prefrontal cortex is highly interconnected with much of the brain, including extensive connections with other cortical, subcortical and brain stem sites. The **dorsal prefrontal cortex** is especially interconnected with brain regions involved with attention, cognition and action, while the **ventral prefrontal cortex** interconnects with brain regions involved with emotion. According to Striedter (2005), the **prefrontal cortex** (PFC) of humans consists of two functionally, morphologically, and evolutionarily different regions: the **ventromedial PFC** (vmPFC) composed of the ***ventral prefrontal cortex*** and the ***medial prefrontal cortex*** present **in all mammals**, and the **lateral prefrontal cortex (LPFC)**, composed of the ***dorsolateral prefrontal cortex*** and the ***ventrolateral prefrontal cortex***, present **only in primates**.

14.2.4 Mental Models of the World: Capturing Environmental Regularities for Prediction and Planning for Future

Think for a moment about the complexity of the environment that animals like us must adapt to. We live in a world that consists of inanimate and animate objects, including our own body, that move and change in space and time. Forces that operate on objects can be physical like gravity and the laws of motion, while for people and animals, additional forces acting on them can be social, emotional, or generated by physiological or other needs and motives. The relations among things can be spatial (e.g. the lion is approaching me from my right), temporal (e.g. right after the bell rang a pellet of food appeared), causal/logical (germs cause disease), predictive (one event signals that another event is about to follow), social (interactions among conspecifics), and can reflect similarity (a bear and a wolf are both large predators) and differences (wolves hunt in packs, bears don't), as well as other entities of even greater levels of abstraction.

To successfully navigate this incredibly complex world and its causal and other contingencies, the brain must have accurate neural representations or mental models, inclusive of all of these objects, forces, and relations, in order to guide adaptive behavior. Consequently, evolution has equipped us with a brain that generates **neural representations (mental models or "cognitive maps")** of things (including ourselves, other people, animals, processes, etc.), the relations among them, and the forces that operate on them in three-dimensional space and time (see Behrens, et al., 2018). Many psychologists and neuroscientists have recognized the importance of mental models in the intelligent organization of behavior. "Mental models are psychological representations that have the same relational structure as what they represent. They have been invoked to explain many important aspects of human reasoning, including deduction, induction, problem solving, [and] language understanding, . . . [and they are] also valuable for providing new understanding of how minds perform abduction, a kind of inference that generates and/or evaluates explanatory hypotheses" which focus on explaining causal mechanisms (Thagard, 2010, p. 447). "[T]hese models underlie deductive, inductive, and abductive reasoning yielding explanations" that permit us to understand the world (Khemlani, et al., 2014) and then to use this understanding to predict, to plan, and to manipulate objects and events in the environment to human advantage. In an important sense, the use of mental models to generate adaptive behavior encompasses much of what we mean by intelligence.

14.2.4.1 Neural Coding and Mental Models

Mental models are thought to be **generated by the activities of large populations of neurons** in the brain. One theory of how this is done proposes that complex ideas, concepts, perceptions, memories and other mental structures are represented by specific patterns of neural firing, over different regions of brain, in enormous numbers of neurons, each "tuned" to fire to a specific feature or to a specific value of a feature or property of the world (recall Hebb's idea of "cell assemblies" discussed in the chapter on learning and memory). Neuroscientists know that this sort of coding of properties exists in many parts of the brain. For example, recall from the chapters on vision and other sensory systems that individual neurons in the visual cortex are "tuned" (i.e. have receptive fields) to specific visual properties such as the orientation of linear borders and their direction of movement, while some neurons in somatosensory cortex, for example, respond to specific angles of specific joints of the skeleton to code limb positions, while other respond to tactile stimulation of specific areas (dermatomes) of skin, while still other neurons in auditory cortex fire to a specific narrow range of frequencies. Paul Churchland, at the University of California, San Diego, postulates that complex stimuli such as human faces or even highly abstract concepts and ideas such as scientific theories may be neurally represented in the brain by complex patterns of firing in enormous populations of neurons, each tuned to fire neuron potentials to specific features and magnitudes of features found in the world or generated by creative recombination of features in imagination.

This kind of coding can be conceptualized as long strings of numbers (each number is referred to as a "scalar," and the entire string of scalars is called a "coding vector") in which each number (scalar) in the string (the vector) corresponds to the firing rate of a specific neuron in a large population of neurons involved in the coding of a perception, a concept, an idea, or a complex set of ideas such as a scientific theory. This means that **coding vectors** for a single idea might be exceedingly long, containing thousands or millions of scalars in a single coding vector, each such scalar representing the firing rate of one of thousands or even millions of neurons involved in the coding of a single concept, idea, or complex of interrelated ideas (see Churchland, 1985, 2013).

Consistent with this view, the coding capacity of a single human brain with its 100 to 200 billion neurons is staggering, well within the parameters suggested by Churchland. As Thagard (2010, p. 450) explains: "A population of neurons represents something by its pattern of firing. The brain is capable of a vast number of patterns: assuming that each neuron can fire 100 times per second, then the **number of firing patterns** of that duration is $(2^{(100)})^{100000000000}$, a number **far larger than** the **number of elementary**

particles in the universe, which is only about 10^{80}. A pattern of activation in the brain constitutes **a representation of something** when there is a stable causal correlation between the firing of neurons in a population and the thing that is represented, such as an object or group of objects in the world."

According to mental model theory, the enormous coding capacity of the human brain permits the human mind to represent concepts, ideas, theories, and other components of human knowledge of great subtlety and complexity, including theories in science about hidden causes of things not directly observable (Churchland, 1985, 2013). "[M]ental models operating at various degrees of abstraction are invaluable for high-level reasoning. . . you can build simplified but immensely useful models of past and future events, as well as of events that your senses do not enable you to observe. Hence science uses abductive inference and conceptual combination to generate representations of theoretical (i.e. non-observable) entities such as electrons, viruses, genes, and mental representations" (Thagard, 2010, p. 457).

Using Churchland's concept of **vector coding**, each neural representation of a perception, a concept, an abstract relation, a thought, or other mental activity could be conceptualized as a specific, very long, coding vector, perhaps consisting of thousands or millions of scalars corresponding to the firing rates in an equal number of neurons which are involved in the coding of that particular mental activity. By this reasoning, mental models of the world would involve patterns of firing over space (neural space; i.e. areas of brain) and time (temporal patterning) in enormous numbers of neurons, corresponding to coding vectors of vast length changing their scalar values and composition moment to moment. This is hard to imagine, but perhaps this is the scale on which neural coding of mental events, such as concepts, relations, ideas, and so on takes place--millions upon millions of neurons firing in ever changing spatial-temporal patterns of neural activity distributed across many regions of brain at any one time.

14.2.4.2 Brain Mechanisms, Mental Representations, Relational Reasoning and Intelligence

Research suggests that reasoning requires coding involving neurons in **prefrontal cortex**. "Reasoning depends on the ability to form and manipulate mental representations of relations between objects and events. . . [and] the integration of multiple relations between mental representations . . .[However,] patients with **prefrontal damage** exhibited a selective and catastrophic deficit in the integration of relations. . . . The integration of relations may be the fundamental common factor linking the diverse abilities that depend on prefrontal function, such as planning, problem solving, and fluid intelligence" (Waltz et al., 1999, p. 119; **fluid intelligence** is another term for **general intelligence**, and is separate from one's store of learned knowledge). These results suggest that although mental models may involve many different brain areas acting together, the **prefrontal cortex** appears to play a special role in high level abstract reasoning requiring the integration of multiple relations. This may explain why **damage in prefrontal cortex** "leads to selective decrements in performance on tasks involving hypothesis testing, categorization, planning, and problem solving, all of which involve relational reasoning" (Waltz et al., 1999, p. 119).

In addition, because the ability to anticipate and prepare for future events improves chances of survival, evolution has favored **brains that** can **predict** future events (Clark, 2013; Nave, et al., 2020; Koenigshofer, 2017) including the future consequences of our own actions. Thus, according to mental model theory, intelligence involves formation by the brain of **mental/neural models or mappings** of the **physical and social worlds** and how they work. This idea is similar to **Tolman's concept of "cognitive maps,"** in which he envisioned mental mappings of the environment even in rats learning a maze (Tolman, 1948). This concept has even been extended to understand the coding of social relationships (Schafer & Schiller, 2020). On this view, **intelligence** involves enormously complex neural representations, models, or mappings of the environment and how it operates, including its causal relations, from which it is possible to make causal inferences and to generate causal explanations. There is evidence that mental models involved in causal inference may depend upon localized regions of cortex. According to Khemlani, et al. (2014) "**mental models for causal inference** are implemented within **lateral prefrontal cortex.**" Furthermore, planning future actions has also been associated by researchers with **lateral prefrontal and parietal cortex**. "**Lateral prefrontal cortex**, the **anterior** extent of the **inferior parietal lobule**, **dorsal anterior cingulate**, and **anterior insula** comprise regions of an extended **frontoparietal** control system" (Spreng, et al., 2015).

Recall Shepard's idea that there has evolved "a mesh between the principles of the mind and the regularities of the world." As discussed above, a primary characteristic of intelligence is that it predisposes animals and humans to discover and exploit regularities in the environment of varying degrees of abstraction. Similarities between things in the world help organize knowledge by the formation of categories (see module that follows on categories, concepts, and schemas). Inference from categories to specific instances of a category is one way to predict future outcomes (if this happens, then that is likely to follow), including the

prediction of the effects of one's own future actions (if I do this, then that is likely to follow). Both of these kinds of predictions are examples of what we mean by "reasoning." The latter is critical for planning and the achievement of long-term goals.

Prediction allows us to plan ahead, to problem solve, and to achieve long-term goals, functions which depend upon **prefrontal cortex (PFC)**. For example, one function of the prefrontal cortex may be to gather facts, memories, and information from various brain regions and then coordinate and organize this divergent information into a workable whole in working memory in the PFC in order to solve a puzzle or find a solution to a problem (D'Esposito & Postle, 2015). However, an alternative theory is that the **PFC** is involved in **selective attention** to information which is localized in posterior brain regions such as **occipital and parietal cortex**. According to this view, the memories for relevant information are stored in the same regions activated during the initial sensory processing prior to memory formation (i.e. memories are stored in the sensory cortex). Information may flow in both directions from posterior regions of cortex to the PFC and from PFC to posterior regions of cortex during a visual cognitive task such as categorizing visual stimuli (Lara & Wallis, 2015). Some studies of sensorimotor processing have shown similar bidirectional interactions within the **frontoparietal network** (Siegal et al., 2015), a brain network which has been identified, using brain functional MRI (fMRI) and Positron Emission Tomography (PET), as a primary neural substrate relevant to **intelligence** (Colom et al., 2010; Hearne, et al., 2016; Jung & Haier, 2007).

The connection of *parietal and frontal cortex* to intelligence has been formalized in the **Parieto-Frontal Integration Theory (P-FIT) of intelligence** (Jung & Haier, 2007). According to Jung and Haier (2007), the P-FIT model of intelligence proposes that the **dorsolateral prefrontal cortex**, the **inferior and superior parietal lobule**, the **anterior cingulate**, and areas within the **temporal and occipital lobes** are crucially involved in intelligence. *White matter* regions (consisting of tracts of myelinated axons) such as the **arcuate fasciculus** are also involved (more details about this theory follow in a later module). This theory attempts to relate these brain areas and their functioning to individual differences among people in intelligence, as measured by standard intelligence tests (see module that follows on the measurement of human intelligence).

14.2.4.3 Representations of Space and Time as Components of Intelligence

As already discussed, similarity relations, predictive relations (event covariation), and cause-effect relations (Koenigshofer, 2017) are not the only environmental features which have been genetically internalized into brain operations by natural selection. In addition, as discussed above, there are fundamental properties of time and three-dimensional space (see Shepard, above) and a large number of other relatively concrete regularities of the world which have also been incorporated into the mental/neural models or cognitive maps of the environment created within our brains. Thus, we, like other animals, have specialized regions of the brain dedicated to time perception and others devoted to creation of three-dimensional representations of the space around us. Recall from above, the **"place cells"** discovered in the **hippocampus** of the rat. These are "tuned" to fire action potentials when the rat is in a particular place in its environment; recall that other cells in the **cerebellum, hippocampus, and entorhinal cortex** appear to be involved in representation of the passage of time (Irvy, et al., 2002; Schafer & Schiller, 2020). Natural selection favored brain circuitry capable of exploiting these universal regularities or invariants in the structure of the world, along with more specific regularities of the physical and social environments acquired through learning, allowing us to generate quite sophisticated mental models of how the world operates. These brain systems form neural groundwork for thinking and intelligence in humans and animals.

14.2.4.4 "From Reflex to Reflection" in the Control of Action

The better our mental/neurological models of the world are, the more accurately we can anticipate future events and thus prepare for them. The better we are able to prepare for future events, whether only moments or years ahead, the more successful, the more adaptive, our behaviors are likely to be. These mental models or cognitive maps are part of the sophisticated guidance systems that shape behavior into adaptive patterns. This allows humans and animals to meet the challenges, and to maximize the opportunities, presented by the environment. The result? Increased biological fitness measured by the successful reproduction of genes (Dawkins, 1976).

In fact, we can see an important evolutionary trend: the brain's increasing ability to anticipate future, beginning with reflexes (e.g. sexual reflexes anticipate coitus) and then classical conditioning, found even in simple animals, in which the organism learns to anticipate or expect the UCS (such as meat in Pavlov's experiments) a few seconds after the CS (the bell). Natural selection for prediction and the anticipation of future ultimately led to evolution of the ability for advanced mental planning and prediction in human thinking (Clark, 2013; Koenigshofer, 2017), sometimes involving anticipation of events decades or even centuries into the

future (e.g. you may anticipate a well paying, interesting career years from now after you have completed college, and that mental image keeps you focused day to day on your goal and the intermediate steps to achieve it; scientists predicted global warning decades ago; Leonardo da Vinci in the 1400's anticipated invention of the helicopter by hundreds of years; Jules Verne anticipated space travel long before the first human space flight; the founders of the U.S. Constitution formed principles of governance which anticipated political behavior of people hundreds of years in the future to this day). This general trend in evolution toward increasing capacity of brains to guide behavior by anticipation of the future can be summarized in the phrase "from reflex to reflection" (Koenigshofer, 2011, 2016).

Recent research has identified a number of brain structures involved in mental representations of future states of the world, specifically of one's own personal future, goals, and plans, essential for intelligent control of behavior. **Prospection** is the ability to mentally represent the future. To reach a future goal it is essential to make plans which organize a sequence of actions leading to achievement of the future goal. Behaviors controlled by mental anticipation of *imagined futures* can be called *prosponses* to distinguish them from *responses* to stimuli in the present or recent past (such as a CS in conditioning) (Koenigshofer, 2016). According to Spreng et al. (2015), *autobiographical planning* involves personal plans directed toward real-world goals. These researchers found that autobiographical planning involves "synchronized activity of **medial temporal lobe memory structures as well as frontal executive regions**, . . . specifically, of the **default** and **frontoparietal control networks**. [The **default network** consists of] the **medial prefrontal cortex** (PFC), **medial parietal cortex**, **including posterior cingulate cortex** (PCC) and **retrosplenial cortex** (RSC), the **posterior inferior parietal lobule** (IPL), **medial temporal lobes** (MTL) [**including hippocampus, amygdala, and parahippocampal** regions], and **lateral temporal cortex**. . . [This network is activated] by self-generated thought and active across multiple functional domains including memory, future-thinking, and social cognition" (Spreng et al., 2015).

Earlier research revealed that the default network may be "a common 'core' network that" . . . "underlies both remembering and imagining." Functional MRI (fMRI) studies "reveal striking overlap in the brain activity associated with remembering actual past experiences and imagining or simulating possible future experiences" (Schacter, et al., 2012, p. 677-678). Clinical observations are consistent with this hypothesis from fMRI studies. For example, *amnesic patients have difficulty imagining the future*. Another study of amnesic patients with **hippocampal damage** revealed impairments when these patients were asked to *imagine* novel experiences (Schacter, et al., 2012). The abilities for imagining (see section below) and future-thinking are very important components of human intelligence.

14.2.4.5 How the World Works and Intelligent Control of Movement

Considering the complexity and variety of the many objects, forces and relations interacting in the physical and social worlds, it is not surprising that intelligence involves the integration of many brain areas acting together. As mentioned earlier, representation of three-dimensional space, including the position of one's own limbs, is dependent upon regions within the **parietal lobe** and one's location in local space is coded by **hippocampal "place cells."** Visual representation of objects depends upon the visual cortex in the **occipital lobe**, while visual recognition requires the **inferotemporal lobe** (lower temporal lobe, also known as **IT cortex**).

Human social interactions require other brain areas. A relatively new specialty called **social cognitive neuroscience** studies the brain systems involved in the information processing involved in social interactions, including some subcortical structures, several of which are part of the **limbic system**. Other researchers study neurocognitive adaptations underlying **intuitive (or "folk") psychology**, which humans use for inferring **social causality** in other humans, helping us understand their motives and predict their likely future actions. Others study **intuitive (or "folk") physics**, for innately inferring **physical causality** among inanimate objects in one's environment helping us anticipate future events (Baron-Cohen, et al., 2001; Kubricht et al., 2017). As already mentioned, developmental psychologists are studying young children's ability to use statistical regularities in the environment to discover the cause-effect relations between physical events (see Gopnik et al., 2001; Gopnik, 2012; Gopnik & Wellman, 2012).

To form mental models of the world to guide intelligent behavior, all of this information and information from additional sources must be combined. Rabini, et al. (2021) provide evidence that the **precuneus area** (see Figure 14.1.6) of the **parietal lobe**, which underwent a *significant enlargement in fairly recent human evolutionary history*, is critically involved in the mental combination of different mental concepts, in part, to create new more complex concepts and ideas, and perhaps to help generate integrated mental models of the environment. Among other things, the **precuneus** region (what is often referred to as the **superior lobule of the parietal lobe** which equates with the outer extension of the **anterior and posterior precuneus**) of the parietal lobe may have set

the stage for the evolution of complex language and thinking by its involvement in the *linking of concepts and ideas into increasingly complex ones* in sophisticated human thought and complex cognition (Rabini, et al., 2021).

Planning for future ultimately involves action in the world and testing out the accuracy of one's mental models against feedback from the world. As Geertz (2016, p. 185) states:

> The process of comparing models with perceptual input from the wider world is known today as “prediction error monitoring” (Frith 2007: 132ff.; Frith & Frith 2010). Our experience of the world is in fact our experience of the brain’s simulation of the world. The brain constantly predicts what is going on in the physical world, the body and, especially, the social world. These simulations are tested against the input of neurological mappings, perceptual input and social mappings. When it detects errors in its predictions, the brain attempts to improve its predictions.

Nave et al. (2020) express similar ideas. But attempts to improve predictions based on feedback from the world is only half the story. Not only does the brain attempt to improve its predictions, but it also directs the modification of behavior, based on feedback from the environment, to make corrections in action needed to keep behavior on track toward the achievement of an imagined goal. Continuous feedback from the environment about the effects of one's own behavior guides behavior toward imagined goals, with needed corrections along the way. For example, if you imagine going to law school and your grade on a midterm exam in one of your pre-law courses is below average, then this feedback from the environment that does not match your mental model of what is required to get accepted into law school can cause you to modify your study habits. This type of monitoring of the match or mismatch between an imagined goal and the actual effects of one's actions with respect to the goal, an important feature of intelligent action, involves interaction between the **prefrontal cortex** and the **hippocampus.** As Numan (2015, p. 323) states:

> Action plans are essential for successful goal-directed behavior, and are elaborated by the prefrontal cortex. When an action plan is initiated, the prefrontal cortex transmits an efference copy (or corollary discharge) to the hippocampus where it is stored as a working memory for the action plan (which includes the expected outcomes of the action plan). The hippocampus then serves as a response intention-response outcome working memory comparator. Hippocampal comparator function is enabled by the hippocampal theta rhythm allowing the **hippocampus** to compare expected action outcomes to actual action outcomes. If the expected and actual outcomes match, the hippocampus transmits a signal to **prefrontal cortex** which strengthens or consolidates the action plan. If a mismatch occurs, the hippocampus transmits an error signal to the prefrontal cortex which facilitates a reformulation of the action plan, fostering behavioral flexibility and memory updating.

This sort of monitoring of the effects of one's actions comparing those effects to the effects desired to reach a goal is an essential part of the guidance systems for movement in complex animals, especially in humans, and is a central feature guiding intelligent action.

Above, we referred to *cognitive instincts*. Another of these instincts is emotional/motivational--what we commonly refer to as curiosity. As argued above, we are innately predisposed to understand cause-effect and to seek causal explanations of the world, "curiosity." As Thagard (2010, p.458-9) states: "abductive inference . . . leads people to respond to surprising observations with a search for hypotheses that can explain them. . . Brains make mental models through complex patterns of neural firing and use them in many kinds of inference, from planning actions to the most creative kinds of abductive reasoning." Understanding causality is not much use if you are not equipped with brain mechanisms which drive the motivation to know why things happen in your environment; that drive to explain, is the drive to identify causes. Knowing the causes of things allows better prediction, which leads to understanding of how to manipulate the environment to adaptive advantage--intelligence put to action.

14.2.5 General Intelligence: Adaptive Specializations to Abstract Relational Regularities

As described above, some of the most biologically important features of the world, of the environment, are abstract and relational. Recall that Spearman believed that *g* was most closely related to the "drawing out" of "relations and correlates,” abstract regularities of the world important in inductive and deductive logic, grasping relationships, inferring rules, and recognizing differences and similarities. Also discussed above, several of these broad, abstract properties of the world were central to the evolution of general intelligence and these have been genetically internalized into brain organization over the course of evolution. As previously noted, these include: 1) Causation: things/events in the world are governed by natural causal laws, and thus, causal relations among its elements are everywhere. Causality in the world holds important adaptive information. Innate knowledge about

causality, by its genetic internalization, gives the mind ability to discover and exploit the biologically important information in causal relations, thereby improving adaptation; 2) Similarity relations: things/events in the world have similarities to one another in multiple dimensions and at varying degrees of abstraction (e.g. "there is nothing new under the sun"). Similarities hold important adaptive information extracted by the grouping of things into categories and concepts of varying degrees of abstraction; 3) Predictive covariation of events: things/events in the world covary or co-occur with one another in predictable ways, offering opportunity for prediction, and serving as an important cue to causality (Gopnik, et al., 2001; Gopnik & Wellman, 2012), both highly significant for successful adaptation. Genetic internalization of knowledge about these abstract relational regularities (causality similarity, and predictive relations) set the groundwork for the evolution of categorization, concept formation, generalization, causal and categorical logic and inference, and analogical reasoning, all of which are potent components of cognition and general intelligence (Koenigshofer, 2017).

From this perspective, **general intelligence** (Spearman's "g;" see discussion above and section 18.11) evolved as a **collection of adaptive specializations** (specific psychological adaptations to a particular class of environmental feature, relation, or adaptive problem; see Gallistel, 1995) **to** these universal, **abstract relational regularities (causality, similarity, and predictive relations)** of the world which repeatedly appear, generation after generation, in virtually all situations, adaptive problems, and adaptive opportunities in the environment. The adaptive information in causality, similarity, and predictive relations among events was far too significant to be missed by natural selection. Specialized brain mechanisms evolved to extract and utilize the information in these abstract relational regularities--forming the foundations of general intelligence in us and many other animals. The high levels of general intelligence that evolved in our species gave us ability to discover and understand general principles of how the world works that could be applied to a wide range of environmental problems and adaptive opportunities. General intelligence, becoming highly developed in recent human evolution (probably within the past 100,000 years; see Mellars, 2005), has allowed our species, unlike any other, to create the scientific method, systems of government and law, machines to do our work, modern medicine, agriculture, cities and nations, and space travel. Highly developed general intelligence combined with our capacities for cooperation and cultural transmission, both facilitated by language, set us apart from all other animals.

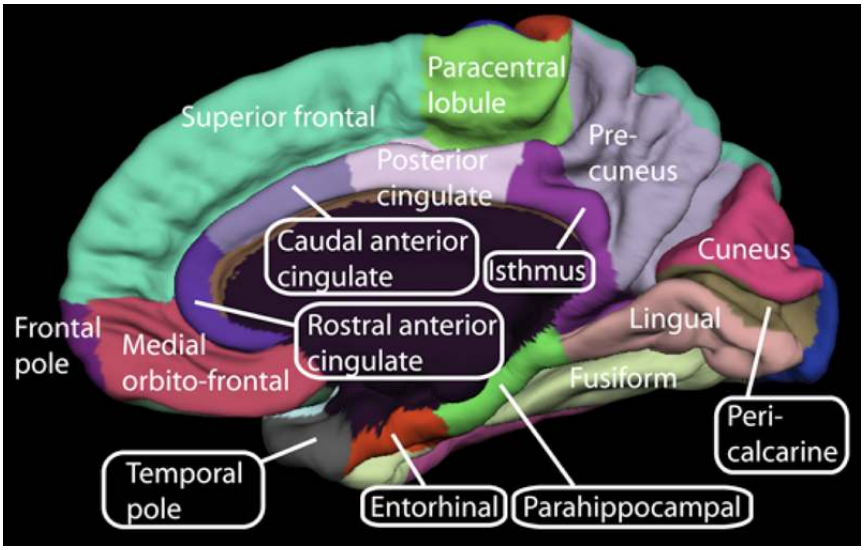


Figure **14.2.10**: Artist's depiction of midline cross-section of human brain showing some of the structures involved in generating some components of intelligence (see text above); the hippocampus proper, buried in temporal lobe cortex, is not shown because it is located behind the plane of the midline view shown here. (Image from Wikimedia Commons; https://commons.wikimedia.org/wiki/F...tex_-_gyri.png; by Patric Hagmann et.al.; licensed under the Creative Commons Attribution 2.5 Generic license; caption by Kenneth A. Koenigshofer, PhD).

14.2.6 General Intelligence in Humans and Non-human Animals: Neural Correlates

Historically, psychologists have focused on ways of measuring intellectual abilities that are important for success in school. Traditional IQ tests were designed to measure these abilities. As discussed above, one important observation from research based on these intelligence tests is that measures of different intellectual skills are correlated: individuals that do well in one type of

intellectual ability tend to do well in all types. These positive correlations among measures of different types of intellectual activity are known among psychologists as "the positive manifold." The "positive manifold" has been interpreted by many psychologists as strong statistical evidence (from the mathematical method, factor analysis; Spearman, 1904) for the general intelligence factor, "g," which exists in addition to more specific intelligences such as verbal fluency and visual-spatial abilities (see models of human intelligence in the next module).

Perhaps unexpectedly, recent research shows that general intelligence (ability to recognize "relations and correlates," according to Spearman,1904, 1925, who originated the concept) is found to some degree in many mammals and even in some birds (Emery & Clayton, 2004). Recent research shows that degrees of general intelligence among mammals is most strongly associated with "the **number of cortical neurons, neuron packing density, interneuronal distance and axonal conduction velocity**—factors that determine general information processing capacity (IPC), as reflected by general intelligence" (Dicke & Roth, 2016). These researchers compare IPC in various species. They report that: "The highest IPC is found in humans, followed by the great apes, Old World and New World monkeys. The IPC of cetaceans and elephants is much lower because of a thin cortex, low neuron packing density and low axonal conduction velocity. By contrast, corvid [rooks, crows, ravens] and psittacid birds [parrots] have very small and densely packed pallial neurons and relatively many neurons, which, despite very small brain volumes, might explain their high intelligence."

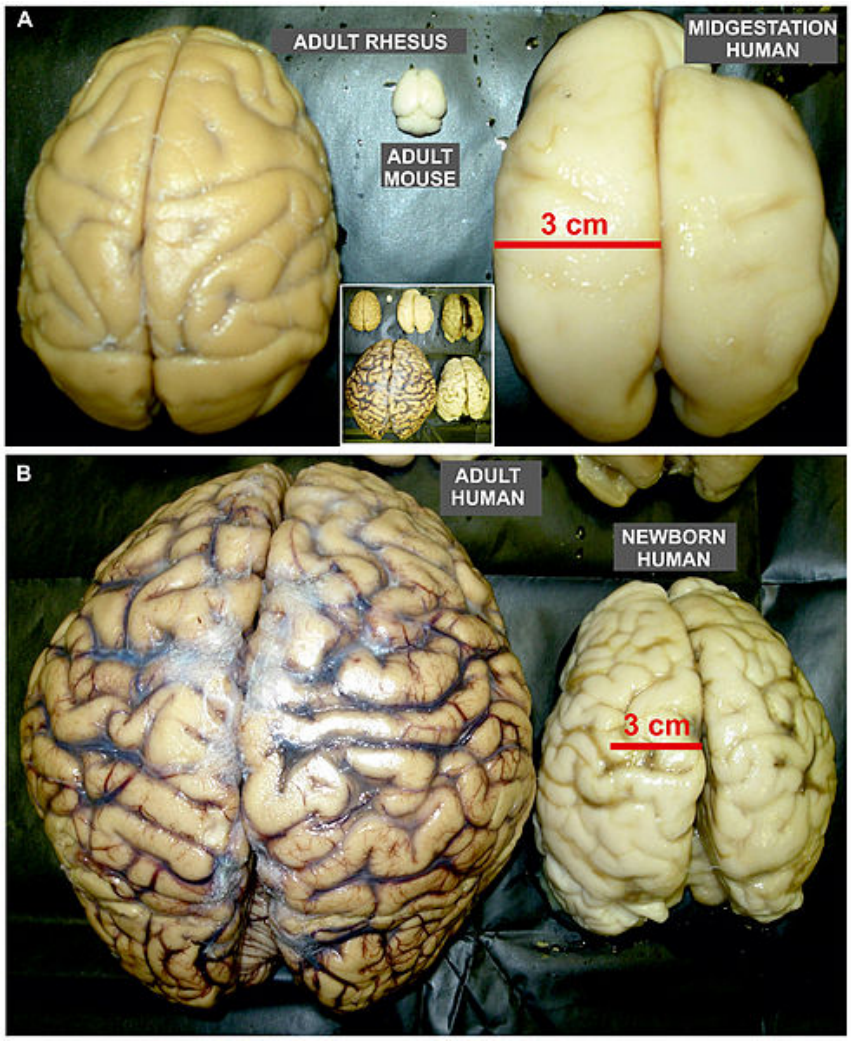


Figure **14.2.11**: Size proportion of mature rodent and non-human primate brain as well as developing and mature human brains. Dorsal view of adult mouse (A), rhesus monkey (A), and human brain (B), as well as human fetal brain around mid-gestation (A) and at term (B). (A) The size of human fetal brain already at mid-gestation has reached the size of the adult rhesus monkey brain. Nevertheless, adult rhesus monkey brain is almost 100 times larger than brain of adult mouse. (B) The adult human brain is around

3–4 times larger than newborn brain which size reaches the size of adult chimpanzee brain. Note that the pattern of gyrification (cerebral gyri--the hill portions of the cortical folds) in human newborn brain is close to the one observed in adult. The inlet in the middle of the figure is the integrative photo of all brains shown in A and B that demonstrate their actual proportions (the right brain in the upper row of insertion is from fetus at the beginning of third trimester of gestation). Image and caption from Wikimedia. https://commons.wikimedia.org/wiki/F...00050-g004.jpg; by Ana Hladnik, Domagoj Džaja, Sanja Darmopil, Nataša Jovanov-Milošević and Zdravko Petanjek; licensed under the Creative Commons Attribution 3.0 Unported license).

An approach to individual differences among humans in intelligence involves **efficiency of neural processing**--higher IQ people show greater processing efficiency (Haier et al., 1992; Van Den Heuvel et al., 2009). Brain imaging studies show that higher IQ individuals show lower brain activation (as indicated by glucose metabolism) than lower IQ individuals when completing the same cognitive task, indicating that individuals with higher intelligence, as measured by IQ tests, do information processing more efficiently, with less expenditure of metabolic energy by the brain, compared to lower IQ individuals whose brains work harder (their brains expend more metabolic energy) at solving the same problem. Haier, et al. (1992, p. 415-416) conclude: "Intelligence is not a function of how hard the brain works but rather how efficiently it works."

Brain processing efficiency may be related to any number of factors: inherited brain circuitry configurations which are more "streamlined" as a result of more efficient neural pruning during brain development (Koenigshofer, 2011, 2016); better conduction over pathways connecting multiple regions of brain; "more long-distance connections that ensure a high level of global communication efficiency within the overall network" (Van Den Heuvel, et al., 2009, p. 7619); less engagement of brain areas not required to solve a particular problem or complete a particular cognitive task (Haier, et al., 1992); and/or as Hearne, et al. (2016) conclude, "higher global network efficiency is related to higher general intelligence measures" with "a key role for connections between **prefrontal and frontal cortices** comprising the **dorsal attention network**" and with activity also in "**posterior cingulate/precuneus** . . . the **superior parietal cortex (fronto-parietal network)** and the **occipital cortices (visual network)**. Resting state functional connectivity between bilateral prefrontal cortices encompassing the dorsal attention network and the **right insula (salience network)** was also associated with intelligence scores."

In humans, **general intelligence** is heritable and is dependent upon **many genes interacting** together (Bouchard, 2014). One can wonder whether some of these many genes determining intelligence might hold information about some of the environmental regularities in the world discussed above, including relational regularities such as causality and similarity. This is an area of research wide open for researchers in behavior genetics or cognitive genetics.

When considering human intelligence, the role of language is a significant factor. Much of human thinking involves the use of words. As Dicke and Roth (2016) state, "The evolution of a syntactical and grammatical language in humans most probably has served as an additional intelligence amplifier, which may have [also] happened in songbirds and psittacids" [parrots] as a consequence of convergent evolution--the evolution of similar characteristics in unrelated or distantly related species because of common selection pressures. As mentioned above, the **precuneus** of the parietal lobe may have played a significant role in the evolution of language by having involvement in the combination of ideas and concepts from different semantic domains (Rabini, et al., 2021).

Studies of brain damaged patients supply further insight into brain mechanisms involved in abstract concept formation and thinking dependent upon language. Patients with one type of **frontotemporal** dementia (FTD) are particularly impaired "in verbal concept formation (i.e. categorization based on abstract similarities between items). . . [with] especially the **left frontal lobe**, thought to be involved in abstract word processing (Lagarde, et al., 2015, p. 456).

14.2.7 Another Component of General Intelligence: Visual Imagery and Imagination

Another important component of general intelligence is ability to visualize--to form and manipulate visual images in imagination. Psychologists who study human intelligence using and analyzing intelligence tests label this the *Broad Visual Perception factor* (*Gv*), “which is an ability to generate, retain, retrieve and transform visual images” (Kvist and Gustafsson, 2008, p.422-423). This is one of the sub-factors of *g* (general intelligence) in Carroll’s (1993) widely accepted three-level model of human intelligence (see module on measuring human intelligence which follows). This visualization ability may have developed as an adaptation or perhaps as an "exaptation" (recruitment of an adaptation to a different function; see Gould and Vrba, 1982). On this view, portions of the visual system, including parts of **visual, motor, and frontal and parietal association cortex**, were recruited over evolutionary time to permit the formation and manipulation of visual images used to mentally test out probable effects of future

behaviors before committing to them in the real physical world. Much of what we call "thought" is of this form (Koenigshofer, 2011, 2016, 2017). Note that some studies described above found activity in the **cortical visual areas and frontal and parietal cortical areas** to be **associated with intelligence**. Other studies, discussed above in the section on mental models, implicate the "default network" in both imagination and memory.

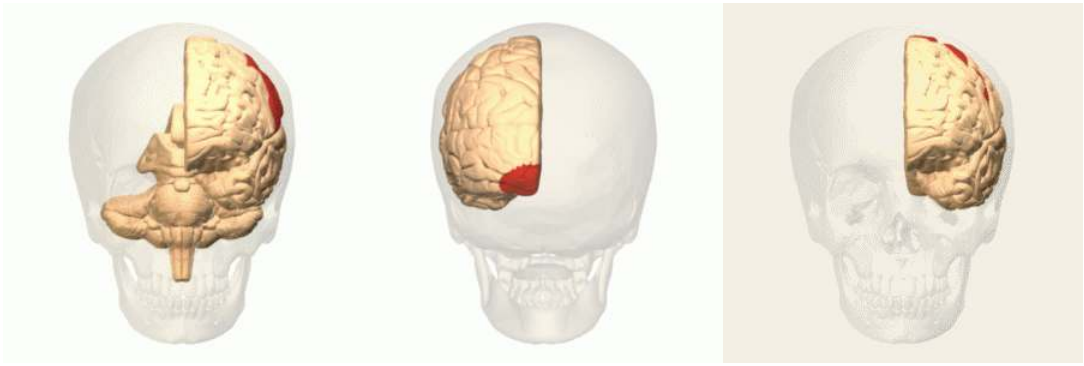


Figure **14.2.12**: (Left) Left Parietal Lobe; (Center) Left Primary Visual Cortex; the Visual association cortex (not colored) immediately surrounds primary visual cortex; (Right) Left Primary Motor Cortex; premotor cortex (not colored) is located just anterior to primary motor cortex. (Images from Wikimedia Commons; https://commons.wikimedia.org/wiki/F..._animation.gif; https://commons.wikimedia.org/wiki/F..._animation.gif; https://commons.wikimedia.org/w/inde...=Go&type=image; by Polygon data were generated by Database Center for Life Science (DBCLS); licensed under the Creative Commons Attribution-Share Alike 2.1 Japan license; caption by Kenneth A. Koenigshofer, PhD).

There is experimental evidence for this claim. For example, brain imaging studies show that when humans imagine a particular movement, the same brain region that becomes active during the actual movement also becomes active during the imagined movement, even though no actual movement occurs. In both cases, the activated brain area shows nearly the same fMRI image (Ganis, et al., 2004; Kanwisher, 2009; Kosslyn, et al., 2006) suggesting that similar circuits are involved in imagining a movement and actually carrying out the movement. Similar results using fMRI have been reported for visual perception and mental imagery in the "mind's eye" with the greatest overlap in activation patterns for perception and imagery in imagination occurring in the **frontal and parietal cortex** (Ganis, et al., 2004). Mental testing of planned actions in the mind's eye (guided by innate, implicit "knowledge" of cause-effect, predictive event covariation, and similarity relations) is highly adaptive because it is safe, fast, and efficient (Koenigshofer, 2017). Testing out possible behaviors and their likely effects in imagination can eliminate harmful or even fatal choices in the real world. In addition, mental testing of possible behavioral choices takes much less time and much less caloric expenditure than testing out behaviors by trial and error in the real physical world (think about mentally searching for your car keys until you find them mentally vs. actually moving to search your house, car, and garage to try to find them--the savings in time and in calories is clear). Because of these adaptive advantages, it is likely that there has been strong selection for neural mechanisms supporting visual imagination--a major component of human general intelligence and an integral part of what is generally referred to as "thinking" (Koenigshofer, 2011, 2016, 2017).

14.2.7.1 **Experimental Support for Mental Imagery**

14.2.7.1.1 Image Scanning

Seminal research in visual imagery was provided by Kosslyn's image-scanning experiments in the 1970s. Using the mental representation of a ship, subjects were instructed to shift their mental focus from one part of the ship to another. The reaction time of the subjects increased with distance between the two parts of the imagined ship, which indicates that we actually create a mental picture of scenes while trying to solve small cognitive tasks such as scanning the mental image of an object. Interestingly, Marmor and Zaback (1976) found that this visual ability can be observed also in people who are congenitally blind. Presuming, that the underlying processes are the same as in sighted subjects, it could be concluded that there is a deeper encoded system that has access to more than just the visual input.

14.2.7.1.2 Mental Rotation Task

Other advocates of visual spatial representation theory, Shepard and Metzler, developed the mental rotation task in 1971. Two spatially complex objects are presented to a participant at different angles and his/her job is to decide whether the objects are identical or not. The results show that the reaction times increases linearly with the rotation angle of the objects. The participants mentally rotate the objects in order to match the objects to one another. This method is called "mental chronometry".

This experiment was crucial for demonstrating the importance of imagery within cognitive psychology, because it showed the similarity of imagery to the processes of perception. For a mental rotation of 40° the subjects needed two seconds on average, whereas for a 140° rotation the reaction time increased to four seconds. Therefore, it can be concluded that people in general have a mental object rotation rate of 50° per second.

14.2.7.1.3 Spatial Frameworks

As mentioned above, researchers generally believe that mental models are perceptually based as indicated by fMRI studies which show extensive overlap between fMRI images of the perception of an object and its visualization. Indeed, people have been found to use spatial frameworks like those created for texts to retrieve spatial information about observed scenes (Bryant, 1991). Thus, people create the same sorts of spatial memory representations no matter if they read about an environment or see it themselves.

14.2.7.1.4 Size and the visual field

If an object is observed from different distances, it is harder to perceive details if the object is far away because the objects fill only a small part of the visual field. Kosslyn did an experiment in 1973 in which he wanted to find out if this is also true for mental images, to show the similarity of the spatial representation and the perception of real environment. He told participants to imagine objects which are far away and objects which are near. After asking the participants about details, he supposed that details can be observed better if the object is near and fills the visual field. He also told the participants to imagine animals with different sizes near by another. For example an elephant and a rabbit. The elephant filled much more of the visual field than the rabbit and it turned out that the participants were able to answer questions about the elephant more rapidly than about the rabbit. After that the participants had to imagine the small animal besides an even smaller animal, like a fly. This time, the rabbit filled the bigger part of the visual field and again, questions about the bigger animal were answered faster. The result of Kosslyn's experiments is that people can observe more details of an object if it fills a bigger part of their mental visual field. This provides evidence that mental images are represented visual-spatially.

14.2.7.2 Visual Imagery, Intelligence, and the Frontoparietal Network

Significantly, visual imagery in the mind’s eye can also be employed to mentally test novel combinations of causes and effects, similarity relations, and predictive relations (covariations) to discover new knowledge about hidden causes of events in the world, eventually leading to the creation of sophisticated scientific models of how the world works—human general intelligence at its best (Koenigshofer, 2017). For example, the famous theoretical physicist, Albert Einstein, used mental imagery extensively to do "thought experiments" that played a large role in his formulation of the theories of general and special relativity. Brain imaging studies using fMRI implicate **parietal cortex** in the use of imagination (Nair et al., 2003). Bruner (2010, p. S84) suggests that “the **parietal lobe system** ‘forms a neural image of surrounding space’ (Mountcastle, 1995 p. 389),” and perhaps of one’s potential future action in that space. Significantly, parietal cortex has strong linkages with **prefrontal cortex** forming a **frontoparietal network:** the **inferior parietal lobule** (further subdivided into the **supramarginal gyrus**, the **temporoparietal junction**, and the **angular gyrus**, see Figure 14.2.8) is primarily connected with **dorsolateral prefrontal cortex** (Bruner, 2010; see Figure 14.2.7), associated, in part, with abilities for abstract thought, while upper parietal regions, according to Bruner, are associated in the scientific literature with functions such as abstract representation, internal mental images, “imagined world[s],. . . and thought experiment” (i.e., imagination). Significantly, a number of other research groups using fMRI brain imaging techniques in humans also find evidence of a critical role of the **frontoparietal network** in human intelligence and general cognitive ability (Colom, et al., 2010; Jung & Haier, 2007; Sripada, et al., 2020; Vendetti & Bunge, 2014; Wendelken, et al., 2017; Yeo et al., 2016).

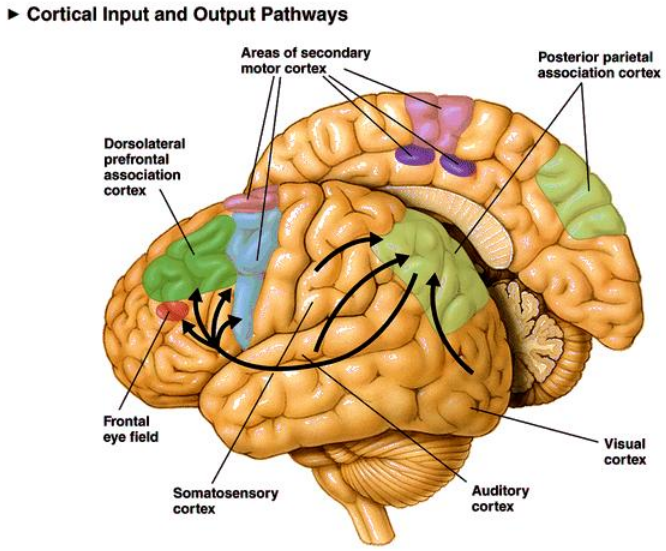


Figure 14.2.13: Connections of posterior parietal association cortex with dorsolateral prefrontal association cortex. (Image from Wikipedia; https://commons.wikimedia.org/wiki/F...ietal_Lobe.jpg; by Paskari; licensed under the Creative Commons Attribution-Share Alike 2.5 Generic, 2.0 Generic and 1.0 Generic license. Caption by Kenneth A. Koenigshofer, PhD.).

The emergence during human evolution of this visualization ability may help account for the development of modern human cognition, first appearing some 60,000-100,000 years ago (Mellars, 2005). Speculatively, superior abilities for imagination of the type described above might account, at least in part, for human competitive advantage over Neanderthals (Koenigshofer, 2017) perhaps contributing to Neanderthal extinction about 30,000 years ago (Watson and Berry, 2009). Consistent with this view, morphological studies suggest enhanced **parietal lobe** development in modern humans compared to Neanderthals (Bruner, 2010); by contrast, recent studies show little relative enlargement of the frontal lobes in humans compared to apes (Barton and Venditti, 2013). At least one post-mortem anatomical study of Einstein's brain suggested unusual enlargement of the **inferior parietal lobules** in parietal lobes along with a number of other atypical features in the **occipital lobes**, and throughout the cerebral cortex (Chen et al., 2014; see Carrillo-Mora et al., 2015). Another study of Einstein's brain by Falk et al. (2013, p. 1304) reports "Einstein's brain has an extraordinary prefrontal cortex, which may have contributed to the neurological substrates for some of his remarkable cognitive abilities. The primary somatosensory and motor cortices near the regions that typically represent face and tongue are greatly expanded in the left hemisphere. Einstein's parietal lobes are also unusual and may have provided some of the neurological underpinnings for his visuospatial and mathematical skills, as others have hypothesized. Einstein's brain has typical frontal and occipital shape asymmetries (petalias) and grossly asymmetrical inferior and superior **parietal lobules**" with relative enlargement of the left inferior parietal module. Chen et al., referred to above, found exceptional enlargement of both **inferior parietal lobules** (which include the **supramarginal gyrus and angular gyrus** (Figure 14.2.8), involved in multiple functions including number processing, "mentalizing," and spatial cognition) in Einstein's brain compared to controls. Einstein was famous for his use of visualization to explore his ideas in theoretical physics through "thought experiments" in his "mind's eye." Additional evidence regarding Einstein's exceptional parietal lobes was provided by studies of cortical glial cells. A study by Diamond et al. (1985) of Einstein's brain found that the only cortical area of the four cortical areas examined with a greater number of glial cells per neuron, compared to controls, was the **left posterior parietal cortex**, a finding interpreted by the authors to mean that this area of Einstein's brain must have been exceptionally active during his lifetime.

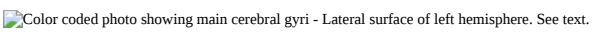


Figure 14.2.14: Cortical gyri, including the left inferior parietal lobule composed of the supramarginal and angular gyri. (Image from Wikimedia Commons; https://commons.wikimedia.org/wiki/F...al_Surface.png; by John A. Beal, Ph.D.; licensed under the Creative Commons Attribution 2.5 Generic license.

This capacity for visualization permitting mental trial and error likely occurs to some degree in non-human animal species as well. Kohler's (1925/1959) classic work showing the use of "insight" by chimpanzees in problem solving offers evidence for this

hypothesis, assuming that insight involves mental trial and error. Additional evidence suggesting mental trial and error in imagination in a variety of animals is provided by experiments showing insight in some orangutans, gorillas, and chimpanzees (Mendes, et al., 2007; Hanus, et al., 2011) and in rooks--birds in the same Corvid family as ravens and crows (Bird & Emery, 2009a, 2009b).

One key feature of more traditional approaches to the study of human intelligence is that they emphasize aspects of human intelligence and thinking, such as language and computation, that are especially important for life in complex modern society. As described in the next module, some psychologists have hypothesized multiple kinds of human intelligence. Others focus primarily on how to measure intelligence, and still others study individual differences in intelligence among people. Because these psychologists tend to be interested in practical application of their tests in school and work settings, they tend to give relatively little attention to theories about the evolutionary origins or biological functions of intelligence. Nevertheless, intelligence tests are very useful predictors of success in environments such as school, military, and work settings. All involve measurement of mental skills important to successful adaptation to the demands of the physical and social environments typical of modern, industrialized societies.

14.2.8 Summary

Intelligence and cognition are complex psychological adaptations involving many different areas of the brain interacting together. Intelligence in the broadest sense involves the formation of sophisticated mental models of the world and how it works. These models include representations of things, events, and the predictive, similarity, and causal relations among them, as well as representations of space and time. Mental models of the physical and social environments permit prediction of future events, providing powerful evolutionary advantage. Mental models include the formation of categories, based on similarities, which allow inferences based on assignment of newly encountered things and events to previously formed categories. Mental models exploit the innate disposition to understand cause-effect relations in the environment, leading to knowledge about what causes what. This knowledge allows manipulation of causes to affect future environmental outcomes toward better adaptation. General intelligence emerges during evolution from genetic internalization of universal, abstract relational features of the world including cause-effect, event covariation, and similarity, each of which plays a role in thought and reasoning about the environment (Koenigshofer, 2017). Spearman (1904, 1925) discovered and was first to describe general intelligence, which he believed consisted of ability to recognize "relations and correlates" in the environment. Another aspect of general intelligence is ability to visualize, to form and manipulate visual images in imagination. This ability may have developed as an exaptation or recruitment of portions of the visual and motor systems, including parietal association cortex. This would allow some visual and motor circuits to form internally generated mental images that can be mentally manipulated to test out probable effects of possible future behaviors before committing to them in the physical world. Because this ability to visualize outcomes in the "mind's eye" is much safer and saves time and calories compared to testing outcomes in actual physical behavior, it is likely that there has been strong selection pressure for ability to manipulate mental images to plan future action. This component of intelligence provides a powerful mechanism for maximizing the adaptive outcomes of behavior and therefore provides enormous selective advantage--a strong impetus for its evolution as a central feature of intelligence (Koenigshofer, 2017).

14.2.9 References

Adams, S. S., & Burbeck, S. (2012). Beyond the octopus: From general intelligence toward a human-like mind. In *Theoretical foundations of artificial general intelligence* (pp. 49-65). Atlantis Press, Paris.

Amodio, P. (2019). Octopus intelligence: the importance of being agnostic. Animal Sentience 26(1). 272

Anderson, J. R. (2009). *How can the human mind occur in the physical universe?*. Oxford University Press.

Baron-Cohen, S., Wheelwright, S., Spong, A., Scahill, V., & Lawson, J. (2001). Are intuitive physics and intuitive psychology independent? A test with children with Asperger Syndrome. *Journal of developmental and learning disorders*, *5* (1), 47-78.

Barton, R. A., and Venditti, C. (2013). Human frontal lobes are not relatively large. *Proceedings of the National Academy of Sciences*, *110* (22), 9001-9006.

Behrens, T. E., Muller, T. H., Whittington, J. C., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron*, *100*(2), 490-509.

Bird, C. D., & Emery, N. J. (2009a). Rooks use stones to raise the water level to reach a floating worm. *Current Biology*, *19* (16), 1410-1414.

Bird, C. D., & Emery, N. J. (2009b). Insightful problem solving and creative tool modification by captive nontool-using rooks. *Proceedings of the National Academy of Sciences*, *106* (25), 10370-10375.

Blakemore, S., Fonlupt, P., Pachot-Clouard, M., Darmon, C., Boyer, P., Meltzoff, A., Segebarth, C., Decety, J. (2001). How the brain perceives causality: an event-related fMRI study, NeuroReport 12: 3741 – 3746.

Bouchard, T. J. (2014). Genes, evolution and intelligence. *Behavior genetics*, 44(6), 549-577.

Broschard, M. B., Kim, J., Love, B. C., Wasserman, E. A., & Freeman, J. H. (2021). Prelimbic cortex maintains attention to category-relevant information and flexibly updates category representations. *Neurobiology of Learning and Memory*, 107524

Bruner, E. (2010). Morphological differences in the parietal lobes within the human genus. *Current Anthropology*, *51* (S1), S77-S88

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences, 36(3), 181–204.

Clark, A. (2016). Surfing uncertainty: Prediction, action, and the embodied mind. New York: Oxford University Press.

Carrillo-Mora, P., Magaña-Vázquez, K., May-López, S. S., & Mondragón-Ramírez, N. A. (2015). What did Einstein have that I don't? Studies on Albert Einstein's brain. *Neurosciences and History*; 3(3): 125-129; https://nah.sen.es/vmfiles/abstract/...5125_129EN.pdf; retrieved 12/23/21.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies* (No. 1). Cambridge University Press.

Cattell, R. B. (1987). Intelligence: Its structure, growth and action. New York: Elsevier Science.

Chen, H., Chen, S., Zeng, L., Zhou, L., & Hou, S. (2014). Revisiting Einstein's brain in Brain Awareness Week. *Bioscience trends*, *8* (5), 286-289.

Churchland, P. M. (1985, 2013). *Matter and consciousness*. MIT press. Cambridge, Mass.

Colom, R., Karama, S., Jung, R. E., & Haier, R. J. (2010). Human intelligence and brain networks. *Dialogues in clinical neuroscience*, *12* (4), 489.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), The adapted mind: Evolutionary psychology and the generation of culture (pp. 163–228). New York: Oxford University Press.

Cosmides, L., & Tooby, J. (2003). "Evolutionary Psychology: Theoretical Foundations." In: *Encyclopedia of Cognitive Science*. Ed. Lynn Nadel. London: Macmillan, 54–64.

Cosmides, L., & Tooby, J. (2007). Evolutionary psychology: A primer. https://www.cep.ucsb.edu/primer.html; retrieved 12/23/2021.

Dawkins, Richard (1976). *The Selfish Gene*. Best Books.

DeLoache, J. S., & LoBue, V. (2009). The narrow fellow in the grass: Human infants associate snakes and fear. *Developmental science*, *12* (1), 201-207.

De Pasquale, R., Beckhauser, T. F., Hernandes, M. S., & Britto, L. R. G. (2014). LTP and LTD in the visual cortex require the activation of NOX2. *Journal of Neuroscience*, *34* (38), 12778-12787.

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual review of psychology*, *66*, 115-142.

Diamond MC, Scheibel AB, Murphy GM Jr, Harvey T. On the brain of a scientist: Albert Einstein. Exp Neurol. 1985;88:198-204.

Dicke U, Roth G. (2016). Neuronal factors determining high intelligence. *Phil. Trans. R. Soc. B* 371: 20150180. http://dx.doi.org/10.1098/rstb.2015.0180; retrieved September 14, 2021.

Daw, N., Rao, Y., Wang, X. F., Fischer, Q., & Yang, Y. (2004). LTP and LTD vary with layer in rodent visual cortex. *Vision research*, *44*(28), 3377-3380.

Durham, W. H. (1991). Coevolution: Genes, culture and human diversity. Palo Alto, CA: Stanford University Press.

Ellis, B. J., & Ketelaar, T. (2002). Commentary: Clarifying the Foundations of Evolutionary Psychology: A Reply to Lloyd and Feldman. *Psychological Inquiry*, *13* (2), 157-164.

Emery, N. J., and Clayton, N. S. (2004). The mentality of crows: convergent evolution of intelligence in corvids and apes. *Science*, *306* (5703), 1903-1907.

Falk, D., Lepore, F. E., & Noe, A. (2013). The cerebral cortex of Albert Einstein: a description and preliminary analysis of unpublished photographs. *Brain*, *136* (4), 1304-1327.

Fonlupt, P. (2003). Perception and judgment of physical causality involve different brain structures, *Cogn. Brain Res.* 17, 248 – 254.

Frith, C. (2007). *Making Up the Mind: How the Brain Creates Our Mental World.* Oxford: Blackwell.

Frith, U. & Frith, C. D. (2010). The Social Brain: Allowing Humans to Boldly Go Where No Other Species Has Been. *Philosophical Transactions of the Royal Society B Biological Sciences 365(1537)*: 165–76

Fugelsang, J. A., & Dunbar, K. N. (2005). Brain-based mechanisms underlying complex causal thinking. *Neuropsychologia*, *43* (8), 1204-1213.

Fugelsang, J. A., Roser, M. E., Corballis, P. M., Gazzaniga, M. S., & Dunbar, K. N. (2005). Brain mechanisms underlying perceptual causality. *Cognitive brain research*, *24* (1), 41-47.

Gallistel C.R. (1995). The replacement of general purpose theories with adaptive specializations. In M.S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (pp. 1255–67). Cambridge, MA: MIT Press.

Gangestad, S. W., & Simpson, J. A. (2000). The evolution of human mating: Trade-offs and strategic pluralism. *Behavioral and Brain Sciences*, *23*, 573–644.

Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Cognitive Brain Research*, *20*(2), 226-241.

Geertz, A. W. (2016). The meaningful brain. In: *Mental Culture: Classical Social Theory and the Cognitive Science of Religion.* Edited by Dimitris Xygalatas and William W. McCorkle Jr., 176-196.

Goldman-Rakic, P. S. (1996). The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*(1346), 1445-1453.

Goodall, J. (2000). *In the shadow of man* (Vol. 4113). Houghton Mifflin Harcourt.

Gopnik, A. (2010). How babies think. *Scientific American*, *303* (1), 76-81.

Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science*, *337* (6102), 1623-1627.

Gopnik, A., and Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child development*, *71* (5),1205-1222.

Gopnik, A., Sobel, D. M., Schulz, L. E., and Glymour, C. (2001). Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental psychology*, *37* (5), 620.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological bulletin*, *138* (6), 1085.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. Psychological Review, 111, 3–32.

Gould, S. J., & Vrba, E. S. (1982). Exaptation—a missing term in the science of form. *Paleobiology*, *8* (1), 4-15.

Haier, R. J., Siegel, B., Tang, C., Abel, L., & Buchsbaum, M. S. (1992). Intelligence and changes in regional cerebral glucose metabolic rate following learning. *Intelligence*, *16* (3-4), 415-426.

Hanus, D., Mendes, N., Tennie, C., & Call, J. (2011). Comparing the performances of apes (Gorilla gorilla, Pan troglodytes, Pongo pygmaeus) and human children (Homo sapiens) in the floating peanut task. *PloS one*, *6* (6), e19555.

Hawkins, J., Ahmad, S., & Cui, Y. (2017). A theory of how columns in the neocortex enable learning the structure of the world. *Frontiers in neural circuits*, 81.

Hayashi, M. J., Kantele, M., Walsh, V., Carlson, S., & Kanai, R. (2014). Dissociable neuroanatomical correlates of subsecond and suprasecond time perception. *Journal of Cognitive Neuroscience*, *26*(8), 1685-1693.

Hearne, L. J., Mattingley, J. B., & Cocchi, L. (2016). Functional brain networks related to individual differences in human intelligence at rest. *Scientific reports*, *6* (1), 1-8.

Huber, L., & Gajdon, G. K. (2006). Technical intelligence in animals: the kea model. *Animal cognition*, *9* (4), 295-305.

Hume, D. (1748/1988). *An Enquiry Concerning Human Understanding*. La Salle, IL: Open Court (Originally published 1748).

Ivry, R. B., & Keele, S. W. (1989). Timing functions of the cerebellum. *Journal of cognitive neuroscience*, *1* (2), 136-152.

Ivry, R. B., Spencer, R. M., Zelaznik, H. N., & Diedrichsen, J. (2002). The cerebellum and event timing. *Annals of the new York Academy of Sciences*, *978* (1), 302-317.

Ivry, R. B., & Spencer, R. M. (2004). Evaluating the role of the cerebellum in temporal processing: beware of the null hypothesis. *Brain*, *127* (8), e13-e13.

Jung, R. E., & Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behavioral and Brain Sciences*, *30* (2), 135-154.

Kanwisher, N. (2009, May). The Specialized Brain. Keynote Address. Annual Meeting of the Association for Psychological Science. San Francisco.

Kaufman, S. B., DeYoung, C. G., Reis, D. L., and Gray, J. R. (2011). General intelligence predicts reasoning ability even for evolutionarily familiar content. Intelligence, 39(5), 311–322.

Khemlani, S. S., Barbey, A. K., & Johnson-Laird, P. N. (2014). Causal reasoning with mental models. *Frontiers in human neuroscience*, *8*, 849.

Koenigshofer, K. A. (2011). *Mind Design: The Adaptive Organization of Human Nature, Minds, and Behavior*. Pearson Education, Boston.

Koenigshofer, K. A. (2016). *Mind Design: The Adaptive Organization of Human Nature, Minds, and Behavior, Revised Edition*. Amazon e-book.

Koenigshofer, K. A. (2017). General Intelligence: Adaptation to Evolutionarily Familiar Abstract Relational Invariants, Not to Environmental or Evolutionary Novelty. *The Journal of Mind and Behavior*, 38 (2), 119-153.

Köhler, W. (1925/1959). *The mentality of apes* (E. Winter, Trans.). New York, USA: Vintage Books

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The Case for Mental Imagery*. Oxford University Press

Kubricht, J. R., Holyoak, K. J., & Lu, H. (2017). Intuitive physics: Current research and controversies. *Trends in cognitive sciences*, *21*(10), 749-759.

Kvist, A. V., and Gustafsson, J. E. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's investment theory. *Intelligence*, *36* (5), 422-436.

Lagarde, J., Valabregue, R., Corvol, J. C., Garcin, B., Volle, E., Le Ber, I., ... & Levy, R. (2015). Why do patients with neurodegenerative frontal syndrome fail to answer:'In what way are an orange and a banana alike?'. *Brain*, *138*(2), 456-471.

Lara, A. H., & Wallis, J. D. (2015). The role of prefrontal cortex in working memory: a mini review. *Frontiers in systems neuroscience*, *9*, 173.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality?. *Cognition*, *25*(3), 265-288.

Lloyd, E. A., & Feldman, M. W. (2002). Commentary: Evolutionary psychology: A view from evolutionary biology. *Psychological Inquiry*, *13* (2), 150-156.

Mather, J. (2019). What is in an octopus's mind? *Animal Sentience 26* (1).

Mather, J. A., and Dickel, L. (2017). Cephalopod complex cognition. *Current Opinion in Behavioral Science 16*, 131–137. doi:10.1016/J.COBEHA.2017.06.008.

Mellars, P. (2005). The impossible coincidence. A single-species model for the origins of modern human behavior in Europe. *Evolutionary Anthropology: Issues, News, and Reviews*, *14* (1), 12-27.

Mendes, N., Hanus, D., & Call, J. (2007). Raising the level: Orangutans use water as a tool. Biology Letters, 3, 453– 455. doi:10.1098/rsbl.2007.0198

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, *24* (1), 167-202.

Nair, D. G., Purcott, K. L., Fuchs, A., Steinberg, F., and Kelso, J. S. (2003). Cortical and cerebellar activity of the human brain during imagined and executed unimanual and bimanual action sequences: a functional MRI study. *Cognitive brain research*, *15* (3), 250-260.

Nave K, Deane G, Miller M, Clark A. (2020). Wilding the predictive brain. WIREs *Cogn Sci.* 2020;11:e1542. https://doi.org/10.1002/wcs.1542

Numan, R. (2015). A prefrontal-hippocampal comparator for goal-directed behavior: the intentional self and episodic memory. *Frontiers in behavioral neuroscience*, *9*, 323.

Oakes, L. M. (1994). Development of infants' use of continuity cues in their perception of causality. *Developmental Psychology*, *30* (6), 869.

Oishi, N., Nomoto, M., Ohkawa, N., Saitoh, Y., Sano, Y., Tsujimura, S., ... & Inokuchi, K. (2019). Artificial association of memory events by optogenetic stimulation of hippocampal CA3 cell ensembles. *Molecular brain*, *12* (1), 1-10.

Panksepp, J., & Panksepp, J. B. (2000). The seven sins of evolutionary psychology. *Evolution and cognition*, *6* (2), 108-131.

Penn, D. C., and Povinelli, D. J. (2007). Causal Cognition in Human and Nonhuman Animals: A Comparative, Critical Review. Annual Review of Psychology 58, 97–118. doi:10.1146/annurev.psych.58.110405.085555.

Penn, D. C., Holyoak, K. J., and Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, *31* (02), 109-130.

Pinker, S. (2007). *The stuff of thought: Language as a window into human nature*. New York: Penguin Books.

Rabini, G., Ubaldi, S., & Fairhall, S. L. (2021). Combining concepts across categorical domains: a linking role of the precuneus. *Neurobiology of Language*, 1-50.

Reader, S. M., Hager, Y., and Laland, K. N. (2011). The evolution of primate general and cultural intelligence. Philosophical Transactions of the Royal Society B: Biological Sciences, 366(1567), 1017–1027. doi: 10.1098/rstb.2010.0342

Roser, M. E., Fugelsang, J. A., Dunbar, K. N., Corballis, P. M., & Gazzaniga, M. S. (2005). Dissociating processes supporting causal perception and causal inference in the brain. *Neuropsychology*, *19* (5), 591.

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: remembering, imagining, and the brain. *Neuron*, *76* (4), 677-694.

Schafer, M., & Schiller, D. (2020). The Brain's Social Road Maps. *Scientific American*, *322* (2), 31-35.

Scholl, B. J., & Nakayama, K. (2002). Causal capture: Contextual effects on the perception of collision events. *Psychological science*, *13* (6), 493-498.

Soto, F. A., & Wasserman, E. A. (2012). Visual object categorization in birds and primates: Integrating behavioral, neurobiological, and computational evidence within a "general process" framework. *Cognitive, Affective, & Behavioral Neuroscience*, *12* (1), 220-240.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *The American Journal of Psychology*, 15(2), 201-292.

Spearman, C. (1925). Some Issues in the Theory of`'g' (including the Law of Diminishing Returns). *Nature*, 116, 436-439.

Shepard, R.N. (1987a). Evolution of a mesh between principles of the mind and regularities of the world. In J. Dupre (Ed.), *The Latest and the Best: Essays on Evolution and Optimality.* Cambridge, MA: MIT Press.

Shepard, R.N. (1987b). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317-1323.

Shepard, R.N. (1992). The perceptual organization of colors: an adaptation to regularities of the terrestrial world? In J.H. Barkow, L. Cosmides, and J. Tooby (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press.

Shepard, R.N. (1994). Perceptual-cognitive universals as reflections of the world. *Psychonomic Bulletin and Review 1*:2-28. .

Shepard, R. N. (2001). Perceptual-cognitive universals as reflections of the world. *Behavioral and brain sciences*, *24* (04), 581-601.

Siegel, M., Buschman, T. J., and Miller, E. K. (2015). Cortical information flow during flexible sensorimotor decisions. *Science* 348, 1352–1355. doi: 10.1126/science.aab0551

Spreng, R. N., Gerlach, K. D., Turner, G. R., & Schacter, D. L. (2015). Autobiographical planning and the brain: Activation and its modulation by qualitative features. *Journal of cognitive neuroscience*, *27* (11), 2147-2157.

Sripada, C., Angstadt, M., Rutherford, S., Taxali, A., & Shedden, K. (2020). Toward a “treadmill test” for cognition: Improved prediction of general cognitive ability from the task activated brain. *Human brain mapping*, *41*(12), 3186-3197.

Striedter, G. F. (2005). *Principles of brain evolution*. Sinauer associates.

Thagard, P. (2010). How brains make mental models. In *Model-based reasoning in science and technology* (pp. 447-461). Springer, Berlin, Heidelberg.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological review*, *55* (4), 189.

Tooby, J., & Cosmides, L. (2015). The theoretical foundations of evolutionary psychology. *The handbook of evolutionary psychology: Foundations (2nd ed., Vol. 1, pp. 3–87). Hoboken, NJ: John Wiley*.

Van Den Heuvel, M. P., Stam, C. J., Kahn, R. S., & Pol, H. E. H. (2009). Efficiency of functional brain networks and intellectual performance. *Journal of Neuroscience*, *29* (23), 7619-7624.

Vendetti, M. S., & Bunge, S. A. (2014). Evolutionary and developmental changes in the lateral frontoparietal network: a little goes a long way for higher-level cognition. *Neuron*, *84* (5), 906-917.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., ... & Miller, B. L. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological science*, *10* (2), 119-125.

Wasserman, E. A., Wasserman, E. A., & Zentall, T. R. (Eds.). (2006). *Comparative cognition: Experimental explorations of animal intelligence*. Oxford University Press, USA.

Watson, J. D., and Berry, A. (2009). *DNA: The secret of life*. New York: Knopf.

Wendelken, C., Ferrer, E., Ghetti, S., Bailey, S. K., Cutting, L., & Bunge, S. A. (2017). Frontoparietal structural connectivity in childhood predicts development of functional connectivity and reasoning ability: A large-scale longitudinal investigation. *Journal of Neuroscience*, *37* (35), 8549-8558.

Yeo, R. A., Ryman, S. G., Pommy, J., Thoma, R. J., & Jung, R. E. (2016). General cognitive ability and fluctuating asymmetry of brain surface area. *Intelligence*, *56*, 93-98.

14.2.10 Attributions

Section 14.2, "Intelligence, Cognition, and Language as Psychological Adaptations," is original material written by Kenneth A. Koenigshofer, PhD, licensed under CC BY 4.0, with the exception of the section titled "Experimental Evidence for Mental

Imagery" which is adapted by Kenneth A. Koenigshofer, Ph.D., from *Cognitive Psychology and Cognitive Neuroscience*, Wikibooks, https://en.wikibooks.org/wiki/Cognit...cience/Imagery; text is available under the Creative Commons Attribution-ShareAlike License

Images from Wikimedia Commons.

This page titled 14.2: Cognition, Intelligence, and Language as Psychological Adaptations is shared under a mixed license and was authored, remixed, and/or curated by Kenenth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

14.3: High Intelligence-Nature or Nurture; Learning Disabilities

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

14.3.1 Learning Objectives

1. Describe how genetics and environment affect intelligence
2. Explain the relationship between IQ scores and socioeconomic status
3. Describe the difference between a learning disability and a developmental disorder

14.3.2 Overview

A young girl, born of teenage parents, lives with her grandmother in rural Mississippi. They are poor—in serious poverty—but they do their best to get by with what they have. She learns to read when she is just 3 years old. As she grows older, she longs to live with her mother, who now resides in Wisconsin. She moves there at the age of 6 years. At 9 years of age, she is raped. During the next several years, several different male relatives repeatedly molest her. Her life unravels. She turns to drugs and sex to fill the deep, lonely void inside her. Her mother then sends her to Nashville to live with her father, who imposes strict behavioral expectations upon her, and over time, her wild life settles once again. She begins to experience success in school, and at 19 years old, becomes the youngest and first African-American female news anchor (“Dates and Events,” n.d.). The woman—Oprah Winfrey—goes on to become a media giant known for both her intelligence and her empathy.

14.3.3 High Intelligence: Nature or Nurture?

Where does high intelligence come from? Some researchers believe that intelligence is a trait inherited from a person’s parents. Scientists who research this topic typically use twin studies to determine the heritability of intelligence. The Minnesota Study of Twins Reared Apart is one of the most well-known twin studies. In this investigation, researchers found that identical twins (100% of their DNA is identical) raised together and identical twins raised apart exhibit a higher correlation between their IQ scores than siblings (50% of their DNA in common) or fraternal twins (50% of their DNA in common) raised together (Bouchard, Lykken, McGue, Segal, & Tellegen, 1990). The findings from this study reveal a genetic component to intelligence (Figure 10.3.1). At the same time, other psychologists believe that intelligence is shaped by a child’s developmental environment. If parents were to provide their children with intellectual stimuli beginning from before they are born, it is likely that they would absorb the benefits of that stimulation, and it would be reflected in intelligence levels.

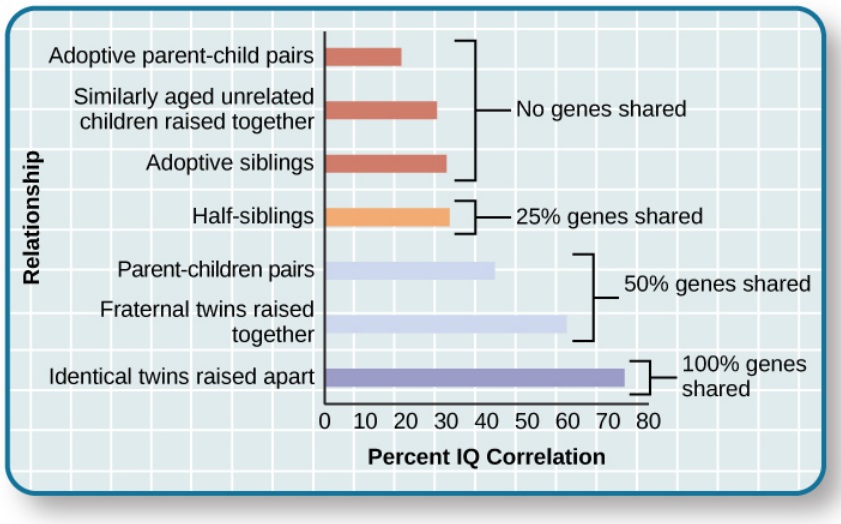


Figure **14.3.1**: The correlations of IQs of unrelated versus related persons reared apart or together suggest a genetic component to intelligence.

The reality is that aspects of each idea are probably correct. One study suggests that genetics seem to be in control of the level of intelligence, environmental influences are also involved (Bartels, Rietveld, Van Baal, & Boomsma, 2002). Certainly, there are behaviors that support the development of intelligence, but the genetic component of high intelligence can not be ignored. As with all heritable traits, however, it is not always possible to isolate how and when high intelligence is passed on to the next generation.

Range of Reaction is the theory that each person responds to the environment in a unique way based on his or her genetic makeup. According to this idea, your genetic potential is a fixed quantity, but whether you reach your full intellectual potential is dependent upon the environmental stimulation you experience, especially in childhood. Think about this scenario: A couple adopts a child who has average genetic intellectual potential. They raise her in an extremely stimulating environment. What will happen to the couple's new daughter? It is likely that the stimulating environment will improve her intellectual outcomes over the course of her life. But what happens if this experiment is reversed? If a child with an extremely strong genetic background is placed in an environment that does not stimulate him: What happens? Interestingly, according to a longitudinal study of highly gifted individuals, it was found that "the two extremes of optimal and pathological experience are both represented disproportionately in the backgrounds of creative individuals"; however, those who experienced supportive family environments were more likely to report being happy (Csikszentmihalyi & Csikszentmihalyi, 1993, p. 187).

Another challenge to determining origins of high intelligence is the confounding nature of our human social structures. It is troubling to note that some ethnic groups perform better on IQ tests than others—and it is likely that the results do not have much to do with the quality of each ethnic group's intellect. The same is true for socioeconomic status. Children who live in poverty experience more pervasive, daily stress than children who do not worry about the basic needs of safety, shelter, and food. These worries can negatively affect how the brain functions and develops, causing a dip in IQ scores. Mark Kishiyama and his colleagues determined that children living in poverty demonstrated reduced prefrontal brain functioning comparable to children with damage to the lateral prefrontal cortex (Kishyama, Boyce, Jimenez, Perry, & Knight, 2009).

The debate around the foundations and influences on intelligence exploded in 1969, when an educational psychologist named Arthur Jensen published the article "How Much Can We Boost I.Q. and Achievement" in the Harvard Educational Review. Jensen had administered IQ tests to diverse groups of students, and his results led him to the conclusion that IQ is determined by genetics. He also posited that intelligence was made up of two types of abilities: Level I and Level II. In his theory, Level I is responsible for rote memorization, whereas Level II is responsible for conceptual and analytical abilities. According to his findings, Level I remained consistent among the human race. Level II, however, exhibited differences among ethnic groups (Modgil & Routledge, 1987). Jensen's most controversial conclusion was that Level II intelligence is prevalent among Asians, then Caucasians, then African Americans. Robert Williams was among those who called out racial bias in Jensen's results (Williams, 1970).

Obviously, Jensen's interpretation of his own data caused an intense response in a nation that continued to grapple with the effects of racism (Fox, 2012). However, Jensen's ideas were not solitary or unique; rather, they represented one of many examples of psychologists asserting racial differences in IQ and cognitive ability. In fact, Rushton and Jensen (2005) reviewed three decades worth of research on the relationship between race and cognitive ability. Jensen's belief in the inherited nature of intelligence and the validity of the IQ test to be the truest measure of intelligence are at the core of his conclusions. If, however, you believe that intelligence is more than Levels I and II, or that IQ tests do not control for socioeconomic and cultural differences among people, then perhaps you can dismiss Jensen's conclusions as a single window that looks out on the complicated and varied landscape of human intelligence.

14.3.4 Learning Disabilities

Learning disabilities are cognitive disorders that affect different areas of cognition, particularly language or reading. It should be pointed out that learning disabilities are not the same thing as intellectual disabilities. Learning disabilities are considered specific neurological impairments rather than global intellectual or developmental disabilities. A person with a language disability has difficulty understanding or using spoken language, whereas someone with a reading disability, such as dyslexia, has difficulty processing what he or she is reading.

Often, learning disabilities are not recognized until a child reaches school age. One confounding aspect of learning disabilities is that they often affect children with average to above-average intelligence. At the same time, learning disabilities tend to exhibit comorbidity with other disorders, like attention-deficit hyperactivity disorder (ADHD). Anywhere between 30–70% of individuals

with diagnosed cases of ADHD also have some sort of learning disability (Riccio, Gonzales, & Hynd, 1994). Let's take a look at two examples of common learning disabilities: dysgraphia and dyslexia.

14.3.4.1 Dysgraphia

Children with dysgraphia have a learning disability that results in a struggle to write legibly. The physical task of writing with a pen and paper is extremely challenging for the person. These children often have extreme difficulty putting their thoughts down on paper (Smits-Engelsman & Van Galen, 1997). This difficulty is inconsistent with a person's IQ. That is, based on the child's IQ and/or abilities in other areas, a child with dysgraphia should be able to write, but can't. Children with dysgraphia may also have problems with spatial abilities.

14.3.4.2 Dyslexia

Dyslexia is the most common learning disability in children that causes difficulties in learning to read. An individual with dyslexia exhibits an inability to correctly process letters. The neurological mechanism for sound processing does not work properly in someone with dyslexia. As a result, dyslexic children may not understand sound-letter correspondence. A child with dyslexia may mix up letters within words and sentences—letter reversals, such as those shown in Figure 10.3.2, are a hallmark of this learning disability—or skip whole words while reading. A dyslexic child may have difficulty spelling words correctly while writing. Because of the disordered way that the brain processes letters and sound, learning to read is a frustrating experience. Some dyslexic individuals cope by memorizing the shapes of most words, but they never actually learn to read (Berninger, 2008).

14.3.4.3 The Brain, Dyslexia, and **Dysgraphia**

Developmental dyslexia is a heritable neurodevelopmental disorder that impairs reading in individuals with normal intelligence and educational opportunities. Reading requires the coordination of many brain systems. This "reading circuit" includes language mechanisms in Broca's and Wernicke's areas in the frontal and temporal lobes, respectively, as well as systems involved in vision, working memory, attention, movement and cognition. Processing of printed words takes place within "the visual word form area" located in the fusiform gyrus (also involved in object and face recognition) of the left hemisphere.

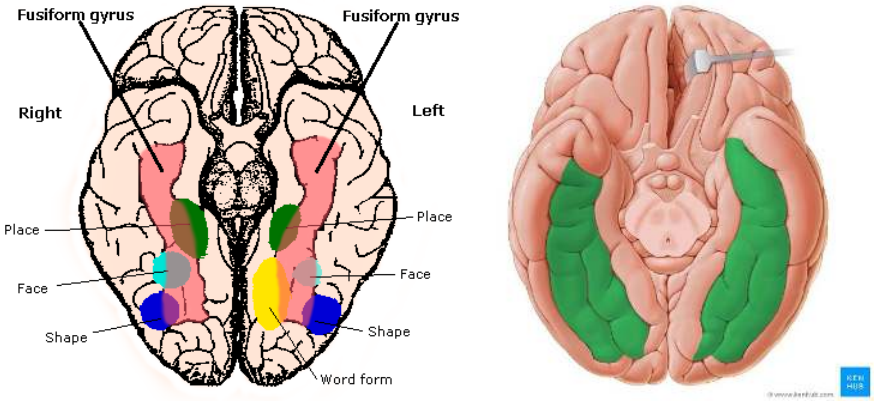


Figure 14.3.2: Fusiform gyrus viewed from bottom of human brain. Note an area for "word form" located in the left fusiform area. (Images from Wikimedia Commons, Fusiform gyrus, retrieved 10/19/21).

Next, a large circuit in the left hemisphere is activated. This circuit includes the **supramarginal gyrus** (mapping of spelling and other writing conventions to language sounds), the **superior temporal gyrus** (phonological sound processing), the **inferior parietal lobule** and the **angular gyrus** (word-meaning processing) and the **inferior frontal gyrus** (phonological sound and semantic meaning processing, working memory). In addition, **subcortical regions** implicated in long-term and working memory, procedural learning and rapid sequential auditory processing (thalamus, basal ganglia and hippocampus) also appear to be involved in reading.

In addition, there is strong evidence for a role of the **visual magnocellular (M) system** in reading and research suggests a significant role of deficits in the visual magnocellular (M) system in **developmental dyslexia** (DD). The M and parvocellular (P) pathways are major parallel visual system streams feeding to lateral geniculate nucleus (visual area) of the thalamus, then onto primary visual (striate) cortex, and then to the extrastriate regions of the occipital lobe (i.e., the 'dorsal' and the 'ventral' visual

streams). The **ventral stream (the "what" pathway)** which involves the inferior temporal cortex is associated with color, form perception, and the discrimination and recognition of objects). The **visual magnocellular or dorsal system (the "where" pathway**; depth and motion perception) contributes to rapid recognition and sequencing of letters by quickly focusing the **ventral attention network** (VAN) on the letter to be identified. Finally, left and right fronto-parietal (attentional) networks critically modulate visual and auditory word pathways by selective attention in both temporal and spatial dimensions. The **visual magnocellular system (the "where" pathway) is related to the fronto-parietal attentional network**. Previous neuroimaging studies have revealed reduced or absent activation within the visual M pathway in DD. Considering the complexity of the circuitry involved in reading, it is likely that a wide range of patterns of neurological deficit in the above mentioned brain areas may contribute to reading difficulties in different individuals (Mascheretti, et al., 2021).

Dysgraphia appears to involve disorder in some **brain areas** which are **not involved in dyslexia**. Richards, et al. (2015) used fMRI to study structural white matter integrity and functional connectivity in children with dysgraphia and dyslexia and found significant differences between control group and both the dysgraphic and dyslexic groups in functional connectivity. Left **occipital temporal gyrus**, supramarginal gyrus, **precuneus**, and **inferior frontal gyrus** were used in these analyses because these brain areas were shown in an analysis of prior research to be related to written word production. Dysgraphia and dyslexia differ in white matter integrity, fMRI functional connectivity, and white matter–gray matter correlations.

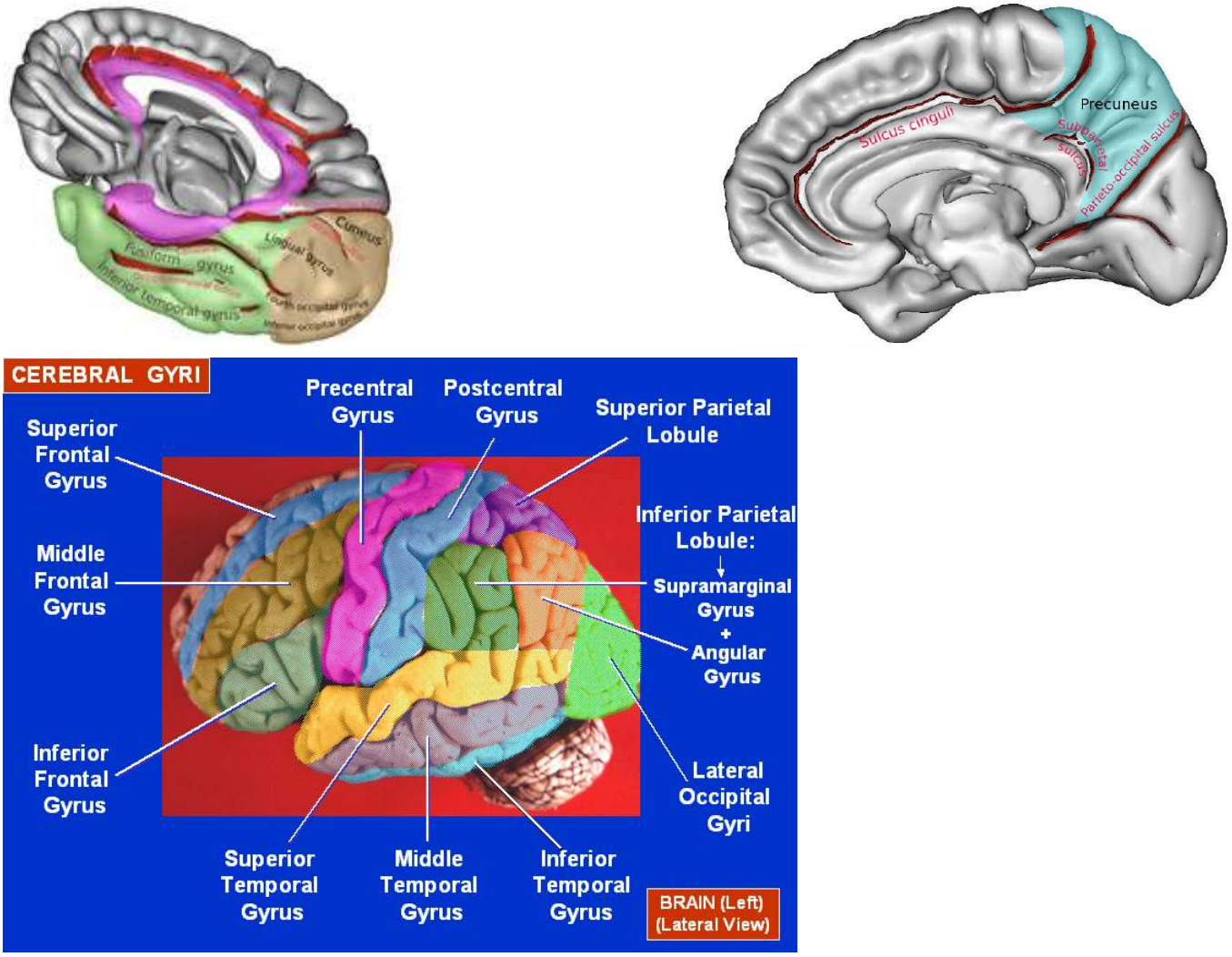


Figure 14.3.3: (top left) Fusiform gyrus (also known as the lateral occipitotemporal gyrus) viewed from bottom of human brain. (top right) Midline view featuring the Precuneus (multiple functions including memory, sensory/perceptual integration, mental imagery strategies) and several other midline sulci. (bottom) Left hemisphere of human brain showing Inferior Frontal gyrus in the Frontal lobe and the Supramarginal gyrus in the Parietal lobe. Decreased white matter integrity and impaired functional connectivity

among these structures is associated with dysgraphia. (Images from Wikipedia Commons, Fusiform gyrus, Precuneus, Inferior Frontal gyrus; retrieved 10/19/21. Caption by Kenneth A. Koenigshofer, PhD, including reference to Science Direct, Precuneus; Richards, et al., 2015).



Figure **14.3.4**: These written words show variations of the word “teapot” as written by individuals with dyslexia. (Image from OER Commons, The Source of Intelligence. Licensed CC BY-NC 4.0).

14.3.5 Summary

Genetics and environment affect intelligence and the challenges of certain learning disabilities. The intelligence levels of all individuals seem to benefit from rich stimulation in their early environments. Highly intelligent individuals, however, may have a built-in resiliency that allows them to overcome difficult obstacles in their upbringing. Learning disabilities can cause major challenges for children who are learning to read and write. Unlike developmental disabilities, learning disabilities are strictly neurological in nature and are not related to intelligence levels. Students with dyslexia, for example, may have extreme difficulty learning to read, but their intelligence levels are typically average or above average.

14.3.6 Review Questions

1. Where does high intelligence come from?
    1. genetics
    2. environment
    3. both A and B
    4. neither A nor B
2. Arthur Jensen believed that _____.
    1. genetics was solely responsible for intelligence
    2. environment was solely responsible for intelligence
    3. intelligence level was determined by race
    4. IQ tests do not take socioeconomic status into account
3. What is a learning disability?
    1. a developmental disorder
    2. a neurological disorder
    3. an emotional disorder
    4. an intellectual disorder
4. Which of the following statements is true?
    1. Poverty always affects whether individuals are able to reach their full intellectual potential.
    2. An individual’s intelligence is determined solely by the intelligence levels of his siblings.
    3. The environment in which an individual is raised is the strongest predictor of her future intelligence
    4. There are many factors working together to influence an individual’s intelligence level.

14.3.7 Critical Thinking Questions

What evidence exists for a genetic component to an individual's IQ?

Describe the relationship between learning disabilities and intellectual disabilities to intelligence.

14.3.8 Personal Application Question

Do you believe your level of intelligence was improved because of the stimuli in your childhood environment? Why or why not?

14.3.9 References

Bartels, M., Rietveld, M., Van Baal, G., & Boomsma, D. I. (2002). Genetic and environmental influences on the development of intelligence. *Behavior Genetics*, *32* (4), 237–238.

Berninger, V. W. (2008). Defining and differentiating dysgraphia, dyslexia, and language learning disability within a working memory model. In M. Mody & E. R. Silliman (Eds.), *Brain, behavior, and learning in language and reading disorders* (pp. 103–134). New York: The Guilford Press.

Bouchard, T. J., Lykken, D. T., McGue, M., Segal, N. L., & Tellegen, A. (1990). Sources of human psychological differences: The Minnesota Study of Twins Reared Apart. *Science*, *250*, 223–228.

Csikszentmihalyi, M., & Csikszentmihalyi, I. (1993). Family influences on the development of giftedness. *Ciba Foundation Symposium*, *178*, 187–206.

Fox, M. (2012, November 1). Arthur R. Jensen dies at 89; Set off debate about I.Q. *New York Times*, p. B15.

Jensen, A. R. (1967). How much can we boost IQ and scholastic achievement?. Harvard Educational Review.

Kishyama, M. M., Boyce, W. T., Jimenez, A. M., Perry, L. M., & Knight, R. T. (2009). Socioeconomic disparities affect prefrontal function in children. *Journal of Cognitive Neuroscience*, *21*(6), 1106–1115.

Mascheretti, S., Peruzzo, D., Andreola, C., Villa, M., Ciceri, T., Trezzi, V., ... & Arrigoni, F. (2021). Selecting the Most Relevant Brain Regions to Classify Children with Developmental Dyslexia and Typical Readers by Using Complex Magnocellular Stimuli and Multiple Kernel Learning. *Brain Sciences*, *11*(6), 722.

Modgil, S., & Routledge, C. M. (Eds.). (1987). *Arthur Jensen: Consensus and controversy.* New York: Falmer Press.

Riccio, C. A., Gonzales, J. J., & Hynd, G. W. (1994). Attention-deficit Hyperactivity Disorder (ADHD) and learning disabilities. *Learning Disability Quarterly, 17*, 311–322.

Richards, T. L., Grabowski, T. J., Boord, P., Yagle, K., Askren, M., Mestre, Z., ... & Berninger, V. (2015). Contrasting brain patterns of writing-related DTI parameters, fMRI connectivity, and DTI–fMRI connectivity correlations in children with and without dysgraphia or dyslexia. *NeuroImage: Clinical*, *8*, 408-421.

Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, public policy, and law, 11* (2), 235–294.

Smits-Engelsman, B. C. M., & Van Galen, G. P. (1997). Dysgraphia in children: Lasting psychomotor deficiency or transient developmental delay? *Journal of Experimental Child Psychology, 67*, 164–184.

Williams, R. L., (1970). Danger: Testing and dehumanizing black children. *Clinical Child Psychology Newsletter, 9* (1), 5–6.

14.3.10 Attributions

Adapted by Kenneth A. Koenigshofer, PhD., from The Source of Intelligence by Rice University, Licensed CC BY-NC 4.0 via OER Commons

---

This page titled 14.3: High Intelligence-Nature or Nurture; Learning Disabilities is shared under a mixed license and was authored, remixed, and/or curated by Kenenth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

14.4: Cognition- Categories, Concepts, Schemas and the Brain

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe cognition.
2. Discuss concepts and prototypes.
3. Explain schemas and how they contribute to the adaptive organization of behavior and its efficiency.
4. Discuss the brain structures involved in category formation.
5. Describe the types of characteristics of things which affect where in the brain their categories are stored.
6. Describe the roles of similarity, prototypes, and typicality in categorization.
7. Explain prosopagnosia.
8. Describe the fuzzy nature of concepts.
9. Explain category hierarchies.
10. Discuss the representation of concepts and knowledge.

14.4.1 Overview

Cognitive psychology is dedicated to examining how people think, including interactions among human thinking, emotion, creativity, language, and problem solving, and how we organize thoughts and information gathered from our environments into meaningful categories. As discussed earlier, the generation of categories, based on similarity, is an example of an ability that has arisen from the genetic internalization, by natural selection, of an enduring fact of the world. That fact is that things in the world are similar to other things, in various properties, and in varying degrees of abstraction (see Section 14.2). A brain unable to form categories would be crippled; every experience would seem unrelated to every other, and the brain's ability to find order in the world would not exist. Much of thinking involves the formation and use of categories. Inferences about the properties of new instances of a category based on knowledge about the category is an essential component of intelligence and thinking in humans and in a number of other animal species. The high level of abstraction that the human brain is capable of provides humans with categories of very high degrees of abstraction, giving great cognitive power to our species compared to other animals. Human concepts can range from classification based on simple concrete properties such as shape or color to high order abstract properties leading to concepts such as mammal, illegal, commerce, electrostatic force, or beauty. **Corticostriatal loops** involving connections between a number of cortical areas and the striatum, composed of several nuclei of the basal ganglia, appear to be crucially involved in category formation in humans.

14.4.2 Concepts and Prototypes

The senses serve as the interface between the mind and the external environment, receiving stimuli and translating it (transducing it via sensory receptors) into nervous impulses that are transmitted to the brain. The brain then processes this information and uses the relevant pieces to create thoughts, which can then be expressed through language or stored in memory for future use. When thoughts are formed, the brain also pulls information from emotions and memories, powerful influences on both our thoughts and behaviors.

Concepts are *abstract representations* or *cognitive structures* formed from classes or *groupings of things*, events, or relations *based on common properties*. Concepts can be about concrete things such as the concepts "car," "bird," or "swimming," or about complex and abstract things, like "justice" or "success". In psychology, for example, Piaget’s stages of development are abstract concepts. Some concepts, like tolerance, are agreed upon by many people, because they have been used in various ways over many years. Other concepts, like the characteristics of your ideal friend or your family’s birthday traditions, are personal and individualized. In this way, concepts touch every aspect of our lives, from our many daily routines to the guiding principles behind the way governments function.

A **prototype** is the *best example* or representation *of a concept*. For example, for the category of civil disobedience, your prototype could be Rosa Parks. Her peaceful resistance to segregation on a city bus in Montgomery, Alabama, is a recognizable example of civil disobedience. Or your prototype could be Mohandas Gandhi, sometimes called Mahatma Gandhi. Mohandas Gandhi served as a nonviolent force for independence for India. Prototypes apply to more concrete concepts as well. In your mind, is the prototype bird a penguin, an eagle, a sparrow, or is there some other type of bird that is the best example of the concept, bird? Which of the birds listed in the previous sentence is *most typical of the category* "birds"? Is "typical" simply a function of frequency of occurrence?

14.4.3 Schemas

A **schema** is a mental construct consisting of a cluster or collection of related concepts (Bartlett, 1932). In other words, a schema is a collection of knowledge and beliefs about some entity or situation that directs behavior and guides expectations. Schemas help organize knowledge and often help us predict event sequences and attributes of things we encounter in the world. For example, the schema "library" suggests the presence of books, desks, shelves, and a quiet place to study. It also suggests a sequence of actions including searching the stacks, selecting a book, and taking the book to a librarian at a check out desk before leaving the library with the book. There are many different types of schemas, and they all have one thing in common: schemas are a method of organizing information that allows the brain to work more efficiently. When a schema is activated, the brain makes immediate assumptions about the person or object being observed and using this *information contained in the schema* to organize behavior.

There are several types of schemas. A **role schema** makes assumptions about how individuals in certain roles will behave (Callero, 1994). For example, imagine you meet someone who introduces himself as a firefighter. When this happens, your brain automatically activates the “firefighter schema” and begins making assumptions and generating *expectations (predictions)* that this person is brave, selfless, and community-oriented. Despite not knowing this person, already you have unknowingly made judgments and formed expectations about him. Notice how the *schema is predictive* in the sense that it allows you to form expectations about something in the future, in this case, what behaviors this person might be expected to engage in. A common feature of many forms of cognition, including schemas, is *projections or expectations about probable events in future time.*

Schemas also help you fill in gaps in the information you receive from the world around you. While schemas allow for more efficient information processing, there can be problems with schemas, regardless of whether they are accurate most of the time: Perhaps this particular firefighter is not brave, he just works as a firefighter to pay the bills while studying to become a children’s librarian. *Schemas involve generalization--inference based on* prior experience with *similar* things in the past. Like schemas, all forms of *generalization* permit *prediction* allowing us to fill in gaps in our direct knowledge including what might be expected to occur in the future. Although there is always the chance that our predictions or inferences based on generalization might be wrong, nevertheless, *schemas and generalizations from them are powerful forms of cognition*. They permit us to form *expectations from incomplete information* about the future and thus allow us to prepare and plan for what is to come next. This *anticipatory property of cognition* is *highly adaptive* and likely has been powerfully *selected* for during the course of *brain evolution*.

An **event schema**, also known as a **cognitive script**, is a knowledge structure about a sequence of events. An event schema can lead to a set of behaviors that can feel like a routine. Think about what you do when you walk into an elevator. First, the doors open and you wait to let exiting passengers leave the elevator car. Then, you step into the elevator and turn around to face the doors, looking for the correct button to push. Like all *schemas*, event schemas are learned from *environmental regularities* we experience in the world.

Event schemas can vary widely among different cultures and countries. For example, while it is quite common for people to greet one another with a handshake in the United States, in Tibet, you greet someone by sticking your tongue out at them, and in Belize, you bump fists (Cairns Regional Council, n.d.) Because event schemas are automatic, they can be difficult to change. Imagine that you are driving home from work or school. This event schema involves getting in the car, shutting the door, and buckling your seatbelt, and putting the key in the ignition, and driving a particular route. How many times have you driven your route home only to remember as you pass your turn that you were intending to stop at the store first?

14.4.4 Concepts, Prototypes, Schemas, and Evolution of General Intelligence

Concepts, prototypes, and schemas all rely upon abstract, higher order features of the world which the brain captures and utilizes when it forms these *knowledge structures*. For example, concepts and prototypes are knowledge structures which *capture*

similarities among individual instances of things (e.g. all birds have beaks, feathers, and wings). As discussed in an earlier section, the *brain* appears to be *innately organized* to find similarities and to *generate higher order representations based on similarity*. This leads to formation of concepts, categories, and predictions or expectations based on partial information. Also, as previously discussed, the brain has evolved to readily recognize and represent ***cause-effect*** and the ***predictive relations*** *(correlation or covariation) of environmental events* (Koenigshofer, 2017), leading to the formation of event schemas, as well as knowledge about the predictive and causal relations among events in the world--*all* of these help generate *sophisticated* ***knowledge*** *and understanding of one's environmen*t, facilitating adaptive organization of behavior and *increasing biological fitness*.

Perhaps one of the strongest examples of human *thinking and intelligence* is *scientific discovery*, a process which demonstrates abilities for *categorization*, often of a *highly abstract* nature, and *causal understanding*, talents which are evident in humans at an early age. As one researcher and his colleagues state,

> "Causal induction, i.e., identifying unobservable mechanisms that lead to the observable relations among variables, has played a pivotal role in modern scientific discovery, especially in scenarios with only sparse and limited data. Humans, even young toddlers, can induce causal relationships surprisingly well in various settings despite its notorious difficulty" (Zhang, et al., 2021, p. 1).

Combined, these innate, evolved properties of the brain help the brain develop understanding of the relations among things in the world, thereby guiding adjustments in behavior for successful adaptation to the environment. Recall from earlier sections that these abilities comprise central components of **general intelligence** (the recognition "of relations and correlates," according to Spearman, 1904, 1925), found not only in humans, but in many animals ranging from crows and ravens to the great apes (see Koenigshofer, 2017). Although it is likely that the abilities that underlie general intelligence are distributed widely in the brain, the **frontal and parietal lobes** of the cerebral cortex appear to be especially important in *human general intelligence* (Bruner, 2010). Also recall that one *additional component of general intelligence* is the *ability to imagine possible future behaviors and their probable outcomes* in visual-like mental images, an ability which may involve **visual and motor cortex and parts of the parietal lobe** involving, as discussed in an earlier module, a **frontoparietal network**. This ability to imagine, as already noted, may play an important role in *response selection and planning* for the future, *key features of human cognition and intelligence* (Koenigshofer, 2017).

14.4.4.1 Brain Mechanisms in Category Formation

How the brain forms categories, and the brain areas involved, is not yet settled. However, evidence from studies of brain damage and brain imaging sheds light on the issue.

One thing is certain. Category formation is essential to cognitive processes. According to Seger and Miller (2010), the *ability to form groupings* of things and events into categories is a *fundamental property of "sophisticated thought."* Intelligence depends upon the ability to form meaningful categories. Disruption of this key feature of thinking and intelligence leads to behavioral and cognitive pathology. The importance of category formation is suggested by Seger and Miller (2010). Without the ability to form categories, "the world would lack any deeper meaning. Experiences would be fragmented and unrelated. Things would seem strange and unfamiliar if they differed even trivially from previous examples. This situation describes many of the cognitive characteristics of neuropsychiatric disorders such as autism" (Seger & Miller, 2010, p. 203).

Category formation, as discussed in earlier sections, reflects a fundamental fact of the world--that things and events in the world are similar to other things and events. Ability of the brain to exploit this fundamental property of the world to allow inference from the properties of a known category to new instances of the category has enormous, favorable consequences for survival and reproduction. A general principle is operating here. Natural selection has organized brain systems to reflect biologically significant regularities of the world (Koenigshofer, 2017; Shepard, 2001). This is consistent with the view expressed earlier that *intelligence and thinking* evolved as sophisticated *guidance systems* which produce neural models or *cognitive maps* of the world for the production of adaptively successful behavior.

The ability to form categories is dependent upon many brain areas which interact during the learning of categories. According to Seger and Miller (2010), these areas include the **visual cortex**, the **prefrontal cortex**, the **parietal cortex**, the **basal ganglia**, and the **medial temporal lobe**, including "interactions within and between **corticostriatal loops** connecting cortex and basal ganglia and between the basal ganglia and the medial temporal lobe." This provides "a balance between acquisition of details of experiences and generalization across them" to form and use categories (Seger & Miller, 2010, p. 203). According to Seger and

Miller (2010), the **inferotemporal (IT) cortex** is likely a participating brain area in visual categorization given that it contains the **fusiform face area (FFA)**, rich in **face cells**, active during learning of new face categories. Furthermore, **IT cortical neurons** in trained monkeys fire selectively to trees or fish with relatively little variation of firing within categories, suggesting that these neurons within the IT cortex **encode specific categories of stimuli**.

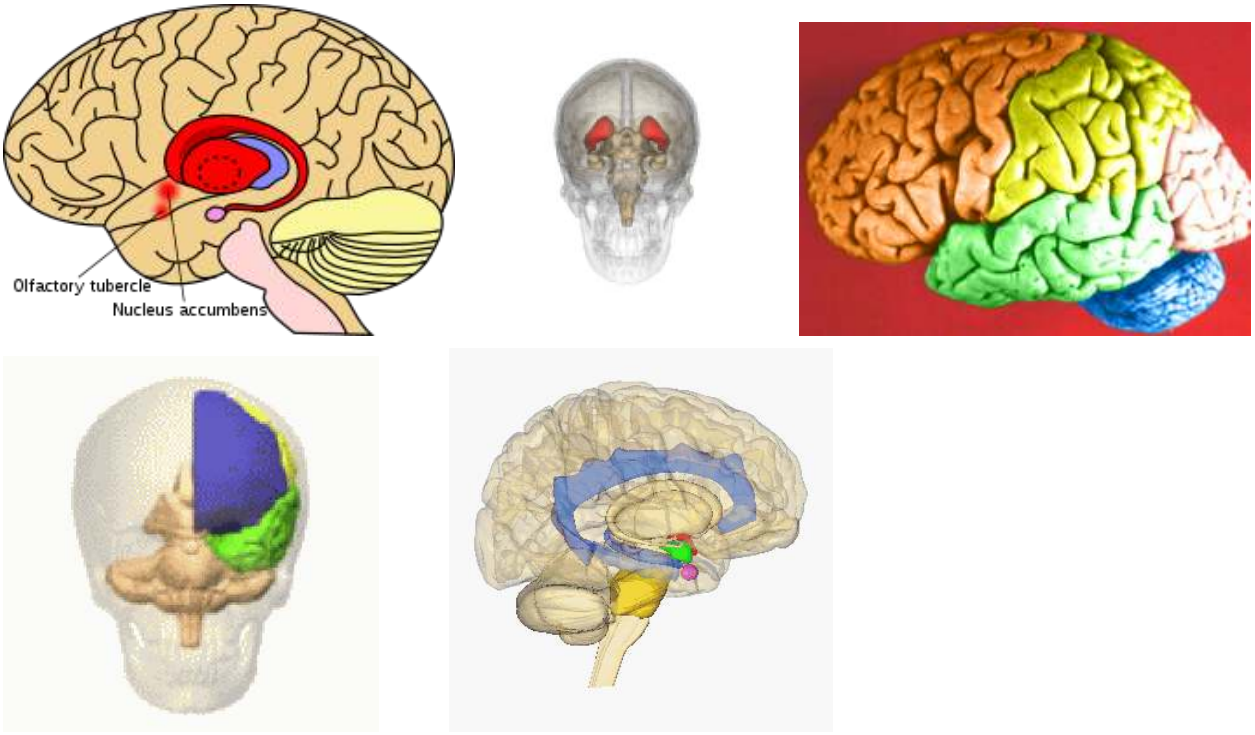


Figure **14.4.1**: (Top Left and center) Striatum (shown in red) is a main input area of the basal ganglia, which receives input primarily from the cerebral cortex. (Right top and bottom left) Cerebral cortex: temporal lobe in green, parietal lobe in yellow, frontal lobe in brown (stationary photo) and in blue (bottom left rotating figure), occipital lobe in pink (stationary photo) and in rust (bottom left rotating figure). (Bottom right rotating figure) Human brain (hypothalamus=red, amygdala=green, hippocampus/fornix=blue, pons=gold, pituitary gland=pink).

(Images from Wikimedia, (top left) Striatum; File:Striatum.svg; https://commons.wikimedia.org/wiki/File:Striatum.svg; licensed under the Creative Commons Attribution-Share Alike 4.0 International license. Rotating (top center), File:Striatum.gif; https://commons.wikimedia.org/wiki/File:Striatum.gif; by Life Science Databases(LSDB); licensed under the Creative Commons Attribution-Share Alike 2.1 Japan license. Cortical lobes. (Top right) File:Brain - Lobes.png; https://commons.wikimedia.org/wiki/F...in_-_Lobes.png; by John A Beal, PhD, Dep't. of Cellular Biology & Anatomy, Louisiana State University Health Sciences Center Shreveport; Modifications: *Hemispheres in color* by DavoO; licensed under the Creative Commons Attribution 2.5 Generic license. (Bottom left, rotating) File:Four lobes animation small.gif; https://commons.wikimedia.org/wiki/F...tion_small.gif; by Database Center for Life Science(DBCLS); licensed under the Creative Commons Attribution-Share Alike 2.1 Japan license. Retrieved 10/25/21. (Bottom right, rotating: File:Rotating brain colored.gif; https://commons.wikimedia.org/wiki/F...in_colored.gif; by lifesciencedb; licensed under the Creative Commons Attribution-Share Alike 2.1 Japan license.).

Mahon and Caramazza (2009) reviewed research on categorization involving brain imaging and brain damage. They note that one generalization from this research is that "object domain and sensory modality jointly constrain the organization of knowledge in the brain." In other words, **localization of particular items of knowledge in the brain depends on the object category (e.g. faces, tools) and also upon which sensory system (e.g. visual, auditory) is involved in representing the knowledge**. Specifically, studies of the effects of brain damage on verbal categorization have been interpreted by some researchers to indicate that broad categories of objects (object domains) may be represented separately in different cortical regions. Research with brain damaged patients with damage in different brain areas has revealed "disproportionate or even selective impairments" for one category

compared to other categories. This supports the view that *different categories of objects are represented in different areas of brain.* Cases of verbal category impairments have been found for the categories "animals," "fruit/vegetables," "conspecifics" (other humans), and "non-living things." For many patients, the deficits included failure to understand knowledge about the concepts, not just in naming them. For example, patients with impairment of the category "animals" could not answer simple questions about the features of specific animals, such as "Does a whale have legs" (Mahon and Caramazza, 2009, p. 28). Another patient had deficits in conceptual knowledge about people as evidenced by a severe inability to name famous people, even though this patient did not have **prosopagnosia** (inability to recognize familiar faces, such as family members or even one's own face in a photograph, while still being able to recognize a familiar person by other sensory modalities such as by voice).

According to Mahon and Caramazza, in addition to the theory that different categories may be localized in different modality-specific regions of brain (e.g. visual areas, somatosensory areas), other theories of categorical knowledge have been proposed, including the idea that **category formation is "constrained by evolutionarily important distinctions such as animate, inanimate, conspecifics, and tools," or** that **categories are based on "statistical regularities in the co-occurrence of object properties in the world,"** implying wide distribution of neural representations of specific categories in the brain.

According to Mahon and Caramazza (2009), Damasio et al. (1996) found that inability to name pictures of famous people was related to "**left temporal pole lesions**," while impairment for naming animals occurred "with (more posterior) lesions of **anterior left ventral temporal cortex**." Additional studies confirmed that impairments naming animals occur with lesions of **anterior temporal cortex**. Studies by Damasio and colleagues and others found deficits for recognizing and naming tools with lesions to the **posterior and lateral temporal cortex,** overlapping the left posterior middle temporal gyrus. fMRI studies reveal that "nonliving things, and in particular tools, differentially activate the **left middle temporal gyrus**" as does mechanical motion. "Living **animate things** such as faces and animals elicit differential neural responses in the **lateral fusiform gyrus**, whereas **nonliving things** (tools, vehicles) elicit differential neural responses in the **medial fusiform gyrus**." Interestingly, brain areas involved in emotional processing and *theory of mind* (attribution of mental states in others) are part of the neural network activated during processing of information about living animate things (Mahon & Caramazza, 2009).

Ishibashi et al. (2016) reviewed "neuroimaging studies . . . to identify tool-related cortical circuits dedicated either to general tool knowledge or to task-specific processes. The results indicate the following: (a) Common, task-general processing regions for tools are located in the **left inferior parietal lobule** (LIPL) and **ventral premotor cortex**; and (b) task-specific regions are located in **superior parietal lobule** (SPL) and **dorsal premotor area** for imagining/executing actions with tools and in **bilateral occipito-temporal cortex** for recognizing/naming tools."

An approach to cognition known as **embodied cognition** hypothesizes that abstract concepts necessarily involve experience with previous sensorimotor interactions involving the whole body and thus may include encoding by sensory and motor areas of the brain. For example, we speak of an idea that we are not grasping as being "over our head," and we speak of affection in terms of warmth as represented in the fact that affection often is expressed in ways that allow us to feel the physical warmth of the body of another for whom we have affection. The concept of justice is often represented by a scale in balance. So, on this view, abstract concepts are often expressed in terms that are analogous to sensory and motor experiences of a body embedded in a concrete physical world.

14.4.5 Summary

In this section, you were introduced to topics within cognitive psychology, which is the study of cognition, or the brain’s ability to think, perceive, plan, analyze, and remember. Concepts and their corresponding prototypes help us quickly organize our thinking by creating categories into which we can sort new information. We also develop schemas, which are clusters of related concepts. Some schemas involve routines of thought and behavior, and these help us function in various situations without having to “think twice” about them. Schemas show up in social situations and routines of daily behavior. Concepts, prototypes, and schemas arise from fundamental dispositions of the brain to form knowledge structures based on similarity, causal relations, and correlations or predictive covariations of events. Natural selection favored brain organization capable of exploiting these abstract properties of the world leading to the evolution of general intelligence. General intelligence equipped humans and many other species with capacity to solve a wide range of adaptive challenges. In humans especially, this involves reasoning, planning, and imagination of possible future actions and their probable outcomes (Koenigshofer, 2017).

14.4.6 References

Bruner, E. (2010). Morphological differences in the parietal lobes within the human genus. *Current Anthropology*, 51(S1), S77-S88.

Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D., & Damasio, A. R. (1996). A neural basis for lexical retrieval. *Nature*, *380* (6574), 499-505.

Koenigshofer, K. A. (2017). General Intelligence: Adaptation to Evolutionarily Familiar Abstract Relational Invariants, Not to Environmental or Evolutionary Novelty. *Journal of Mind & Behavior, 38* (2).

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: a cognitive neuropsychological perspective. *Annual review of psychology*, *60*, 27-51.

Ishibashi, R., Pobric, G., Saito, S., & Lambon Ralph, M. A. (2016). The neural network for tool-related cognition: an activation likelihood estimation meta-analysis of 70 neuroimaging contrasts. *Cognitive Neuropsychology*, *33*(3-4), 241-256.

Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annual review of neuroscience*, *33*, 203-219.

Shepard, R. N. (2001). Perceptual-cognitive universals as reflections of the world. *Behavioral and brain sciences*, *24* (04), 581-601.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *American Journal of Psychology, 15* (2), 201-292.

Spearman, C. (1925). Some issues in the theory of`'g' (including the law of diminishing returns). *Nature*, *116*, 436-439.

Zhang, C., Jia, B., Edmonds, M., Zhu, S. C., & Zhu, Y. (2021). ACRE: Abstract Causal REasoning Beyond Covariation. arXiv preprint arXiv:2103.14232. https://arxiv.org/pdf/2103.14232.pdf retrieved April 2, 2021.

14.4.7 Attributions

- Adapted by Kenneth A. Koenigshofer, PhD., from What is Cognition? by OpenStax Colleg licensed CC BY-NC 4.0 via OER Commons
- "Concepts, Prototypes, Schemata, and Evolution of General Intelligence" and "Brain Mechanisms of Category Formation" is original work written by Kenneth Koenigshofer, PhD, licensed under CC BY-NC 4.0

This page titled 14.4: Cognition- Categories, Concepts, Schemas and the Brain is shared under a mixed license and was authored, remixed, and/or curated by Kenenth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

14.5: Brain Mechanisms and Intelligence

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe the brain structures included in the P-FIT model of the biology of intelligence.
2. Describe the Default-Mode Network (DMN) and its relationship to the biology of intelligence.
3. Discuss the type of attention that is probably involved when someone is taking an I.Q. test and how this is related to interpretation of brain imaging studies of the brain networks involved in human intelligence.

14.5.0.1 Overview

According to Hearne, et al. (2016, p.1), "Human intelligence can be broadly defined as the capacity to understand complex ideas, adapt effectively to the environment and engage in complex reasoning[1]. Measures of intelligence can be related to performance on virtually any cognitive task, from sensory discrimination[2] to challenging cognitive tasks such as the identification of patterns in the Raven's Progressive Matrices test[3]. Importantly, scores on intelligence tests can accurately predict various life outcomes, including academic success[4], job performance[5] and adult morbidity and mortality[6]." In spite of the adaptive importance of intelligence in human life, the brain mechanisms involved in intelligence are not well understood. However, research makes it increasingly clear that complex functions such as cognition and intelligence involve many brain areas acting together in interconnected neural networks comprised of multiple brain areas. Furthermore, interactions between different networks of brain areas are also important if we are to understand how the brain creates cognition and intelligence. Functional magnetic resonance imaging (fMRI) has been used to investigate the relationship between individual differences in intelligence and brain activity during cognitive activities such as working memory and reasoning. As discussed in prior modules of this chapter, brain imaging studies implicate neural interactions in a **fronto-parietal network** underlying many of the functions associated with human intelligence. These observations have resulted in the influential **Parieto-Frontal Integration Theory** of intelligence (P-FIT) which identifies a number of interacting brain structures associated with individual differences in human intelligence including frontal and parietal lobes and other structures. Additional research (Hearne, et al., 2016) shows that individual differences in intelligence between people are also associated with the degree of neural interaction between the fronto-parietal network and the **default-mode network**. Greater connectivity between these two neural networks when the brain is at rest (i.e. in the absence of any specific cognitive task) is correlated with higher intelligence scores in individuals compared to those with lesser connectivity between the fronto-parietal and default mode networks. Interactions with a **dorsal attention network** (DAN) must also be considered for a more complete understanding of the complex network interactions involved in cognition and intelligence.

14.5.1 Brain Mechanisms and Intelligence

Because of the complexity and number of interacting factors in human intelligence, understanding the brain mechanisms in human intelligence is a monumental task for researchers. The figure below showing Carroll's influential model of human intelligence is a reminder of the large number of interacting abilities involved. The various factors and abilities shown are derived mathematically using *factor analysis* based on analysis of patterns of correlations in the performances of large numbers of people on measures of intelligence such as standard I.Q. tests (e.g. Wechsler Adult Intelligence Scale, WAIS).

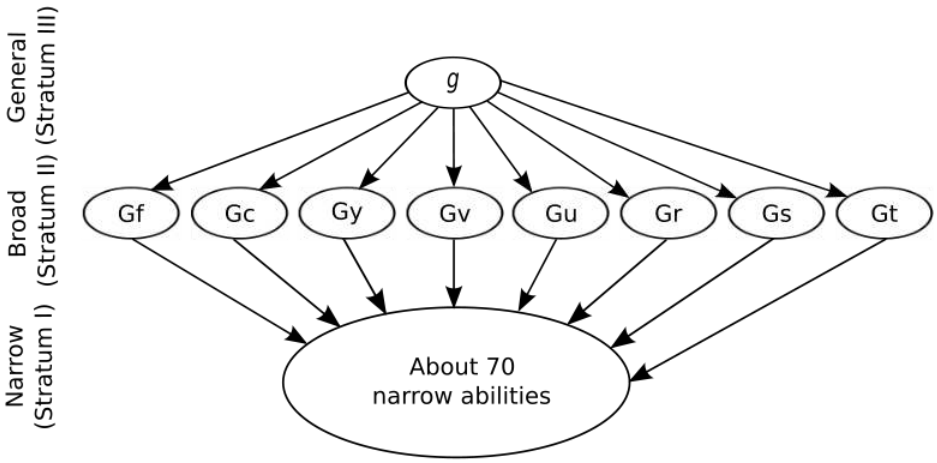


Figure 14.5.1: At the top of Carroll's three-stratum model of human intelligence is the g-factor, general intelligence. Notice the large number of Stratum I abilities which comprise the factors in g and the more specific g-factors in Stratum II. Abbreviations: fluid intelligence (Gf), crystallized intelligence (Gc), general memory and learning (Gy), broad visual perception (Gv), broad auditory perception (Gu), broad retrieval ability (Gr), broad cognitive speediness (Gs), and processing speed (Gt). (Image and abbreviations from Wikimedia Common; File:Carroll three stratum.svg; https://commons.wikimedia.org/wiki/F...ee_stratum.svg; by Victor Chmara; made available under the Creative Commons CC0 1.0 Universal Public Domain Dedication).

In spite of this complexity, neuroscientists using brain imaging methods have identified brain regions associated with differences in measured I.Q. or other measures of general intelligence. As mentioned earlier in this chapter, Jung and Haier (2007) reviewed brain imaging studies and identified a set of brain areas correlated with individual differences in intelligence and reasoning among people. They called this set of interconnected brain regions the **fronto-parietal network** and dubbed their model the **Parietal-Frontal Integration Theory (P-FIT)**. "The P-FIT model includes, by Brodmann areas (BAs): the **dorsolateral prefrontal cortex** (BAs 6, 9, 10, 45, 46, 47), the **inferior** (BAs 39, 40) **and superior** (BA 7) **parietal lobule**, the **anterior cingulate** (BA 32), and **regions within** the **temporal** (BAs 21, 37) and **occipital** (BAs 18, 19) **lobes**. White matter regions (i.e., **arcuate fasciculus**) are also implicated" (Jung and Haier, 2007, p. 135).

The arcuate fasciculus interconnects regions of temporal and parietal cortex with frontal cortex (see Figure 14.8.2). This bundle of axons connects Wernicke's and Broca's areas, in temporal and frontal lobes, respectively, which are involved in language comprehension and language production (see modules on language later in this chapter). Surprisingly, this does not mean that language ability is necessary for cognition and intelligence.

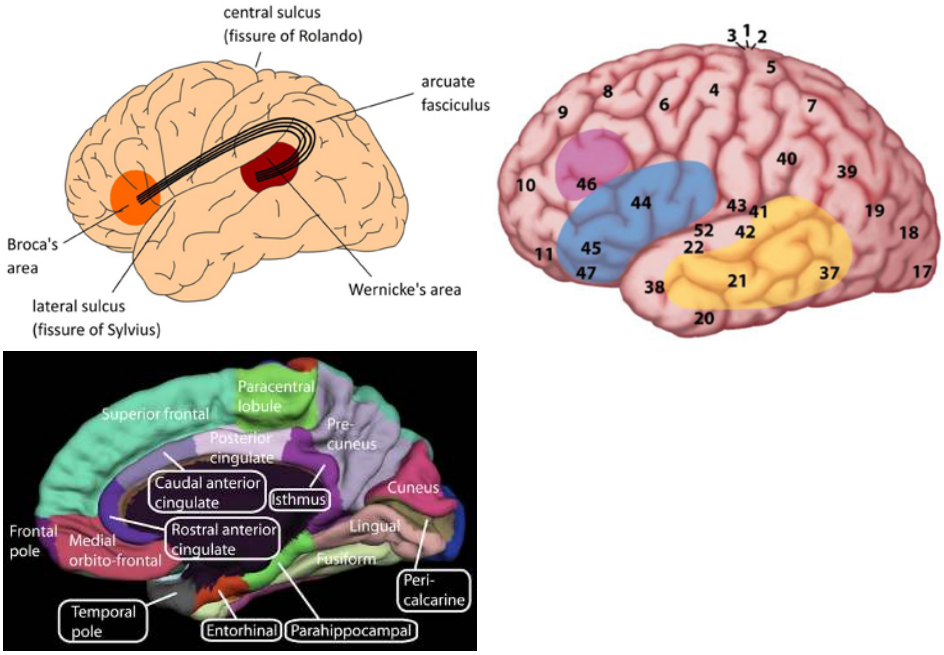


Figure **14.5.2**: (Left): The Arcuate Fasciculus. Lateral view of left hemisphere. (Center): The center figure also displays a lateral view of the left hemisphere. The numbers indicate Brodmann areas (BA). These are areas with differences in the cytoarchitectonics (i.e., composition of cell types). The memory areas are in the temporal cortex (in yellow) including the angular gyrus in parietal cortex. Broca's area (Brodmann areas 44 and 45) and adjacent cortex (Brodmann areas 47 and 6) in the frontal lobe involved in language. Control operations recruit another part of the frontal lobe (in pink), and the Anterior Cingulate Cortex (ACC; not shown in the center figure), as well as areas involved in attention. (Right): Medial view of cerebral cortex showing major gyri. (Images from Wikimedia Commons; Left image: File:The classical Wernicke-Lichtheim-Geschwind model of the neurobiology of language fpsyg-04-00416-g001.jpg; https://commons.wikimedia.org/wiki/F...00416-g001.jpg; by Peter Hagoort; licensed under the Creative Commons Attribution 3.0 Unported license. Center image and caption: File:The MUC (Memory, Unification, Control) model of language fpsyg-04-00416-g002.jpg; https://commons.wikimedia.org/wiki/F...00416-g002.jpg; by Peter Hagoort; Hagoort P (2013) MUC (Memory, Unification, Control) and beyond. Front. Psychol. 4:416. doi: 10.3389/fpsyg.2013.00416 http://journal.frontiersin.org/article/10.3389/fpsyg.2013.00416/full); licensed under the Creative Commons Attribution 3.0 Unported license. Right image: File:Medial surface of cerebral cortex - gyri.png; https://commons.wikimedia.org/wiki/F...tex_-_gyri.png; by Patric Hagmann et.al., Hagmann P, Cammoun L, Gigandet X, Meuli R, Honey CJ, et al. (2008) Mapping the Structural Core of Human Cerebral Cortex. PLoS Biol 6(7): e159. doi:10.1371/journal.pbio.0060159[1]; licensed under the Creative Commons Attribution 2.5 Generic license).

For example, "patients with even severe damage to the language network can retain high intelligence and the ability to perform challenging cognitive tasks, like arithmetic (Varley et al., 2005) and causal reasoning (e.g., Varley and Siegal, 2000; see Fedorenko and Varley, 2016, for a review)" (Assem, et al., 2020, p. 139). This finding strongly suggests that **cognitive tasks like arithmetic and causal reasoning do not depend upon language or language networks.** This finding is consistent with the theory that causal reasoning is a component of general intelligence evolved by "genetic internalization" of causality, similarity, and predictive relations into functional organization of the brain, and that general intelligence is found in many non-human animal species even though they lack language (Koenigshofer, 2017).

Using fMRI, Duncan (2010, 2013) has identified a frontal and parietal pattern of neural activity associated with diverse cognitive tasks and general fluid intelligence. Duncan calls this pattern "multi-demand" (MD) because this activation pattern is a "salient part of the brain's response to many different kinds of cognitive challenge. . . . [T]his multiple-demand (MD) pattern extends over a specific set of regions in **prefrontal and parietal cortex**, in particular: cortex in and around the **posterior** part of the **inferior frontal sulcus** (IFS), [cortex' in the **anterior insula** and the adjacent **frontal operculum** (AI/FO), in the **pre-supplementary motor area** and adjacent **dorsal anterior cingulate** (preSMA/ACC), and [cortex] in and around the **intraparietal sulcus** (IPS). A smaller region of accompanying activity is **sometimes** seen in **rostrolateral prefrontal cortex** (RPFC)" (Duncan, 2010, p. 172). A

similar pattern of activation is seen in tasks associated with **general fluid intelligence**. Furthermore, "recent lesion data suggest that deficits in fluid intelligence are specifically associated with damage to the MD network (Duncan, 2010, p. 172).

Using fMRI, Gray, et al. (2003) found that individual differences in general fluid intelligence (gF) were associated with differences in activation of attentional mechanisms in the **lateral prefrontal and parietal cortex** during a working memory task. These results are consistent with the P-FIT model of intelligence and suggest that at least some of the differences in measured intelligence among people may be due to differences in brain networks that mediate attention.

According to Assem, et al. (2020, p. 131), "A distributed frontoparietal Multiple Demand (MD) network [Duncan, 2010, 2013] has long been implicated in intelligent behavior, and its damage has been associated with lower intelligence and difficulties in problem solving. . . [It] has been linked to our ability to engage in goal-directed behaviors, solve novel problems, and acquire new skills. . . . Damage to this network as a result of stroke, degeneration or head injury leads to poorer executive abilities (attention, working memory, and inhibitory control) and lower fluid intelligence (Glascher et al., 2010; Roca et al., 2010; Woolgar et al., 2010) . . . and aberrant functioning of this network has been reported in a variety of neurological and psychiatric disorders."

Emphasizing interactions between various brain networks in human intelligence, Hearne, et al. (2016, p. 7) state: "How the brain self-reorganizes to achieve optimal configurations of functional networks across individuals with varying levels of intelligence is an open question. Recent neuroimaging work has suggested that transient cooperation between different neural systems, including **fronto-parietal**, **cingulo-opercular and default-mode networks**, is integral to **complex cognitive tasks** such as reasoning[42,43], memory recollection[44] and working memory performance[39,45]. Future studies should test the notion that individual differences in intelligence rely on dynamic, context-specific, reconfigurations of local activity and connectivity within a diffuse system comprising fronto-parietal, cingulo-opercular and default-mode regions[46]."

Hearne, et al. (2016, p. 1) "revealed a novel contribution of across-network interactions between **default-mode and fronto-parietal networks** to individual differences in intelligence at rest (i.e. in the absence of any specific cognitive task). Specifically, [they] found that greater connectivity in the resting state was associated with higher intelligence scores." The **default-mode network** includes the medial prefrontal cortex, posterior cingulate cortex, and the inferior parietal lobule. Other structures sometimes considered part of this network are the middle temporal lobe and the precuneus.

What is the significance of connectivity between networks at rest, when no specific cognitive task is being performed? The answer to this question reflects another method used by neuroscientists to understand the functional and neuroanatomical relationships between networks of interconnected brain areas. Dixon et al. (2017) describe the method this way: "Resting state functional connectivity has emerged as a powerful, non-invasive tool for delineating the functional network architecture of the human brain." In other words, connectivity between brain networks at rest helps neuroscientists determine how various brain networks are "wired" together and how they functionally interact with one another to produce cognition and intelligence.

In this light, remember that Hearne, et al. (2016, p.1) found that "greater connectivity in the resting state [between default-mode and fronto-parietal networks] was associated with higher intelligence scores." To help us understand why this might be so, we have to know something about the suspected role that the default-mode network plays in cognition. Dixon, et al. (2017, p.632) explain it this way: "The default network (DN) is involved in a variety of internally-directed processes, including self-reflection, autobiographical memory, future event simulation, conceptual processing, and spontaneous cognition . . . and exhibits decreased activation during many cognitive tasks that demand external perceptual attention." The finding of decreased activation during cognitive tasks requiring attention to input from the external environment suggested to some researchers that there might be **two competing mutually inhibiting networks**, one for **internal reflection**, and **the other**, **a "task positive" or "task-related"** network **for attention to** the **external world**. However, later research found "co-activation and positive functional connectivity between the DN and the frontoparietal control network (FPCN)—a component of the “task positive” network—during some task conditions, including mind wandering, spontaneous thought, autobiographical future planning, creativity, memory recall, working memory guided by information unrelated to current perceptual input, social working memory, and semantic decision making. . . . On the other hand, studies have generally found anticorrelation [increase in activity in one network is associated with a decrease in the other] between the DN and other components of the “task positive” network, particularly the dorsal attention network (DAN)" (Dixon, et al., 2017, p. 633).

The **“dorsal attention network” (DAN) or “frontoparietal network”** directs visual attention and short-term memory processes . . . Moreover, this network is distinct from a cingulo-opercular cognitive control network (CCN). Yet, no consensus has been reached

regarding the precise components of the DAN. On the basis of task-based and resting-state fMRI studies, the DAN in humans is typically defined to include all or some of the following four regions: (1) intraparietal sulcus (IPS)/ superior parietal lobule; (2) superior pre-central sulcus (sPCS) containing the homolog of primate frontal eye fields; (3) inferior pre-central sulcus (iPCS), alternately known as inferior frontal junction; and (4) the motion-sensitive area MT complex (MT) . . . [Moreover,] subcortical structures, such as superior colliculus and pulvinar, are often implicated in attentional functions" and parts of the cerebellum are also involved in the DAN" (Brissenden, et al., 2016, p. 6083-4).

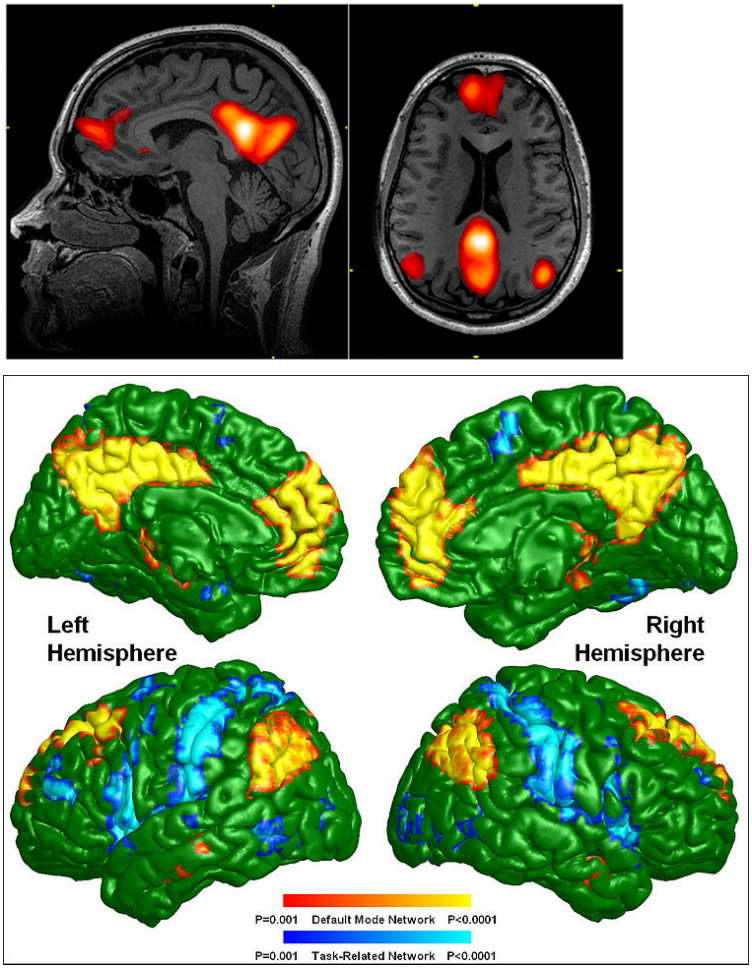


Figure **14.5.3**: (Left): The Default Mode Network, midsagittal and horizontal cross-section views. (Right): Default Mode Network contrasted with Task-Related Network. On a green background, the default mode network is highlighted in warm colors (red and yellow) and the task-related network is highlighted in cold colors (blue and light blue). Top, medial views. Bottom, lateral views. (Images from Wikimedia Commons; Left: File:Default mode network-WRNMMC.jpg; https://commons.wikimedia.org/wiki/F...ork-WRNMMC.jpg; by John Graner, Neuroimaging Department, National Intrepid Center of Excellence, Walter Reed National Military Medical Center; in the **public domain** in the United States because it is a work prepared by an officer or employee of the United States Government as part of that person's official duties. Right: File:Default mode and task-related maps for healthy subjects.jpg; https://commons.wikimedia.org/wiki/F...y_subjects.jpg; by Shim G, Oh JS, Jung WH, et al., Shim G, Oh JS, Jung WH, et al. Altered resting-state connectivity in subjects at ultra-high risk for psychosis: an fMRI study. Behavioral and Brain Functions: BBF. 2010;6:58. doi:10.1186/1744-9081-6-58. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2959003/; Shim G, Oh JS, Jung WH, et al. Altered resting-state connectivity in subjects at ultra-high risk for psychosis: an fMRI study. Behavioral and Brain Functions: BBF. 2010;6:58. doi:10.1186/1744-9081-6-58. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2959003/).

"[A]ttention is a selection mechanism that serves to choose a particular source of stimulation, internal train of thoughts, or a specific course of action for priority processing . . . However, in certain situations attention is necessary to supervise goal-directed

action . . . [and] is necessary for detecting errors, and controlling behavior in dangerous and novel or unpracticed conditions. Thus, attention mechanisms are also central to the generation of voluntary behavior, which often involves inhibition of automatic responses. . . . [Furthermore, attention can be] mostly driven by external stimulation or . . . [by] endogenous processes such as voluntary intentions or expectations. . . . [or] attention can also be directed to an object because of its relevance to our current goals" (Rueda, et al., 2015, p.183). Additionally, sustained attention on a task (such as taking an I.Q. test) must be maintained during distraction. Sustained, focused attention is clearly involved in a person's performance on many measures of intelligence. Rueda, et al. (2015) identify **three types of attention networks: 1) alerting**, such as attention to a sudden loud noise (mediated by norepinephrine); **2) orienting**, shifts of attention from one location to another, involving **parietal and frontal cortices** and which uses acetylcholine; and **3) executive attention**, attention required to deal with distraction and incongruous (conflicting) information especially if incompatible with the task, goal, or problem at hand; executive attention involves **anterior cingulate and prefrontal cortices**, and is modulated by levels of brain dopamine and serotonin.

The close association of attention with ability to concentrate during working memory and problem solving raises the possibility of a confound between attention and intelligence in some of the brain imaging studies supporting the P-FIT and similar models of the biology of intelligence discussed above. In other words, the networks identified with intelligence may turn out to be more accurately defined as attention networks. If so, some of the brain imaging research in intelligence may be accessing attentional networks, rather than intelligence networks. In this case, variations in I.Q. associated with the functioning of these networks may primarily reflect differences among people in attentional mechanisms which then affect performance on I.Q. tests. Attention and working memory, though certainly essential to problem solving, reasoning, and intelligent action, do not capture the reasoning processes themselves. When interpreting research in this area, it is important to be aware of the specific features of the cognitive tasks under study. For example, does any particular imaging study use control procedures that can separate effects of attention and working memory from the reasoning processes of intelligence themselves? It may turn out that this separation is so difficult to accomplish that, in actual practice, the overlap between attention networks and intelligence networks may be hard to tease out. However, attentional networks have been identified such as the *dorsal attention network* (DAN) and a cingulo-opercular cognitive control network (CCN) (and others mentioned above), and the extent to which these do not overlap the networks associated with general fluid intelligence, researchers may be closer to identifying networks involved in reasoning and intelligence as distinct from the mechanisms of attention.

In the meantime, we do know that variations among people in I.Q. and other measures of intelligence are correlated with differences in their brain activity in the fronto-parietal network (as proposed by the P-FIT model and Duncan's MD network), and with the degrees of functional connectivity at rest between this network and the default-mode network. However, additional research is needed to clarify the underlying reasons for these correlations between measures of intelligence and levels of activity in these brain networks and their connectivity.

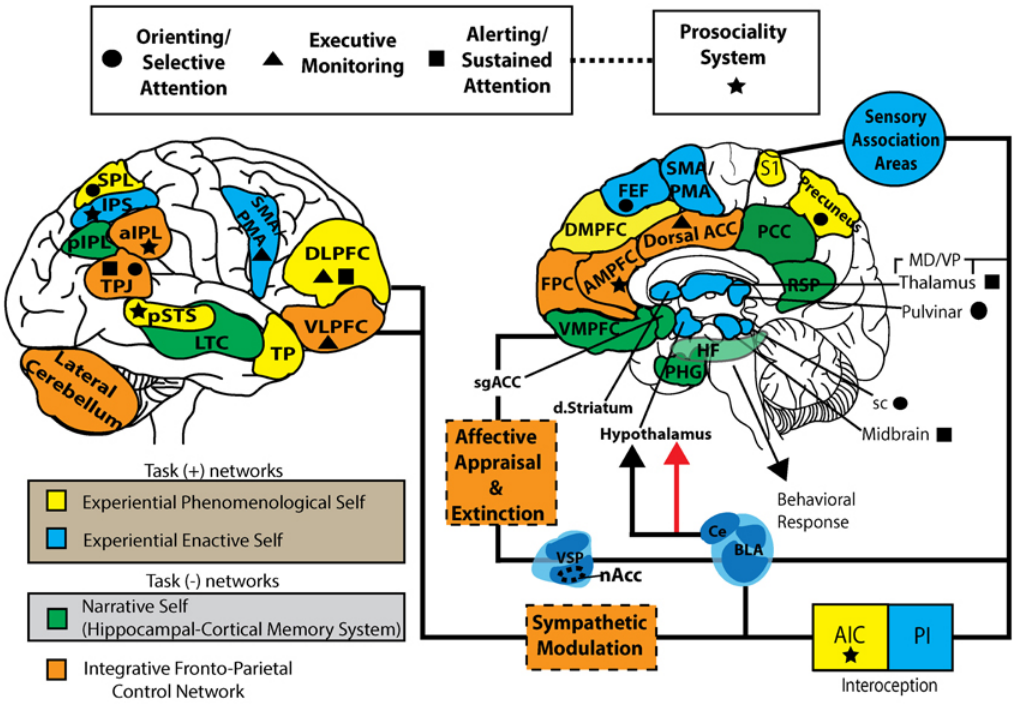


Figure 14.5.4: One model of the complex interactions between various brain areas involved in cognitive control of behavior generating various neural models of the intelligent self in interaction with the internal (mental) and external (sensory) worlds (the foregoing caption by the author of this module, Kenneth A. Koenigshofer, Ph.D.). (The following caption is from the authors of this image). This working model represents a parcellation of task positive (self-specifying: EPS and EES), task negative (NS), and integrative fronto-parietal control networks. EPS, experiential phenomenological self; EES, experiential enactive self; NS, narrative self; FPCN, fronto-parietal control network; FEF, frontal eye fields; DMPFC, dorsal-medial prefrontal cortex; AMPFC, anterior medial prefrontal cortex; VMPFC, ventromedial prefrontal cortex; PHG, parahippocampal gyrus; HF, hippocampal formation; RSP, retrosplenial cortex; PCC, posterior cingulate cortex; Dorsal ACC, dorsal anterior cingulate cortex; DLPFC, dorsolateral prefrontal cortex; VLPFC, ventrolateral prefrontal cortex; TP, temporal pole, LTC, lateral temporal cortex; TPJ, temporoparietal junction; sPL, superior parietal lobe; pIPL, posterior inferior parietal lobe; aIPL, anterior inferior parietal lobe; nAcc, nucleus accumbens; VSP, ventrostriatal pallidum; dstriatum, dorsal striatum; S1, primary somatosensory cortex; AIC, anterior insular cortex; PIC, posterior insular cortex; sgACC, subgenual anterior cingulate cortex; VMpo, ventromedial posterior nucleus; sc, superior colliculus; BLA, basolateral amygdala; CE, central nucleus. (Image from Wikimedia Commons; File:S-ART Mindfulness and brain1.jpg; https://commons.wikimedia.org/wiki/F...and_brain1.jpg; by Vago DR and Silbersweig DA; licensed under the Creative Commons Attribution 3.0 Unported license).

14.5.2 Cellular Basis of Differences in Human Intelligence

Most of the research on the neurological basis of intelligence in humans has focused either on 1) gene loci associated with individual differences in intelligence or 2) is based on using whole brain imaging to identify brain regions correlated with individual differences in intelligence. Some studies show common genes for differences in brain volume and differences in intelligence, while others suggest that genes that facilitate growth of neurons are associated with higher IQ (see section below on genes and intelligence). Functional and structural MRI studies highlight correlations between specific areas of cortex, including frontal and temporal lobes, and measures of g, general intelligence (see discussion of g in prior modules in this chapter). Higher order association cortex in frontal and temporal lobes contains large numbers of pyramidal neurons with large and complex dendritic branching that might account for individual differences in cortical thickness, synaptic integration, and perhaps IQ (Goriounova, et al., 2018). These neurons and their connections are "the principal building blocks for coding, processing, and information storage in the brain and give rise to cognition (Salinas & Sejnowski, 2001). Given their vast number in the human neocortex, even the slightest change in efficiency of information transfer by neurons may translate into large differences in mental ability" (Goriounova, et al., 2018, p. e41715).

As shown in Figure 14.5.5 (below), Goriounova, et al. (2018) found that IQ scores are positively correlated with cortical thickness of the temporal lobe, the length of pyramidal neuron dendrites, and complexity of dendritic branching. "Thus, larger and more complex pyramidal neurons in temporal association area may partly contribute to thicker cortex and link to higher intelligence" (p. e41721).

These same researchers also found evidence that larger temporal lobe pyramidal neurons generated action potentials faster than smaller ones and that faster action potential (AP) onset generates better temporal tracking of high frequency inputs leading to more efficient information coding and transfer in larger temporal lobe pyramidal neurons. They report that "human cortical pyramidal neurons from individuals with higher IQ scores generate faster APs" and therefore "neurons from individuals with higher IQ scores are better equipped to process synaptic signals at high rates and at faster time scales, which is necessary to encode large amounts of information accurately and efficiently" (Goriounova, et al., 2018, p. e41724). These researchers state that their results help explain why faster reaction times in even simple tasks (as a measure of mental processing speed) are consistently associated with higher general intelligence measured by IQ scores on intelligence tests and other measures of Spearman's g (see discussion of g, general intelligence factor, in prior modules of this chapter).

They also note that "Pyramidal cells are integrators and accumulators of synaptic information. Larger dendrites can physically contain more synaptic contacts and integrate more information. Indeed, human pyramidal neuron dendrites receive twice as many synapses as in rodents (DeFelipe et al., 2002) and corticocortical whole-brain connectivity positively correlates with the size of dendrites in these cells" (Goriounova, et al., 2018, p. e41726). Goriounova et al. (2018) claim that larger and more complex dendritic arrays in larger pyramidal neurons are accompanied by increases in integration in cortical areas. "Human pyramidal cells in layers 2 and 3 [of the cerebral cortex] have 3-fold larger and more complex dendrites than in macaque or mouse" which "suggest evolutionary pressure on both dendritic structure and AP waveform and emphasize specific adaptations of human pyramidal cells in association areas for cognitive functions" (p. e41727). They reason, therefore, that larger pyramidal neurons, with more extensive dendritic branching and faster onset of action potentials, are necessary for higher order cortical processing associated with human cognition. Based on this conclusion, they suggest that differences among individuals in "neuronal complexity" might associate with differences among people in mental ability (Figure 14.8.5).

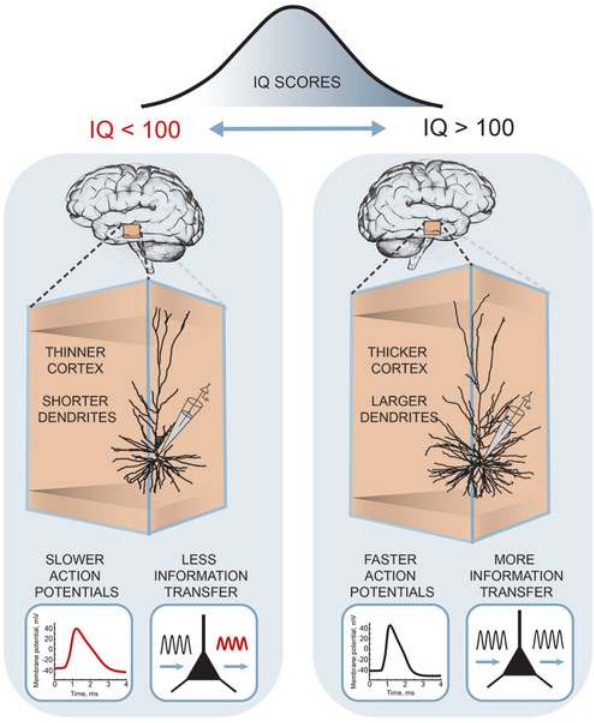


Figure **14.5.5**: A cellular basis of human intelligence. Higher IQ scores associate with larger dendrites, faster action potentials during neuronal activity and more efficient information tracking in pyramidal neurons of temporal cortex. The figure is based on the results from Goriounova et al.(2018). (Image from Wikimedia Commons; File:Cellular basis of IQ.png;

https://commons.wikimedia.org/wiki/F...asis_of_IQ.png; by Natalia A. Goriounova and Huibert D. Mansvelder; licensed under the Creative Commons Attribution 4.0 International license).

14.5.3 Summary of Other Brain Features Correlated with Differences in Human Intelligence

The following table summarizes characteristics of the brain associated with individual differences among humans in general intelligence, as measured by IQ and other psychological tests of intelligence. Based on Goriounova and Mansvelder (2019).

==========

Table 14.5.1: Neurological features (unshaded rows) associated with intelligence (rows 3 and 4 are continuations of rows 1 and 2).

==========

NEUROLOGICAL TRAIT	overall brain volume	cortical thickness in frontal, parietal and temporal association cortex	rapid increase in cortical thickness in childhood with rapid decrease in cortical thickness by early adolescence as pruning of excess synapses occurs increasing processing efficiency	efficiency of processing in distributed cortical areas (frontal, parietal, temporal) as measured by lower neural activity and lower brain metabolic energy required for mental task performance by high I.Q. individuals
CORRELATION OF NEUROLOGICAL TRAIT WITH INTELLIGENCE AS MEASURED BY PSYCHOLOGICAL TESTS (r =, when stated by Goriounova & Mansvelder)	positive, r= between .24 to .33 (correlated with overall brain volume)	positive correlations with IQ (top of this column above, i.e. cortical thickness)	both changes (rapid increase followed by rapid decrease) in cortical thickness positively correlated with high I.Q.	positive correlations with general "fluid" intelligence
NEUROLOGICAL TRAIT (continued)	cortical structure and cortical thickness in lateral areas of temporal lobes and in temporal pole	prefrontal cortex: structure, function and connectivity	functional integrity of white matter (pathways composed of myelinated axons interconnecting brain regions), especially between right frontal and temporal cortex; mental retardation associated with severe damage to white matter	left hemisphere gray matter volume and white matter (myelinated axons) connectivity between left posterior orbital frontal cortex (OFC) and rostral anterior cingulate cortex (rACC)
CORRELATION OF NEUROLOGICAL TRAIT WITH INTELLIGENCE AS MEASURED BY PSYCHOLOGICAL TESTS (continued)	positive correlation with general "crystallized" intelligence dependent on verbal abilities, semantic working memory and acquired knowledge (correlated with structure and cortical thickness of temporal lobes and temporal pole)	positive correlation especially with reasoning ability (i.e. correlated with prefrontal structure, function, and connectivity)	positive correlation with intelligence (i.e. correlated with white matter integrity)	positive correlations with general "fluid" intelligence, "g" (see section 18.11)

==========

Summary Conclusions from Table 14.5.1 (above): "**Conclusions on Gross Brain Distribution of Intelligence:** intelligence is supported by a distributed network of brain regions in many, if not all, higher-order association cortices, also known as parietal-frontal network (Jung and Haier, 2007). This network includes a large number of regions—the dorsolateral prefrontal cortex, the parietal lobe, and the anterior cingulate, multiple regions within the temporal and occipital lobes and, finally, major white matter tracts. Some limited division of function can be observed, implicating frontal and parietal areas in fluid intelligence, temporal lobes in crystallized intelligence and white matter integrity in processing speed. Although brain imaging studies have identified

anatomical and functional correlates of human intelligence, the actual correlation coefficients have consistently been modest, around 0.15–0.35" (Goriounova and Mansvelder, 2019). Note that the amount of variation (variance) in a variable accounted for by its correlation with another variable is equal to the correlation coefficient squared. For example, in this case, a correlation of 0.35 means that only .35 x .35 = .1225 (12.25%) of the variation in intelligence among individuals studied is attributed to its correlation with anatomical and functional features of the brains studied in brain imaging studies, according to Goriounova and Mansvelder (2019). Thus, in the best case, they find that about 88% of the variance in intelligence among individuals studied is accounted for by observed differences among individuals in anatomical and functional features of the brains of those studied using brain imaging. Perhaps the brain imaging is detecting brain regions involved in only some components of intelligence (perhaps attention?) while missing others such as processing of similarity or causal relations, for example.

14.5.4 Genes and Intelligence

Attempts to find single genes that determined intelligence completely failed. The conclusion was that intelligence must be a highly polygenic trait. However, most of the genes that affect intelligence are in non-coding regions of DNA that do not code for protein structure but are instead regulatory genes (which turn on and off other genes) that are involved in generation of cortical neurons during brain development. Different genes correlated with intelligence primarily affect prenatal brain developmental processes including "the proliferation of neural progenitor cells and their specialization, the migration of new neurons to the different layers of the cortex, the projection of axons from neurons to their signaling target and dendritic sprouting" (Goriounova and Mansvelder, 2019).

Genes also affect cell-cell interactions. "Many of the identified genes that play a role in neurodevelopment might contribute to synaptic function and plasticity. Brain function relies on highly dynamic, activity-dependent processes that switch on and off genes. These can lead to profound structural and functional changes and involve formation of new and elimination of unused synapses, changes in cytoskeleton, receptor mobility and energy metabolism. Cognitive ability may depend on how efficient neurons can regulate these processes . . . some candidate genes . . . are specifically involved in axon guidance during neuronal development" (Goriounova and Mansvelder, 2019). This may suggest that efficiency in connections between different circuits and brain areas may be a significant factor in intelligence. Axons that don't go to the right places during brain development would lead to processing inefficiencies.

Other genes that may be involved in differences among individuals in intelligence affect signaling pathways involved in neuron proliferation and migration (during prenatal brain development; see section in Chapter 4 on development of the nervous system) and synaptic communication throughout development.

Other genes that may be involved in differences in intelligence among people are genes that organize pre-synaptic neuron activities especially synaptic vesicles and their release of transmitter.

Others regulate cAMP and CREB involved in gene transcription that for neurons plays a role in synaptic plasticity, learning and memory.

Still others play a role in voltage-gated calcium channel function. Recall that calcium channels open when an action potential reaches an axon ending stimulating movement and binding of synaptic vesicles to the pre-synaptic axonal membrane followed by transmitter release.

Interestingly, most of the energy consumed by the brain (20% of the body's total) is for generation of post-synaptic potentials (PSPs). "Notably, the emergence of higher cognitive functions in humans during evolution is also associated with the increased expression of energy metabolism genes" and "cognitive ability associates with genetic variation in several genes that code for regulators of mitochondrial function" essential to energy metabolism.

"In addition, genes involved in lipid metabolism (BTN2A1 and BTN1A1) and glucose and amino acid metabolism (GPT) . . . [as well as] "microtubule-associated proteins [which] . . . affect recycling of synaptic receptors and neurotransmitter release . . . linked to intelligence by several studies . . . are among the candidate genes of intelligence" (Goriounova and Mansvelder, 2019). The genes affecting microtubule-associated proteins are known to be altered in Alzheimer's, Parkinson's, and Huntington's Disease.

The highest expression of the genes associated with intelligence occur "within pyramidal neurons in hippocampal area CA1 and cortical somatosensory regions . . . [and significant expression] in medium spiny neurons. . . . Pyramidal neurons are the most abundant neuronal types in neocortex and hippocampus, structures associated with higher executive functions, decision-making,

problem-solving and memory. Striatal medium spiny neurons constitute 95% of all neuronal types within the striatum, a structure responsible for motivation, reward, habit learning and behavioral output" (Goriounova and Mansvelder, 2019).

14.5.5 A Central Role for Pyramidal Neurons and Implications of Cross-Species Comparisons

According to Goriounova and Mansvelder (2019), "Genetic studies indicate that expression of genes associated with intelligence accumulates in cortical pyramidal neurons (Savage et al., 2018; Coleman et al., 2019). Comparisons of key cellular properties of pyramidal neurons across species may offer insights into functional significance of such differences for human cognition . . . compared to rodents and macaques, human layer 2/3 pyramidal cells have threefold larger and more complex dendrites (Mohan et al., 2015). Moreover, these large dendrites also receive two times more synapses than rodent pyramidal neurons (DeFelipe et al., 2002). Apart from structural differences, human pyramidal neurons display a number of unique functional properties. human excitatory synapses recover 3–4 times faster from depression than synapses in rodent cortex, have more speedy action potentials and transfer information at up to nine times higher rate than mouse synapse (Testa-Silva et al., 2014)." In humans, larger pyramidal neurons, with longer dendrites, more complex dendritic branching, more synaptic connections, and faster onset of action potentials, combine to make possible the processing of greater amounts of information, with greater efficiency and integration among brain areas, than in other animals, and these large neurons may play a central role in the vast differences between human and non-human animal intelligence (see Figure 14.8.5).

14.5.6 **Brain Correlates of Creativity**

Jung and Haier (2013) report a number of brain correlates of intelligence and creativity. However, they argue against the idea of one brain area for one cognitive function. Instead, as discussed in the prior module, they argue that brain networks involving multiple brain areas are involved in cognition, especially in complex psychological processes such as intelligence and creativity. Nevertheless, they recognize that brain injury and lesion studies reveal brain structures that are necessary, though not sufficient, for certain psychological functions. They give three examples: 1) Phineas Gage, who survived an iron rod passing through his **frontal lobe** resulting in personality and emotional changes as well as impaired judgement and loss of many social inhibitions; 2) "Tan," whose brain damage led to identification of Broca's area for language expression; and 3) H.M., whose bilateral surgical removal of **temporal lobe** structures including **hippocampus** revealed the role of hippocampus and related structures in formation of new long-term explicit memories and their retrieval.

Within this context, Jung and Haier (2013) note some interesting observations from post-mortem examination of the brain of the famous theoretical physicist, Albert Einstein (whose work led to the equation, $E=mc^2$), and what it might suggest about brain mechanisms in creativity. Einstein's brain was unremarkable in many ways. Its size and weight were within the normal range for a man of his age, and frontal and temporal lobe morphology and corpus callosum area were no different from control brains. However, there was one pronounced difference. According to Jung and Haier, Einstein's brain was missing the parietal operculum, the typical location of the secondary somatosensory cortex, resulting in a larger inferior parietal lobule. In Einstein's brain, the **inferior parietal lobule** was approximately 15% wider than in the brains of normal controls. According to Jung and Haier, this region of brain is associated with "visuospatial cognition, mathematical reasoning, and imagery of movement . . . and its expansion was noted in other cases of prominent physicists and mathematicians." They add that further examination of this area of Einstein's brain revealed that rather than more neurons, this region of his brain had a much larger number of **glial cells**, which provide nutrition to neurons, perhaps indicating an unusually large amount of activity among neurons in this region of his brain.

Significantly, as described in the prior module, **parietal cortex** has strong linkages with **prefrontal cortex** forming a **frontoparietal network**: the **inferior parietal lobule** is primarily **connected with dorsolateral prefrontal cortex** (Bruner, 2010), associated, in part, with abilities for abstract thought, while **upper parietal regions**, according to Bruner, as discussed in module 14.2, are associated in the literature with functions such as abstract representation, internal mental images, “imagined world[s],. . . and thought experiment” (i.e., imagination). Jung and Haier detail another study of Einstein's **right prefrontal association cortex**, where researchers found **greater packing density of neurons** (same number of neurons in a smaller space), which was interpreted as **shorter conduction times between cortical neurons in Einstein's brain** compared to control brains. Jung and Haier conclude that **Einstein's brain differed from controls in the frontoparietal network**. These authors have proposed that the **frontoparietal network is crucial to human intelligence**; furthermore they hypothesized that differences among people in the efficiency of neural communication between the frontal and parietal regions of cortex accounts for differences in intelligence in humans (Jung & Haier, 2007). In part, this idea is based on their finding that **high IQ people** show **less activity** in

these brain areas during a complex cognitive task, while lower IQ people show more brain activity, suggesting that high IQ is related to **efficiency** in neural information processing operations. Moreover, higher IQ and ability for abstraction are both inversely correlated with cerebral glucose metabolic rate (Haier et al., 1988, 1992, 2003, 2004), suggesting an efficiency model of individual differences in *g* in which superior **ability for abstraction increases processing efficiency**. In their **Parietal-Frontal Integration Theory (P-FIT)** of the neural basis of intelligence, after sensory processing, information "is then fed forward to the **angular, supramarginal, and inferior parietal cortices**, wherein structural symbolism and/or abstraction are generated and manipulated. The **parietal cortex** then **interacts with frontal regions** that serve to **hypothesis test various solutions** to a given problem." They add that "the **anterior cingulate** is involved in response selection as well as inhibition of competing responses. This process is critically dependent on the fidelity of underlying **white matter** needed to facilitate rapid and error-free transmission of data **between frontal and parietal lobes**" (Jung & Haier, 2013, p. 239). They also note that research in genetics shows that "intelligence and brain structure (i.e., gray and white matter) share common genes" (p. 240).

Regarding creativity specifically, these authors refer to a theory by Flaherty (2005) which proposes a **frontotemporal system** driven by **dopaminergic limbic activity** which provides the drive for creative expression whether art, music, writing, science, etc. and as measured by tests of divergent thinking. Jung and Haier (2013) explain that the **temporal lobe normally inhibits the frontal lobe** so that lesion or mild dysfunction of the temporal lobe releases activity from the frontal lobe by disinhibition causing increased interactions of frontal lobe with other brain regions, sometimes leading to increased creative outputs from neurological patients with left side damage. They argue that this and other data from "three structural studies point to a decidedly left lateralized, frontosubcortical, and disinhibitory network of brain regions underlying creative cognition and achievement" (p. 244). They add that this model, which still requires much more empirical investigation, "appears to include the frontal and temporal lobes, with cortical “tone” being modulated via interactions between the frontal lobes, basal ganglia and thalamus (part of the dopamine system) through white-matter pathways" (p. 244). Although this model is speculative for such a complex form of cognition as creativity, it can guide continuing research into how humans develop creative intellectual and artistic products.

Inferior parietal lobule

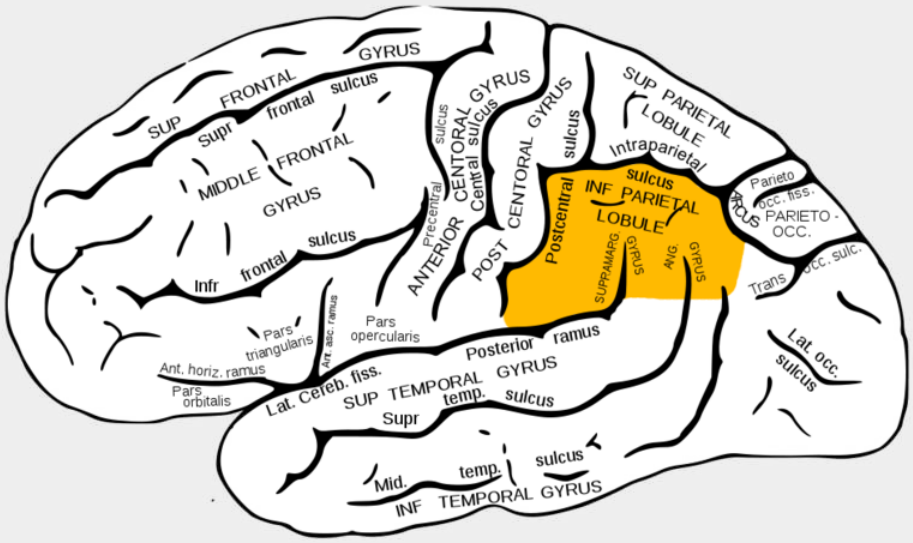


Figure **14.5.6**: Lateral surface of left cerebral hemisphere. Inferior parietal lobule is shown in orange. (Image from Wikimedia Commons; File:Gray726 inferior parietal lobule.png;https://commons.wikimedia.org/wiki/F...tal_lobule.png; by Gray, vectorized by Mysid, colourd by was_a_bee.; this work is **public domain**. This applies worldwide).

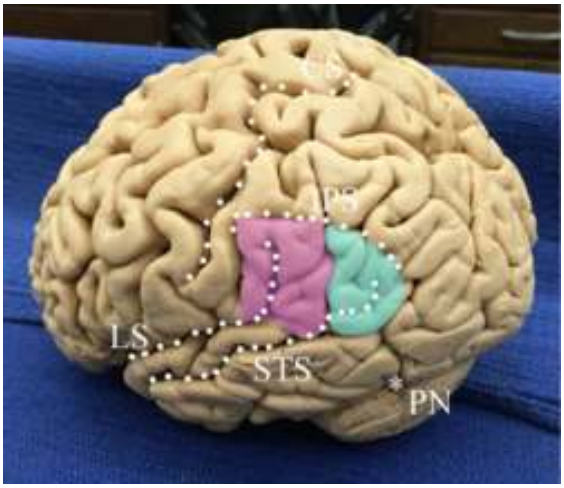


Figure **14.5.7**: Superficial anatomy of the inferior parietal lobule. Purple: Supramarginal gyrus. Blue: Angular gyrus. **LS**: Lateral sulcus (Sylvian fissure). **CS**: Central sulcus. **IPS**: Intraparietal sulcus. **STS**:Superior temporal sulcus. **PN**: Preoccipital notch. (Image and caption from Wikimedia Commons; File:Superficial anatomy of the inferior parietal lobule (IPL).png); https://commons.wikimedia.org/wiki/F...bule_(IPL).png; by Joshua D. Burks Lillian B. Boettcher Andrew K. Conner Chad A. Glenn Phillip A. Bonney Cordell M. Baker Robert G. Briggs Nathan A. Pittman Daniel L. O'Donoghue Dee H. Wu Michael E. Sughrue; licensed under the Creative Commons Attribution 4.0 International license).

14.5.7 Intelligence Testing and Conceptions of Human Intelligence

The development of tests to measure intelligence has had a major impact on the development of ideas about the nature and structure of human intelligence, and its biological basis in the brain. Most theories of human intelligence are based on data derived from intelligence tests, data which is analyzed using factor analysis, a mathematical method for analyzing patterns of correlations among different measures of mental abilities. In section 14.2, we have already discussed how this method, invented and used by Spearman (1904), revealed the "g" factor in human intelligence.

To understand current thinking and research about the biological basis of human intelligence, it is essential to gain at least a general familiarity with the major theoretical models of human intelligence psychologists have developed. For this purpose, this chapter has a supplementary section entitled, "Traditional Models of Human Intelligence." It is highly recommended that students read that supplementary section. The theories we examine in that section are based to a large extent on intelligence testing and factor analysis, while others are more intuitive. That section also introduces key historical figures, major theories of intelligence, and common assessment strategies used to measure human intelligence.

In section 14.2, we discussed a number of enduring, across-generation, universal regularities of the environment which have been incorporated by evolution into brain organization and intelligence. As described in that section, these enduring facts about how the world works include innate, genetically internalized information about objects in three-dimensional space, the passage of time, daily cycles of light and dark, causality relations (forming basis for causal logic and inference), similarity relations (leading to category formation and categorical logic and inference), and predictive relations, based on covariation of events, allowing human and animal brains to mentally project the organism into future time. All of these invariant properties of the world must be included in the brain's neural models or cognitive maps of the world if the brain is to effectively guide adaptive behavior.

When we examine the traditional models of intelligence in the above referenced supplementary section, you will recognize that each focuses on only one, or a few of the facets of intelligence discussed in the evolutionary approach taken in section 14.2. In a sense, each theory discussed in that section is akin to the fable of the blind men trying to describe an elephant. Each blind man only knows that part of the elephant which he happens to feel and so each man has a different and incomplete understanding of the whole. Likewise, each theory of intelligence focuses on only part of the complex of processes that we collectively refer to as "intelligence." Nevertheless, each theory makes a contribution, and each, in one or more ways, is related to the evolutionary discussion in section 14.2.

For example, as you will see, emotional intelligence, including Gardner's intra- and inter-personal intelligence, is related to neural representations of the contingencies of the social environment--brain mechanisms for which are the focus of the new field of **social cognitive neuroscience.** Gardner's multiple intelligences include spatial intelligence related to representation of objects in three-dimensional space, abilities which require portions of **parietal cortex** and **hippocampus**. At Level III in Carroll's theory of intelligence is "g," general intelligence, related to representations of causal, similarity, and predictive relations, likely involving the **frontoparietal network** (Jung & Haier, 2007). Of these theories, the first, Carroll's three-stratum theory of human intelligence is by far the most widely accepted and most productive in terms of explanatory power and empirical evidence. With this background, you will be better prepared to understand the traditional models of intelligence and, perhaps more importantly for this course, you will gain a better understanding of the biological bases of intelligence and thinking, the primary focus of this chapter.

14.5.8 References

Assem, Moataz, Blank, Idan A, Mineroff, Zachary, Ademoğlu, Ahmet and Fedorenko, Evelina. (2020). "Activity in the fronto-parietal multiple-demand network is robustly associated with individual differences in working memory and fluid intelligence." *Cortex*, 131.

Brissenden, J. A., Levin, E. J., Osher, D. E., Halko, M. A., & Somers, D. C. (2016). Functional evidence for a cerebellar node of the dorsal attention network. *Journal of Neuroscience*, *36*(22), 6083-6096.

Dixon, M. L., Andrews-Hanna, J. R., Spreng, R. N., Irving, Z. C., Mills, C., Girn, M., & Christoff, K. (2017). Interactions between the default network and dorsal attention network vary across default subsystems, time, and cognitive states. *Neuroimage*, *147*, 632-649.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in cognitive sciences*, *14* (4), 172-179.

Duncan, J. (2013). The Structure of Cognition: Attentional Episodes in Mind and Brain. *Neuron* 80:35–50. Available at: http://dx.doi.org/10.1016/j.neuron.2013.09.015

Goriounova, N. A., Heyer, D. B., Wilbers, R., Verhoog, M. B., Giugliano, M., Verbist, C., ... & Mansvelder, H. D. (2018). Large and fast human pyramidal neurons associate with intelligence. *Elife*, *7*, e41714.

Goriounova, N. A., & Mansvelder, H. D. (2019). Genes, cells and brain areas of intelligence. *Frontiers in human neuroscience*, 44.

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature neuroscience*, *6* (3), 316-322.

Haier, R. J., Siegel Jr, B. V., Nuechterlein, K. H., Hazlett, E., Wu, J. C., Paek, J., Browning, H.L. and Buchsbaum, M. S. (1988). Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with positron emission tomography. *Intelligence*, *12* (2),199-217.

Haier, R. J., Siegel, B., Tang, C., Abel, L., and Buchsbaum, M. S. (1992). Intelligence and changes in regional cerebral glucose metabolic rate following learning. *Intelligence 16* (3), 415.

Haier, R. J., White, N. S., and Alkire, M. T. (2003). Individual differences in general intelligence correlate with brain function during nonreasoning tasks. *Intelligence*, *31*(5), 429-441.

Haier R.J, Jung R., Yeo R., Head K., Alkire M.T. (2004). Structural brain variation and general intelligence. *NeuroImage*, *23*(1), 425-433.

Hearne, L. J., Mattingley, J. B., & Cocchi, L. (2016). Functional brain networks related to individual differences in human intelligence at rest. *Scientific reports*, *6* (1), 1-8.

Jung, R. E., & Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behavioral and brain sciences*, *30* (2), 135-154.

Jung, R. E., & Haier, R. J. (2013). Creativity and Intelligence: Brain Networks That Link and Differentiate the Expression of Genius

Koenigshofer, K. A. (2017). General Intelligence: Adaptation to Evolutionarily Familiar Abstract Relational Invariants, Not to Environmental or Evolutionary Novelty. *The Journal of Mind and Behavior*, *38* (2), 119-153.

Rueda, M. R., Pozuelos, J. P., & Cómbita, L. M. (2015). Cognitive neuroscience of attention from brain mechanisms to individual differences in efficiency. *AIMS Neuroscience*, *2* (4), 183-202.

Salinas E. & Sejnowski TJ. (2001). Correlated neuronal activity and the flow of neural information. *Nature Reviews Neuroscience 2*:539–550. DOI: https://doi.org/10.1038/35086012, PMID: 11483997

Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, *15*(2), 201-292.

14.5.9 Attributions

"Brain Mechanisms and Intelligence," written by Kenneth A. Koenigshofer, Ph.D., is licensed under CC BY 4.0.

This page titled 14.5: Brain Mechanisms and Intelligence is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

CHAPTER OVERVIEW

15: Language and the Brain

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

15.1: Introduction to Language
15.2: Broca's Area, Wernicke's Area, and Additional Language-Processing Areas in the Brain
15.3: Other Brain Models of Spoken and Written Language
15.4: Handedness, Language, and Brain Lateralization
15.5: The Right Hemisphere's Contribution to Language

This page titled 15: Language and the Brain is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

15.1: Introduction to Language

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. List some of the adaptive functions of animal communication; how do these compare to the adaptive functions of human language
2. Explain why memory is a component of language function
3. List brain structures involved in various types of memory
4. List the stages of speech production during the first 12 months after birth
5. Describe Alzheimer's Disease and the Autism Spectrum Disorders and brain changes associated with each

15.1.1 Overview

In this section, we examine communication in other species, the role of memory in language function, language acquisition, and two disorders, Alzheimer's Disease and Autism, both of which involve language dysfunction as well as other alterations of function.

15.1.2 Communication in Animals

by Kenneth A. Koenigshofer, Ph.D., Chaffey College

Humans are highly social animals, along with many other social species, including social mammals such as wolves, lions, elephants, baboons, chimpanzees, gorillas, water buffalo, dolphins, sea lions, and whales, and social insects such as bees, ants, and wasps. Living in groups has many adaptive advantages such as group hunting, predator defense, food gathering and other forms of cooperation, helping, and, in some mammals such as the mothers in a lion pride, even cooperative nursing of the young within their social group.

Contrast social species with species which are solitary, such as tigers which have no interactions with other tigers except during mating, and when females nurse and care for their offspring. Although both lions and tigers are big cat species, they are highly dissimilar in regard to social life. One reason for this vast difference between lions and tigers is that lions evolved into social creatures because they lived in the open plains of the savannah where group hunting was far more successful than solitary hunting. By contrast, tigers evolved in dense forest where plenty of cover favored quiet stalking and stealth, best accomplished by solitary hunting strategies.

It will come as no surprise to you that animals that live in groups communicate with one another, in part to coordinate their activities, to strengthen social bonds within the group, to warn of danger, to share the location of a food source, to express emotions, to attract mates, to reinforce dominance hierarchies, and so on.

Animal communication can involve any sensory modality. For example, lions of the same group who have been separated greet one another by rubbing their bodies together as they walk past one another. Humans use touch to express affection and sometimes sexual interest. Chimpanzees, like humans, hug and even kiss one another to express affection. Many species mark territory by using the smell of their urine (think of dogs, but male lions do the same, and some species of monkey have scent glands that release a smelly substance on the leaves of the trees they move through). Many animals use visual signals such as bright coloration in many species of birds and tropical fish; other species use visual gestures and facial expressions. For example, chimpanzees raise a hand high above their head and shake it vigorously to signal their displeasure and to threaten other chimps. Bees communicate the direction and distance to a food source by their waggle dance. Wolves and dogs, lions, tigers, primates, and many other species bare their teeth and lower their head as a visual signal communicating threat. Finally, many species utilize auditory means of communication such as screeches, howls, croaks, barks, songs (birds and whales), clicks (dolphins), screams, crying (visual and auditory cues) and laughing in humans, and many other sounds such as those made by crickets or other insects.



Figure 15.1.1: Here a chimpanzee uses visual and auditory signals to communicate threat. A coyote howls to call its pack. (Images from Wikimedia Commons; (left) File:Chimpanzee, Kibale, Uganda (15244558084).jpg; https://commons.wikimedia.org/wiki/F...244558084).jpg; by Rod Waddington; licensed under the Creative Commons Attribution-Share Alike 2.0 Generic license. (right) File:Howl (cropped).jpg; https://commons.wikimedia.org/wiki/F..._(cropped).jpg; by USFWS Mountain-Prairie; licensed under the Creative Commons Attribution 2.0 Generic license).

All of these various forms of communication use sensory signals received by others who interpret their meaning. However, none have the special properties of human language. Human languages use arbitrary symbols (words do not have innate meaning but acquire meaning by learning; there are approximately 600,000 words in English), and words can be strung into sentences by rules of grammar which permit an infinite number of different and complex meanings to be communicated, a situation which linguists sometimes refer to as "the infinite use of finite means," an infinite number of sentences can be constructed from a finite number of words using a finite number of grammatical rules. Although some animals such as chimpanzees, bonobo chimpanzees, and gorillas have been taught American Sign Language with interesting results, the full bloom of human language appears to require specialized brain circuitry present only in specific, although somewhat widespread, areas of the human brain.

15.1.3 Memory and Language

"You need memory to keep track of the flow of conversation" (Goldstein, 2005).

The interaction between memory and language does not seem very obvious at first, but this interaction is necessary when trying to have a conversation. Memory is also required to know and recall the meanings of words.

This is not a simple process which can be learned within days. In childhood everybody learns to communicate, but it is a process lasting for years.
The connection between memory and language becomes most obvious when an impairment occurs when certain brain areas are damaged.

15.1.4 Basics of Memory and Language Interactions

15.1.5 **Memory**

Explicit memory, also known as declarative, can be subdivided into semantic and episodic memory. Procedural memory and priming effects are components of the implicit memory.

Table 14.9.1. Brain regions important in memory and language and their interaction.

Brain regions	Memory
Frontal lobe, parietal lobe, dorsolateral prefrontal cortex	Short-term Memory/ Working Memory
Hippocampus	Short-term Memory → Long-term Memory
Medial temporal lobe (neocortex)	Declarative Memory
Amygdala, Cerebellum	Procedural Memory

15.1.6 Language

Language is an essential system for communication which highly influences our life. This system uses sounds, symbols and gestures for the purpose of communication. Visual and auditory systems are the entrance-pathway for language to the brain. The motor system is responsible for speech and writing production, the exit-pathway for language. In this sense, language processing in the brain (like other types of cognition) occurs between processing by the sensory and motor systems. Most of the knowledge about brain mechanism for language comes from studies of language deficits resulting from brain damage. Though there are about 10,000 different languages and dialects in the world, all of them express the subtleties of human experience and emotion.

15.1.7 Acquisition of language

A phenomenon which occurs daily and in everybody's life is the acquisition of language. Theorists like Catherine Snow and Michael Tomasello think that acquisition of language begins at birth. Babbling in the first six months of life activates brain regions later involved in speech production.

The ability to understand the meaning of words begins before the first birthday, and progresses faster than the ability to speak. Babies show comprehension of more complex sentences, even though they may still be in the one-word stage of speech development.

The different stages of speech production in the first year of life are listed in the table below.

Age	Stage of Acquisition	Example
6th month	Stage of babbling: - systematic combining of vowels and consonants	
7th – 10th month	Stage of repetitive syllable-babbling: - higher part of consonants → paired with a vowel – monosyllabic reduplicated babbling	da, ma, ga mama, dada, gaga
11th – 12th month	Stage of variegated babbling: - combination of different consonants and vowels	bada, dadu
12th month	Usage of first words - : - prephonological → consonant-vowel (-consonant)	car, hat

Researchers like Charlotte Bühler (1928), a German psychologist, think that speaking the first word occurs around the tenth month, whereas Elizabeth Bates et al. (1992) proposed a period between eleven and 13 months. The one-word stage described above can last from two to ten months. Until the second year of life a vocabulary of about 50 words develops, four times more than the child actually uses in speech. Two thirds of the language processed is still babbling. After this stage of learning, vocabulary increases

rapidly. The so-called vocabulary spurt causes an increment of about one word every two hours. From that point on children learn to have fluent conversations with a simple grammar usually containing some errors. Over the first three years of life, the length of sentences and the grammatical output improves.

Children first learn to conjugate verbs and to decline nouns using regular rules. To produce irregular forms is more difficult, because they have to be learned and stored in Long-term memory one by one. The observation of speech is important for acquisition of grammatical skills. Around the third birthday the complexity of language increases exponentially.

15.1.8 Disorders

15.1.8.1 **Alzheimer's Disease**

Discovered in 1906 by Alois Alzheimer, this disease is the most common type of dementia. Alzheimer's is characterized by symptoms such as loss of memory, loss of language skills and impairments in skilled movements. Additionally, other cognitive functions such as planning or decision-making which are connected to the frontal and temporal lobe can be also be impaired. The correlation between memory and language in this context is very important because the two work together in order to establish and maintain conversations. When both are impaired, communication becomes a difficult task. People with Alzheimer's have reduced working memory capability, so they cannot keep in mind all of the information they have heard during a conversation. They also forget words which they need to denote items, express their desires, and to understand what they are told. Affected persons also change their behavior; they become anxious, suspicious or restless and they may have delusions or hallucinations.

In the early stages of the disorder, affected persons become less energetic and may suffer little loss of memory. They are still able to dress themselves, to eat and to communicate enough to get by. Middle stages of the disease are characterized by problems of navigation and orientation. Affected persons may not be able to find their way home or they may even forget where they live. In the late stages of the disease, the patients' ability to speak, read and write is severely impaired. They are no longer able to denote objects and to talk about their feelings and desires. So their family and the nursing staff have great difficulty finding out what the patients want to tell them. In the end-state, persons with Alzheimer's disease do not show any response or reaction. They lie in bed, have to be fed and are totally helpless. Most of them die after four to six years following diagnosis, although the disease can last from three to twenty years. It is sometimes difficult to distinguish Alzheimer's from other related disorders. Only after death when observing the shrinkage of the brain can one definitely diagnose Alzheimer's disease.

In the Alzheimer brain:
· The cortex shrivels up, damaging areas involved in thinking, planning and remembering.
· Shrinkage is especially severe in the hippocampus, which, as discussed in earlier modules, plays a key role in formation of new memories.
· Ventricles (fluid-filled spaces within the brain) grow larger as the surrounding brain tissue dies away.

Long before the first symptoms appear, nerve cells that store and retrieve information have already begun to degenerate. There are two theories about the causes of Alzheimer's disease. The first describes plaques, protein fragments, which impair the synaptic connections between nerve cells. They arise when little fragments released from nerve cell walls associate with other fragments from outside the cell. These combined fragments, the plaques, attach to the outside of nerve cells and destroy the synaptic connections. Then the nerve cells start to die. The second theory explains that tangles limit the functions of nerve cells. They are twisted fibers of another protein that form inside brain cells and destroy a vital cell transport system made of proteins. But scientists have not yet found out the exact role of plaques and tangles.

- Alzheimer tissue has many fewer nerve cells and synapses than a healthy brain.
- Plaques, abnormal clusters of protein fragments, build up between nerve cells.
- Dead and dying nerve cells contain tangles, which are made up of twisted fibers of another protein.

Alzheimer's progress is separated into three stages: In the early stages (1), tangles and plaques begin to evolve in brain areas where learning, memory, thinking and planning takes place. This may begin 20 years before diagnosis. In the middle stages (2), plaques and tangles start to spread to areas for speaking and understanding speech. The sense of where your body is in relation to objects around you is impaired. This may last from 2–10 years. In advanced Alzheimer's disease (3), most of the cortex is damaged, so that the brain starts to shrink seriously and cells begin to die. The people affected lose their ability to speak and communicate and they do not recognize their family or people they know. This stage may generally last from one to five years.

Today, more than 18 million people suffer from Alzheimer’s disease. Alzheimer’s is often only related to older people. Five percent of the people older than 65 years and fifteen to twenty percent of the people older than 80 years suffer from Alzheimer’s. But people in their late thirties and forties can also be affected by this heritable disease. Though heritable, the probability of getting Alzheimer’s when one's parents suffer from the typical older-generation-Alzheimer’s is not very high.

15.1.8.2 **Autism**

Autism is a condition of neurodevelopment, which causes neurodevelopmental disorders in several ways. For more than a decade, autism has been studied in the context of Autistic Spectrum Disorders, including mild and severe autism, as well as Asperger's syndrome. Individuals with autism, for example, have restricted perception and problems in information processing. The often associated intellectual giftedness (savants) only holds for a minority of people with autism; most possess normal or below average intelligence.

There are different types of autism:

- Asperger’s syndrome – usually arising by the age of three years
- infantile autism – arising between nine and eleven months after birth

Two different types of infantile autism are low functioning autism (LFA) and high functioning autism (HFA). LFA describes children with an IQ lower than 80, while HFA refers to those with an IQ higher than 80. The disorders in both types are similar, but they are more severe in children with LFA.

The disorders are mainly defined by the following symptoms:

1. the inability for normal social interaction, e.g. normal relations with other children, perhaps related to impairments in Theory of Mind (TOM), the ability to "read/understand the minds" and intentions of others
2. the inability for ordinary communication, e.g. disorder of spoken language/idiosyncratic language
3. stereotypical behavior, e.g. stereotypical and restricted interests with an atypical content

To investigate the inability of children with autistic disorder to manage normal communication and language, the University of Pittsburgh performed experiments to provide possible explanations for some of their symptoms. Sentences, stories or numbers were presented to children with autism and to normal children. The researchers concluded that the disorders in people with HFA and LFA are caused by an impairment in declarative memory. This impairment leads to difficulties in learning and remembering sentences, stories or personal events, whereas the ability to learn numbers is still available. It has been shown that these children are not able to link words they have heard to their general knowledge, thus the words are only partially learned, often with an idiosyncratic meaning. This may in part explain why LFA and HFA affected children differ in their way of thinking from normal children. It is often difficult for them to understand others and vice versa (perhaps due to deficits in TOM or one or more social processing modules in the brain). Furthermore scientists believe that the process of language learning depends on an initial vocabulary of fully meaningful words. It is assumed that these children do not possess such a vocabulary, thus their language development is impaired. In a few cases the acquisition of language fails completely, therefore in some cases the children are not able to use language in general. The inability to learn and use language may be a consequence of an impairment of declarative memory. This might also cause a low IQ because much of the process of human learning is language-mediated. In HFA the IQ is not significantly lower than the IQ of normal children. This milder form of autism correlates well with their better understanding of word meanings.

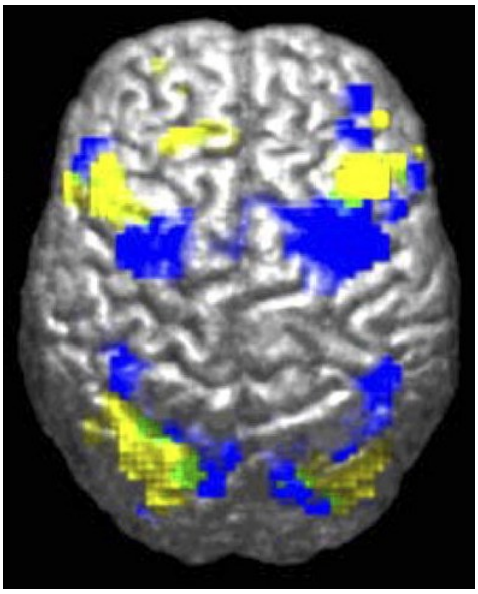


Figure **15.1.2**: fMRI-derived image of difference between brains of autistic and control groups. Activation during visuomotor coordination: Autism Group [yellow], Control Group [Blue], Overlap (both groups) [green]. Even researchers who study autism can display a negative bias against people with the condition. For instance, researchers performing functional magnetic resonance imaging (fMRI) scans systematically report changes in the activation of some brain regions as deficits in the autistic group — rather than evidence simply of their alternative, yet sometimes successful, brain organization. (Image and caption from Wikimedia Commons; File:Powell2004Fig1A.jpeg; https://commons.wikimedia.org/wiki/F...2004Fig1A.jpeg; by Ralph-Axel Müller, Modifications made by Eubulides; licensed under the Creative Commons Attribution 2.5 Generic license. Laurent Mottron, *Changing perceptions: The power of autism* ; Nature 479, 33–35 (03 November 2011) ; doi:10.1038/479033a En ligne : 2011-11-02).

The causes of autism are not yet known. It is still not clear whether the Spectrum Disorders are caused by genetic abnormalities, or non-genetic neurological factors such as brain damage or biochemical abnormalities. One possibility is that during brain development prior to birth new neuron migration and/or neural pruning may be impaired leading to atypical brain circuit formation (Koenigshofer, 2011, 2016).

15.1.9 References

Goldstein, E. G. (2005). *Cognitive Psychology - Connecting Mind, Research, and Everyday Experience*, p. 137, Thomson Wadsworth.

Pinker, S. *The Language Instinct*, p.269f

Koenigshofer, K.A. (2011, 2016 revised e-book edition). *Mind Design: The Adaptive Organization of Human Nature, Minds, and Behavior.* Pearson Learning Solutions, Boston.

15.1.10 Resources

Books

Steven Pinker: The Language Instinct; The Penguin Press, 1994, ISBN 0140175296
Gisela Klann-Delius: Spracherwerb; Sammlung Metzler, Bd 325; Verlag J.B.Metzler; Stuttgart, Weimar, 1999; ISBN 3476103218
Arnold Langenmayr: Sprachpsychologie - Ein Lehrbuch; Verlag für Psychologie, Hogrefe, 1997; ISBN 3801710440
Mark F. Bear, Barry W. Connors, Michael A. Paradiso: Neuroscience - Exploring The Brain; Lippincott Williams & Wilkins, 3rd edition, 2006; ISBN 0781760038

Links

http://de.Wikipedia.org/wiki/Ged%C3%A4chtnis
http://en.Wikipedia.org/wiki/Memory
http://www.cogsci.rpi.edu/CSJarchive/proceedings/2006/docs/p822.pdf
http://www.psychology.uiowa.edu/faculty/gupta/pdf/gupta.brain+lang2003.pdf
http://www.cogsci.rpi.edu/CSJarchive/1980v04/i03/p0243p0284/MAIN.PDF
http://www.quarks.de/gedaechtnis/gedaechtnis.pdf
http://www.alz.org/alzheimers_disease_alzheimers_disease.asp
http://en.Wikipedia.org/wiki/Autism
http://www.apa.org/journals/releases/neu20121.pdf
http://www.esrc.ac.uk/ESRCInfoCentre/Plain_English_Summaries/knowledge_communication_learning
/index22.aspx?ComponentId=17702&SourcePageId=11748

15.1.11 Summary

Although forms of communication exist even in species which are primarily solitary, social species communicate more frequently with other members of their species. Animal communication can utilize signals from any of the sensory modalities. Vocal communication is common in many species, however only our species has developed a form of communication with the complexity of human language. Language use requires memory. Different areas of the brain are involved in different types of memory (as was covered previously in Chapter 10). Normal children begin to comprehend language before the end of their first year. Language expression begins with several stages of babbling typically followed by the production of first words by a year old. Both Alzheimer's and Autism are disorders of the brain that involve language difficulties. fMRI studies reveal that persons with autism have brain organization different from that in persons without autism.

15.1.12 Attributions

"Communication in Animals" is written by Kenneth A. Koenigshofer, Ph.D., Chaffey College

"Memory and Language" adapted by Kenneth A. Koenigshofer, Ph.D., Chaffey College, from Ch. 7, Cognitive Psychology and Cognitive Neuroscience; Wikibooks; https://en.wikibooks.org/wiki/Cognit...e_Neuroscience ; licensed under the Creative Commons Attribution-ShareAlike License.

This page titled 15.1: Introduction to Language is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

15.2: Broca's Area, Wernicke's Area, and Additional Language-Processing Areas in the Brain

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe the classic Wernicke-Lichtheim-Geschwind model of the neurobiology of language.
2. Identify the temporal planum and specific gyri involved in language processing and list the role of each
3. Explain the idea of a universal grammar and what it implies about human language acquisition and a "language acquisition device" (LAD)
4. On what neuroanatomical basis are the Brodmann areas defined?
5. Describe the functions of the three sub-areas within Wernicke's area
6. Which gyrus gives evidence of involvement of the right hemisphere in language processing and what is that evidence?
7. Briefly describe what is meant by a distributed model of cognitive function; provide examples from neurobiological models of intelligence (from previous modules) and of language function

15.2.1 The Classical Wernicke-Lichtheim-Geschwind model of the neurobiology of language

The brain areas included in the best known model of how language is processed by the brain are **Broca's area** in the frontal lobe, **Wernicke's area** in the temporal lobe, and the **arcuate fasciculus** which connects the two. In this model, Broca's area is associated with language expression and Wernicke's area with language comprehension. Damage to any of these produces various forms of **aphasia** (language disorder) with different symptoms. Damage to Broca's area produces **Broca's aphasia** characterized by difficulties with language expression. **Wernicke's aphasia** involves difficulty with speech and language comprehension. Damage to the arcuate fasciculus produces **conduction aphasia** characterized by impaired ability to repeat simple phrases while expression and comprehension remain otherwise unimpaired.

Table 10.14.1: Features of two forms of aphasia (based on National Aphasia Association; https://www.aphasia.org/aphasia-reso...QaAhX6EALw_wcB).

Broca's (Expressive) aphasia	Speech output is severely reduced and is limited mainly to short utterances of less than four words. Finding the right words is often extremely difficult and clumsy. Comprehension can be relatively preserved. Difficulty producing grammatical sentences. May be able to read but limited in writing. Difficulty understanding complex grammatical sentences. Cognitive capabilities not related to speech and language may be fully preserved.
Wernicke's (Receptive) aphasia Additional forms of aphasia are global, mixed transcortical, transcortical motor, transcortical sensory, and anomic aphasia.	Damage in brain areas important for processing meaning of words and spoken language. Profound language comprehension deficits, even for single words or simple sentences. Reading and writing are often severely impaired. Often speak using grammatically correct sentences with normal rate and prosody but sentences make little sense and may include non-existent or incorrect and irrelevant words. Show little awareness of the lack of meaning of their speech. Cognitive capabilities not related to speech and language may be fully preserved.

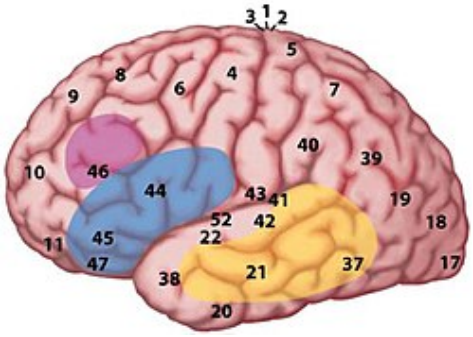


Figure 15.2.1: Lateral view of the human left cerebral cortex. The numbers indicate Brodmann areas. Brodmann areas are cortical regions differentiated by cytoarchitectonics (i.e., composition of cell types).

Broca's area is generally defined as comprising Brodmann areas 44 and 45, which lie anterior to the premotor cortex in the inferior posterior portion of the frontal lobe. Though both area 44 and area 45 contribute to verbal fluency, each seems to have a separate function, so that Broca's area can be divided into two functional units.

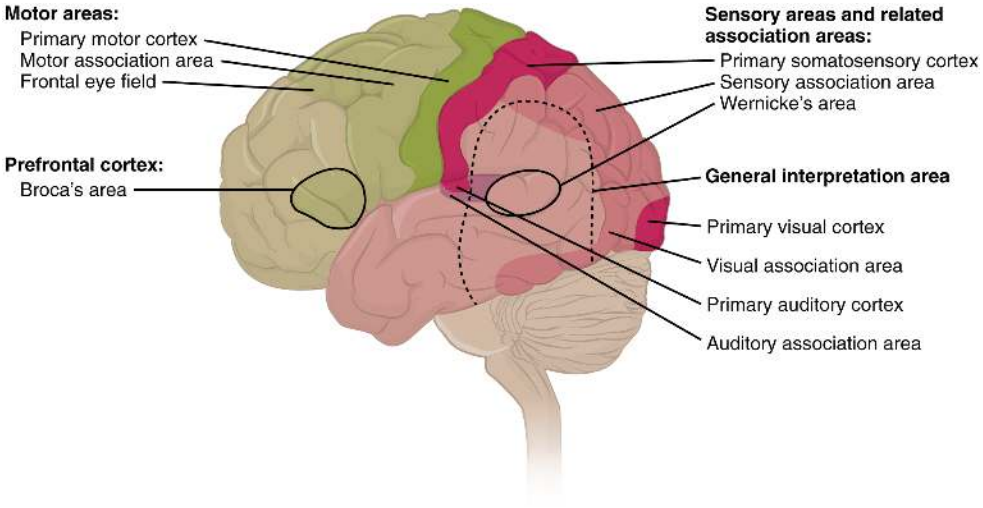


Figure 15.2.2: Locations of Broca's and Wernicke's areas and surrounding cortical regions. (Image from Wikimedia Commons; retrieved 10/18/21).

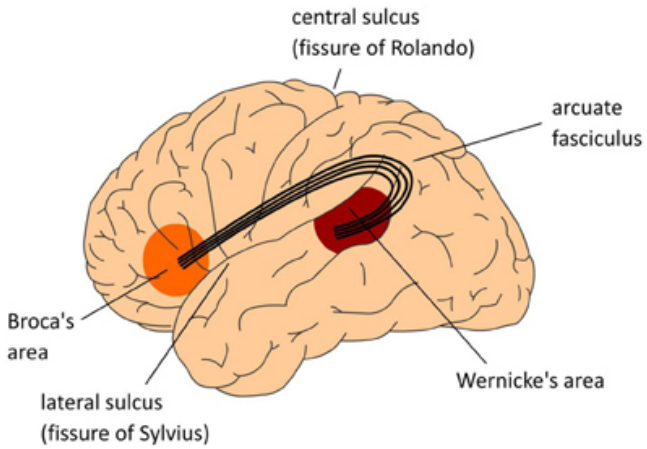


Figure 15.2.3: The classical **Wernicke-Lichtheim-Geschwind model of the neurobiology of language**. In this model Broca's area is crucial for language production, Wernicke's area subserves language comprehension, and the necessary information exchange between these areas (such as in reading aloud) is done via the arcuate fasciculus, a major fiber bundle connecting the language

areas in temporal cortex (Wernicke's area) and frontal cortex (Broca's area). The language areas are bordering one of the major fissures in the brain, the so-called Sylvian fissure. Collectively, this part of the brain is often referred to as perisylvian cortex. (Image and caption from Wikimedia Commons; retrieved 10/18/21).

Area 44 (the **posterior** part of the **inferior frontal gyrus**) seems to be involved in phonological processing and in language production as such; this role would be facilitated by its position close to the motor centres for the mouth and the tongue. Area 45 (the **anterior** part of the **inferior frontal gyrus**) seems more involved in the semantic aspects of language. Though not directly involved in accessing meaning, Broca's area therefore plays a role in verbal memory (selecting and manipulating semantic elements).

Wernicke's area lies in the **left temporal lobe** and, **like Broca's area**, is **no longer regarded as a single, uniform anatomical/functional region** of the brain. By analyzing data from numerous brain-imaging experiments, researchers have now distinguished **three sub-areas within Wernicke's area**. The first responds to spoken words (including the individual's own) and other sounds. The second responds only to words spoken by someone else but is also activated when the individual recalls a list of words. The third sub-area seems more closely associated with producing speech than with perceiving it. All of these findings are still compatible, however, with the **general role of Wernicke's area**, which relates to the representation of phonetic sequences, regardless of whether the individual hears them, generates them himself or herself, or recalls them from memory.

Wernicke's area, of which the **temporal planum** is a key anatomical component, is located on the **superior temporal gyrus**, in the superior portion of Brodmann area 22. This is a strategic location, given the language functions that Wernicke's area performs. The temporal planum lies **between** the **primary auditory cortex** (Brodmann areas 41 and 42) **and the inferior parietal lobule**.

This lobule is composed mainly of **two distinct regions:** caudally, the **angular gyrus** (area 39), which itself is bounded by the visual occipital areas (areas 17, 18, and 19), and dorsally, the **supramarginal gyrus** (area 40) which arches over the end of the lateral sulcus, adjacent to the inferior portion of the somatosensory cortex.

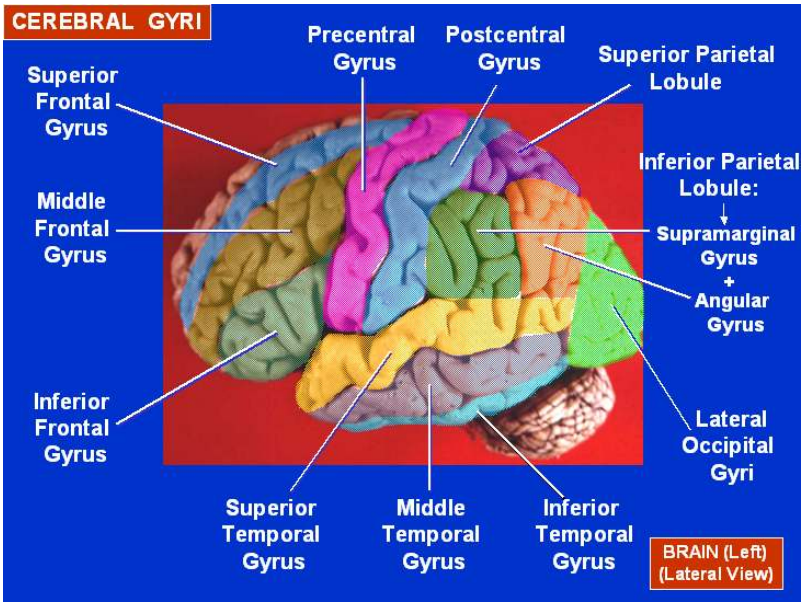


Figure **15.2.4**: . Gyri ("hills") on the cerebral cortex, several of which are involved in language processing. (Image from Wikimedia Commons; File:Cerebral Gyri - Lateral Surface.png; https://commons.wikimedia.org/wiki/F...al_Surface.png; by John A Beal, PhD. Dep't. of Cellular Biology & Anatomy, Louisiana State University Health Sciences Center Shreveport; licensed under the Creative Commons Attribution 2.5 Generic license).

The supramarginal gyrus seems to be involved in phonological and articulatory processing of words, whereas the angular gyrus (together with the **posterior cingulate gyrus**) seems more involved in semantic processing. The **right angular gyrus** appears to be **active as well as the left**, thus revealing that the right hemisphere also contributes to semantic processing of language.

Together, the **angular and supramarginal gyri** constitute a **multimodal associative area** that receives auditory, visual, and somatosensory inputs. The neurons in this area are thus very well positioned to process the phonological and semantic aspect of

language that enables us to identify and categorize objects.

The **language areas of the brain are distinct from the circuits responsible for auditory perception of the words we hear or visual perception of the words we read.** The auditory cortex lets us recognize sounds, an essential prerequisite for understanding language. The visual cortex, which lets us consciously see the outside world, is also crucial for language, because it enables us to read words and to recognize objects as the first step in identifying them by a name.

There are wide variations in the size and position of Broca's area and Wernicke's area as described by various authors.

Brain areas such as these, which perform high-level integration functions, are more heterogeneous than areas that perform primary functions. This greater heterogeneity might reflect greater sensitivity to environmental influences and greater plasticity (Links to an external site.) (ability to adapt to the environmental influences). The functional organization of language would even appear to vary within the same individual at various stages of his or her life!

Any attempt to define the precise boundaries of a particular area of the brain, such as Broca's area or Wernicke's area, will involve some serious problems. But we do know that the **cytoarchitectonic areas described by Brodmann provide better anatomical correlates for brain functions than do the shape of the brain's convolutions**. That said, a cortical area such as Broca's cannot be precisely described by reference to Brodmann areas alone. Though many authors regard **Broca's area** as consisting of **Brodmann areas 44 and 45**, other authors say it consists only of area 44, still others only of area 45, and yet others of areas 44, 45, and 47.

Broca's area may also include the most ventral portion of Brodmann area 6, as well as other parts of the cortex lying deep within the lateral sulcus (sulcus is a "valley" between gyri). It is even possible that only certain parts of these areas are actually dedicated to language.

Brain-imaging studies have shown to what a large extent **cognitive tasks** such as those involving language **correspond to a complex pattern of activation of various areas distributed throughout the cortex.** That a particular area of the brain becomes activated when the brain is performing certain tasks therefore does not imply that this area constitutes the only clearly defined location for a given function. In the more **distributed model of cognitive functions** that is now increasingly accepted by cognitive scientists, all it means is that the neurons in this particular area of the brain are more involved in this particular task than their neighbors. It in no way excludes the possibility that other neurons located elsewhere, and sometimes even quite far from this area, may be just as involved.

Thus, just because the content of a word is encoded in a particular neuronal assembly does not necessarily mean that all of the neurons in this assembly are located at the same place in the brain. On the contrary, **understanding or producing a spoken or written word can require** the simultaneous contribution of several modalities (auditory, visual, somatosensory, and motor). Hence the **interconnected neurons** in the assembly responsible for this task may be **distributed across the various cortexes dedicated to these modalities.**

In contrast, the neuronal assemblies involved in encoding grammatical functions appear to be less widely distributed.

It may therefore be that the **brain processes language functions in two ways simultaneously**: in **parallel** mode by means of **distributed networks**, and in **serial** mode by means of **localized** convergence zones in the brain.

15.2.2 Language Acquisition

Language acquisition in humans is based on our **capacities for abstraction and for applying rules of syntax**—capacities that other animals lack. For example, brain-imaging experiments have shown that Broca's area becomes active when subjects are learning actual rules of grammar in another language, but not when they are exposed to fictitious rules that actually violate the grammar of that language.

These findings suggest that in Broca's area, biological constraints interact with experience to make the acquisition of languages possible. **Broca's area** may thus represent the neuronal substrate of the **"universal grammar"** shared by all of the world's languages. The idea of the universal grammar is that in spite of the great variation in human languages, there still exists basic universal commonalities shared by all human languages and that these basic similarities in grammar reflect **genetically determined** features of the brain areas that acquire language. The linguist, Noam Chomsky, proposed that the **universal grammar** shared by all

languages originates from innate brain organization that biologically prepares humans to learn language naturally and easily despite its great complexity. This idea is expressed in Chomsky's proposal of a **Language Acquisition Device (LAD)**--innate, genetically organized brain circuitry found in every normal human specialized for the easy acquisition of a human language that one is exposed to during childhood. Though popular among many psychologists and linguists for some time, more recently others have proposed that language acquisition does not require an LAD and that general learning mechanisms may be sufficient to account for language acquisition (Behme & Deacon, 2008).

Summary

In the classical **Wernicke-Lichtheim-Geschwind model of the neurobiology of language,** Broca's area is crucial for language production, Wernicke's area subserves language comprehension, and the information exchange between these areas (such as in reading aloud) is done via the arcuate fasciculus. Damage to any of these produces various forms of **aphasia**. Distributed models of cognitive function suggest that language function involves many areas of the brain not included in the classic neurobiological model of language function.

15.2.3 References

Behme, C., & Deacon, S. H. (2008). Language learning in infancy: Does the empirical evidence support a domain specific language acquisition device?. *Philosophical Psychology*, *21*(5), 641-671.

15.2.4 Attributions

Adapted by Kenneth A. Koenigshofer, Ph.D., Chaffey College, from Broca's Area, Wernicke's Area, and other Language-Processing Areas in the Brain by Bruno Dubuc under a Copyleft license.

This page titled 15.2: Broca's Area, Wernicke's Area, and Additional Language-Processing Areas in the Brain is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

15.3: Other Brain Models of Spoken and Written Language

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Contrast Mesulam's model with the Wernicke-Lichtheim-Geschwind model of the brain's processing of language
2. Describe and explain the distributed model of language processing by the brain
3. Explain why in a split brain patient objects presented in the right visual field are processed by the left side of the brain and can be verbally identified and why objects presented in the left visual field cannot be verbally identified
4. Describe the major evidence for and the basic facts of brain lateralization
5. Outline the auditory and visual processing of language
6. Describe the features and anatomical structures involved in three types of aphasia
7. Describe the types of alexia and agraphia and their causes
8. Briefly explain what the processing of sign language, music, and other symbolic systems reveals about language processing

15.3.1 Overview

In this module we consider other models of language representation and function in the brain. The first follows the traditional locationist model which posits specific circuits involved in language processing. The second hypothesizes a non-locationist view which emphasizes distributed networks engaged in parallel processing of language.

15.3.2 An Alternative to the Wernicke-**Lichtheim**-Geschwind Model

In the 1980s, American neurologist Marsel Mesulam proposed an alternative to the Wernicke-Lichtheim-Geschwind model for understanding the brain's language circuits. Mesulam's model posits a hierarchy of networks in which information is processed by levels of complexity.

For example, when you perform simple language processes such as reciting the months of the year in order, the motor and premotor areas for language are activated directly. But when you make a statement that requires a more extensive semantic and phonological analysis, other areas come into play first.

When you hear words spoken, they are perceived by the **primary auditory cortex**, then processed by unimodal associative areas of the cortex: the **superior and anterior temporal lobes** and the **opercular part of the left inferior frontal gyrus.**

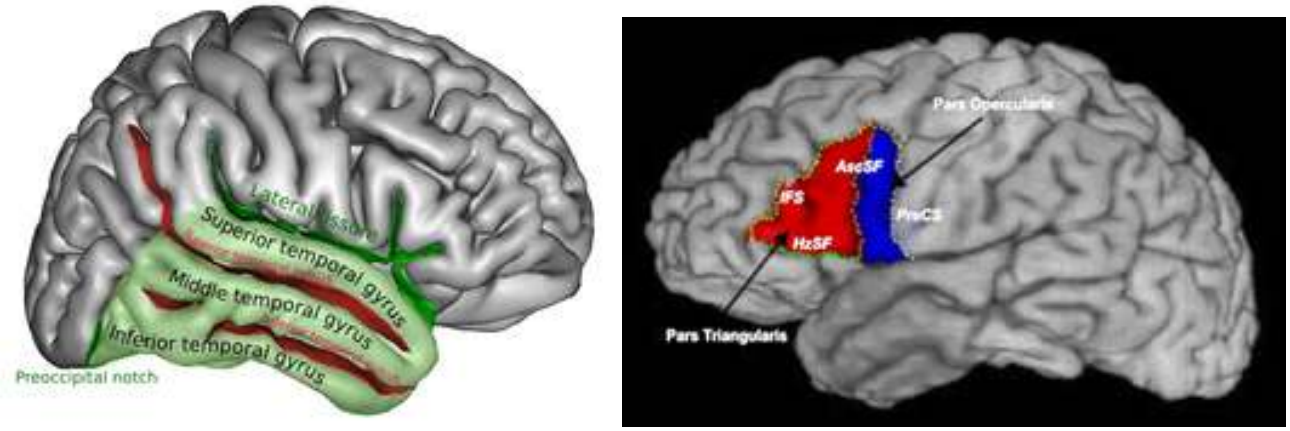


Figure 15.3.1: (Left) Right **temporal lobe**. (Right) Broca's area (left hemisphere) of the **inferior frontal gyrus** divided into the pars triangularis (PTri, red) and pars opercularis (POp, blue). **PreCS**: Precentral sulcus. **IFS**: Inferior frontal sulcus. **AscSF** Ascending

ramus of Sylvian fissure. **HzSF** Horizontal ramus of Sylvian fissure (Images and caption for image on the right from Wikimedia Commons, retrieved 10/18/21).

According to Mesulam's model, these unimodal areas then send their information on to two separate sites for integration. One of these is the temporal pole of the **paralimbic system**, which provides access to the long-term memory system and the emotional system. The other is the posterior terminal portion of the **superior temporal sulcus**, which provides access to meaning. The triangular and orbital portions of the **inferior frontal gyrus** also play a role in semantic processing. One important idea in Mesulam's model is that the function of a brain area dedicated to language is not fixed but rather varies according to the "neural context". In other words, the function of a particular area depends on the task to be performed, because these areas do not always activate the same connections between them. For instance, the left inferior frontal gyrus interacts with different areas depending on whether it is processing the sound of a word or its meaning.

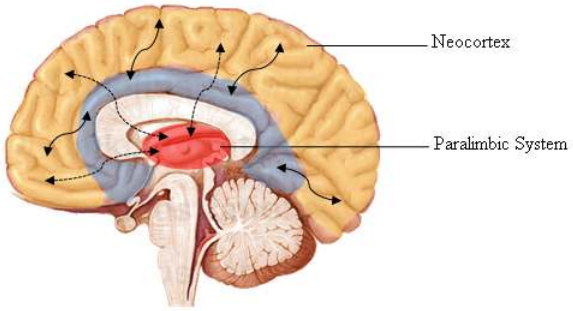


Figure **15.3.2**: Paralimbic system, area of three-layered cortex that consists of the pyriform cortex, entorhinal cortex, and parahippocampal cortex on the medial surface of the temporal lobe, and the cingulate cortex just above the corpus callosum. The paralimbic cortex lies close to the limbic structures, and is directly connected with them. (Image, Wikimedia. Caption from Kolb & Whishaw: *Fundamentals of Human Neuropsychology*, 2003).

This networked type of organization takes us beyond the "one area = one function" equation and explains many of the sometimes highly specific language disorders. For example, some people cannot state the names of tools or the colors of objects. Other people can explain an object's function without being able to say its name, and vice versa.

Mesulam does, however, still believe that there are two "epicenters" for semantic processing, i.e., Broca's area and Wernicke's area. This new conception of these two areas is consistent with the fact that they often work synchronously when the brain is performing a word processing task, which supports the idea that there are very strong connections between them.

Mesulam's concept of epicenters resembles that of convergence zones as proposed by other authors: zones where information obtained through various sensory modalities can be combined. This combining process is achieved through the forming of cell assemblies: groups of interconnected neurons whose synapses have been strengthened by their simultaneous firing, in accordance with Hebb's law. This concept of language areas as convergence zones where neuronal assemblies are established thus accords a prominent place to epigenetic influences in the process of learning a language.

Unquestionably, one of these convergence zones is the **left inferior parietal lobule**, which comprises the **angular gyrus and the supramarginal gyrus** . In addition to receiving information from the right hemisphere , the left inferior parietal lobule also integrates emotional associations from the **amygdala** and the **cingulate gyrus**.

Some scientists believe that over the course of evolution, language remained under limbic control until the inferior parietal lobule evolved and became a convergence zone that provides a wealth of inputs to Broca's area . Some scientists also think that it was the emergence of the inferior parietal lobule that gave humans the ability to break down the sounds that they heard so as to make sense of them and, conversely, to express sounds in a sequential manner so as to convey meaning. In this way, primitive emotional and social vocalizations would have eventually come to be governed by grammatical rules of organization to create what we know as modern language.

15.3.3 Distributed Models of Language Function

Lastly, a number of researchers now reject classic locationist models of language such as Geschwind's and Mesulam's. Instead, they conceptualize language, and cognitive functions in general, as being distributed across anatomically separate areas that process information in parallel (simultaneously, rather than serially, from one "language area" to another).

Even those researchers who embrace this view that linguistic information is processed in parallel still accept that the primary language functions, both auditory and articulatory, are localized to some extent.

This concept of a parallel, distributed processing network for linguistic information constitutes a distinctive epistemological paradigm that is leading to the reassessment of certain functional brain imaging (fMRI) studies.

The proponents of this paradigm believe that the extensive activation of various areas in the left hemisphere and the large number of psychological processes involved make it impossible to associate specific language functions with specific anatomical areas of the brain. For example, the single act of **recalling words** involves a highly **distributed network** that is located primarily in the left brain and that includes the **inferolateral temporal lobe,** the **inferior posterior parietal lobule,** the **premotor areas of the frontal lobe,** the **anterior cingulate gyrus,** and the **supplementary motor area.** According to this paradigm, with such a widely distributed, parallel processing network, there is **no way to ascribe specific functions to each of these structures** that contribute to the performance of this task.

The **brain does seem to access meanings by way of categories that it stores in different physical locations**. For example, if the **temporal pole** (the anterior end of the temporal lobe) is damaged, the category "famous people" is lost; if a lesion occurs in the **intermediate and inferior parts of the temporal lobe**, the category "animals" disappears. It also seems that the networks involved in encoding words activate areas in the motor and visual systems. The task of **naming tools** activates the **frontal premotor areas**, while that of **naming animals** activates the **visual areas**. But in both cases, Broca's area and Wernicke's area are not even activated.

Among those scientists who argue that the brain's language processing system is distributed across various structures, some, such as Philip Lieberman, believe that the **basal ganglia** play a very important role in language. These researchers further believe that other subcortical structures traditionally regarded as involved in motor control, such as the **cerebellum** and the **thalamus**, also contribute to language processing. These views stand in opposition to Chomsky's on the exceptional nature of human language and fall squarely within an adaptationist (emphasizing natural selection), evolutionary perspective.

15.3.4 References

Mesulam, 2000 In: M.M. Mesulam, Editor, Principles of Behavioral and Cognitive Neurology (2nd ed.), Oxford University Press, New York (2000)

15.3.5 Lateralization, Auditory and Visual Language Processing

15.3.6 **Introduction**

What is happening inside my head when I listen to a sentence? How do I process written words? This chapter will take a closer look on brain processes concerned with language comprehension. Dealing with natural language understanding, we distinguish between the neuroscientific and the psycholinguistic approach. As text understanding spreads through the broad field of cognitive psychology, linguistics, and neurosciences, our main focus will lay on the intersection of two latter, which is known as neurolinguistics.

Different brain areas need to be examined in order to find out how words and sentences are being processed. For long time scientist were restricted to draw conclusions from certain brain lesions to the functions of corresponding brain areas. During the last 40 years techniques for brain imaging and ERP-measurement have been established which allow for a more accurate identification of brain parts involved in language processing.

Scientific studies on these phenomena are generally divided into research on auditory and visual language comprehension; we will discuss both. Not to forget is that it is not enough to examine English: To understand language processing in general, we have to look at non-Indo-European and other language systems like sign language. But first of all we will be concerned with a rough localization of language in the brain.

15.3.7 Lateralization of language

Although functional lateralization studies and analyses find individual differences in personality or cognitive style don't favor one hemisphere or the other, some brain functions occur in one or the other side of the brain. Language tends to be on the left and attention on the right (Nielson, Zielinski, Ferguson, Lainhart & Anderson, 2013). There is a lot of evidence that each brain hemisphere has its own distinct functions in language comprehension. Most often, the right hemisphere is referred to as the non-dominant hemisphere and the left is seen as the dominant hemisphere. This distinction is called lateralization (from the Latin word lateral, meaning sidewise) and reason for it first was raised by experiments with split-brain patients. Following a top-down approach we will then discuss the right hemisphere which might have the mayor role in higher level comprehension, but is not quite well understood. Much research has been done on the left hemisphere and we will discuss why it might be dominant before the following sections discuss the fairly well understood fundamental processing of language in this hemisphere of the brain.

15.3.7.1 Functional asymmetry

Anatomical differences between left and right hemisphere

Initially we will consider the most apparent part of a differentiation between left and right hemisphere: Their differences in shape and structure. As visible to the naked eye there exists a clear asymmetry between the two halves of the human brain: The right hemisphere typically has a bigger, wider and farther extended frontal region than the left hemisphere, whereas the left hemisphere is bigger, wider and extends farther in it’s occipital region (M. T. Banich,"Neuropsychology", ch.3, pg.92). Significantly larger on the left side in most human brains is a certain part of the temporal lobe’s surface, which is called the planum temporale. It is localized near Wernicke’s area and other auditory association areas, wherefore we can already speculate that the left hemisphere might be stronger involved in processes of language and speech treatment.

In fact such a left laterality of language functions is evident in 97% of the population (D. Purves, "Neuroscience", ch.26, pg.649). But actually the percentage of human brains, in which a "left-dominance" of the planum temporale is traceable, is only 67% (D. Purves, "Neuroscience", ch.26, pg.648). Which other factors play aunsolved yet.

Evidence for functional asymmetry from "split brain" patients

In hard cases of epilepsy a rarely performed but popular surgical method to reduce the frequency of epileptic seizures is the so-called corpus callosotomy. Here a radical cut through the connecting "communication bridge" between right and left hemisphere, the corpus callosum, is done; the result is a "split-brain". For patients whose corpus callosum is cut, the risk of accidental physical injury is mitigated, but the side-effect is striking: Due to this eradicative transection of left and right half of the brain these two are not longer able to communicate adequately. This situation provides the opportunity to study differentiation of functionality between the hemispheres. First experiments with split-brain patients were performed by Roger Sperry and his colleagues at the California

Institute of Technology in 1960 and 1970 (D. Purves, "Neuroscience", ch.26, pg.646). They lead researchers to sweeping conclusions about laterality of speech and the organization of the human brain in general.

A digression on the laterality of the visual system

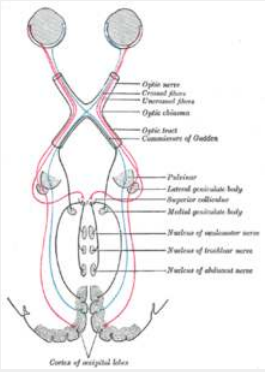


Figure **15.3.3**: Visual system showing wiring from each half of the eye to brain. A visual stimulus, located within the left visual field, projects onto the nasal (inner) part of the left eye's retina and onto the temporal (outer) part of the right eye's retina. Images on the temporal retinal region are processed in the visual cortex of the same side of the brain (ipsilateral), whereas nasal retinal information is mapped onto the opposite half of the brain (contralateral). The stimulus within the left visual field will completely arrive in the right visual cortex to be processed and worked up. In "healthy" brains this information furthermore reaches the left hemisphere via the corpus callosum and can be integrated there. In split-brain patients this current of signals is interrupted; the stimulus remains "invisible" for the left hemisphere. (Image from Wikimedia Commons; File:Gray722.png; by Henry Vandyke Carter; This work is in the **public domain** in the United States).



Figure **15.3.4**: Split Brain Experiments. (Image from Wikimedia Commons; File:SplitBrainExperiments.jpg; https://commons.wikimedia.org/wiki/F...xperiments.jpg; by TilmanATuos at English Wikibooks; released into the **public domain** by its author).

The experiment we consider now is based on the laterality of the visual system: What is seen in the left half of the visual field will be processed in the right hemisphere and vice versa. Aware of this principle a test operator presents the picture of an object to one half of the visual field while the participant is instructed to name the seen object, and to blindly pick it out of an amount of concrete objects with the contralateral hand.

It can be shown that a picture, for example the drawing of a die, which has only been presented to the left hemisphere, can be named by the participant ("I saw a die"), but is not selectable with the right hand (no idea which object to choose from the table). Contrarily the participant is unable to name the die, if it was recognized in the right hemisphere, but easily picks it out of the heap of objects on the table with the help of the left hand.

These outcomes are clear evidence of the human brain's functional asymmetry. The left hemisphere seems to dominate functions of speech and language processing, but is unable to handle spatial tasks like vision-independent object recognition. The right hemisphere seems to dominate spatial functions, but is unable to process words and meaning independently. In a second experiment evidence arose that a split-brain patient can only follow a written command (like "get up now!"), if it is presented to the left hemisphere. The right hemisphere can only "understand" pictorial instructions.

The following table (D. Purves, "Neuroscience", ch.26, pg.647) gives a rough distinction of functions:

Left Hemisphere	Right Hemisphere
• analysis of right visual field • language processing ◦ writing ◦ speech	• analysis of left visual field • spatial tasks • visuospatial tasks • object and face recognition

First it is important to keep in mind that these distinctions comprise only functional dominances, no exclusive competences. In cases of unilateral brain damage, often one half of the brain takes over tasks of the other one. Furthermore it should be mentioned that this experiment works only for stimuli presented for less than a second. This is because not only the corpus callosum, but as well some subcortical comissures serve for interhemispheric transfer. In general both can simultaneously contribute to performance, since they use complement roles in processing.

A digression on handedness

An important issue, when exploring the different brain organization in the two hemispheres, is handedness, which is the tendency to use the left or the right hand to perform activities. Throughout history, left-handers, which only comprise about 10% of the population, have often been considered being something abnormal. They were said to be evil, stubborn, defiant and were, even until the mid 20th century, forced to write with their right hand.
One most commonly accepted idea, as to how handedness affects the hemispheres, is the brain hemisphere division of labour. Since both speaking and handiwork require fine motor skills, the presumption here is that it would be more efficient to have one brain hemisphere do both, rather than having it divided up. Since in most people, the left side of the brain controls speaking, right-handedness predominates. The theory also predicts that left-handed people have a reversed brain division of labour.
In right handers, verbal processing is mostly done in the left hemisphere, whereas visuospatial processing is mostly done in the opposite hemisphere. Therefore, 95% of speech output is controlled by the left brain hemisphere, whereas only 5% of individuals control speech output in their right hemisphere. Left-handers, on the other hand, have a heterogeneous brain organization. Their brain hemisphere is either organized in the same way as right handers, the opposite way, or even such that both hemispheres are used for verbal processing. But usually, in 70% of the cases, speech is controlled by the left-hemisphere, 15% by the right and 15% by either hemisphere. When the average is taken across all types of left-handedness, it appears that left-handers are less lateralized.
After, for example, damage occurs to the left hemisphere, it follows that there is a visuospatial deficit, which is usually more severe in left-handers than in right-handers. Dissimilarities may derive, in part, from differences in brain morphology, which concludes from asymmetries in the planum temporale. Still, it can be assumed that left-handers have less division of labour between their two hemispheres than right-handers do and are more likely to lack neuroanatomical asymmetries.
There have been many theories as to find out why people are left-handed and what its consequences may be. Some people say that left-handers have a shorter life span or higher accident rates or autoimmune disorders. According to the theory of Geschwind and Galaburda, there is a relation to sex hormones, the immune system, and profiles of cognitive abilities that determine, whether a person is left-handed or not. Concludingly, many genetic models have been proposed, yet the causes and consequences still remain a mystery (M.T.Banich, "Neuropsychology", ch.3, pg. 119).

15.3.7.2 The right hemisphere

The role of the right hemisphere in text comprehension

The experiments with "split-brain" patients and evidence that will be discussed soon suggest that the right hemisphere is usually not (but in some cases, e.g. 15% of left handed people) dominant in language comprehension. What is most often ascribed to the right hemisphere is cognitive functioning. When damage is done to this part of the brain or when temporal regions of the right hemisphere are removed, this can lead to cognitive-communication problems, such as impaired memory, attention problems, and

poor reasoning (L. Cherney, 2001). Investigations lead to the conclusion that the right hemisphere processes information in a gestalt and holistic fashion, with a special emphasis on spatial relationships. Here, an advantage arises for differentiating two distinct faces because it examines things in a global manner and it also reacts to lower spatial, and also auditory, frequency. The former point can be undermined with the fact that the right hemisphere is capable of reading most concrete words and can make simple grammatical comparisons (M. T. Banich,“Neuropsychology“, ch.3, pg.97). But in order to function in such a way, there must be some sort of communication between the brain halves.

Prosody - the sound envelope around words

Consider how different the simple statement "She did it again" could be interpreted in the following context taken from Banich: LYNN: Alice is way into this mountain-biking thing. After breaking her arm, you'd think she'd be a little more cautious. But then yesterday, she went out and rode Captain Jack's. That trail is gnarly - narrow with lots of tree roots and rocks. And last night, I heard that she took a bad tumble on her way down. SARA: *She did it again* Does Sara say that with rising pitch or emphatically and with falling intonation? In the first case she would ask whether Alice has injured herself again. In the other case she asserts something she knows or imagines: That Alice managed to hurt herself a second time. Obviously the sound envelope around words - prosody - does matter.

Reason to belief that recognition of prosodic patterns appears in the right hemisphere arises when you take into account patients that have damage to an anterior region of the right hemisphere. They suffer from **aprosodic** speech, that is, their utterances are all at the same pitch. They might sound like a robot from the 80ties. There is another phenomena appearing from brain damage: **dysprosodic** speech. In that case the patient speaks with disordered intonation. This is not due to a right hemisphere lesion, but arises when damage to the left hemisphere is suffered. The explanation is that the left hemisphere gives ill-timed prosodic cues to the right hemisphere, thus proper intonation is affected.

Beyond words: Inference from a neurological point of view

On the word level, the current studies are mostly consistent with each other and with findings from brain lesion studies. But when it comes to the more complex understanding of whole sentences, texts and storylines, the findings are split. According to E. C. Ferstl’s review “The Neuroanatomy of Text Comprehension. What’s the story so far?” (2004), there is evidence for and against right hemisphere regions playing the key role in pragmatics and text comprehension. On the current state of knowledge, we cannot exactly say how and where cognitive functions like building situation models and inferencing work together with “pure” language processes.

As this chapter is concerned with the neurology of language, it should be remarked that patients with right hemisphere damage have difficulties with inferencing. Take into account the following sentence:

With mosquitoes, gnats, and grasshoppers flying all about, she came across a small black bug that was being used to eavesdrop on her conversation.

You might have to reinterpret the sentence until you realize that "small black bug" does not refer to an animal but rather to a spy device. People with damage in the right hemisphere have problems to do so. They have difficulty to follow the thread of a story and to make inferences about what has been said. Furthermore they have a hard time understanding non-literal aspects of sentences like metaphors, so they might be really horrified when they hear that someone was "Crying her eyes out".

The reader is referred to the next chapter for a detailed discussion of Situation Models

15.3.7.3 The left hemisphere

Further evidence for left hemisphere dominance: The Wada technique

Before concerning concrete functionality of the left hemisphere, further evidence for the dominance of the left hemisphere is provided. Of relevance is the so-called Wada technique, allowing testing which hemisphere is responsible for speech output and usually being used in epilepsy patients during surgery. It is not a brain imaging technique, but simulates a brain lesion. One of the hemispheres is anesthetized by injecting a barbiturate (sodium amobarbital) in one of the patient’s carotid arteries. Then he is asked to name a number of items on cards. When he is not able to do that, despite the fact that he could do it an hour earlier, the concerned hemisphere is said to be the one responsible for speech output. This test must be done twice, for there is a chance that the patient produces speech bilaterally. The probability for that is not very high, in fact, according to Rasmussen & Milner 1997a

(as referred to in Banich, p. 293) it occurs only in 15% of the left-handers and none of the right-handers. (It is still unclear where these differences in left-handers' brains come from.)

That means that in most people, only one hemisphere "produces" speech output – and in 96% of right-handers and 70% of left-handers, it is the left one. The findings of the brain lesion studies about asymmetry were confirmed here: Normally (in healthy right-handers), the left hemisphere controls speech output.

Explanations of left hemisphere dominance

Two theories why the left hemisphere might have special language capacities are still discussed. The first states that dominance of the left hemisphere is due to a specialization for **precise temporal control of oral and manual articulators**. Here the main argument is that gestures related to a story line are most often made with the right and therefore by the left hemisphere controlled hand whilst other hand movements appear equally often with both hands. The other theory says that the left hemisphere is dominant because it is specialized for **linguistic processing** and is due to a single patient - a speaker of American Sign Language with a left hemisphere lesion. He could neither produce nor comprehend ASL, but could still communicate by using gestures in non-linguistic domains.

How innate is the organizational structure of the brain?

Not only cases of left-handers but also brain imaging techniques have shown examples of bilateral language processing: According to ERP studies (by Bellugi et al. 1994 and Neville et al. 1993 as cited in E. Dabrowska, "Language, Mind an Brain" 2004, p. 57), people with the Williams' syndrome (WS) also have no dominant hemisphere for language. WS patients have a lot of physical and mental disorders, but show, compared to their other (poor) cognitive abilities, very good linguistic skills. And these skills do not rely on one dominant hemisphere, but both of them contribute equally. So, whilst the majority of the population has a dominant left hemisphere for language processing there are a variety of exceptions to that dominance. That there are different "organization possibilities" in individual brains Dabrowska (p. 57) suggests that the organizational structure in the brain could be less innate and fixed as it is commonly thought.

15.3.8 Auditory Language Processing

This section will explain where and how language is processed. To avoid intersections with visual processes we will firstly concentrate on spoken language. Scientists have developed three approaches of conceiving information about this issue. The first two approaches are based upon brain lesions, namely aphasia, whereas the recent approach relies on results of on modern brain-image techniques.

15.3.8.1 Neurological Perspective

The Neurological Perspective describes which pathways language follows in order to be comprehended. Scientists revealed that there are concrete areas inside the brain where concrete tasks of language processing are taking place. The most known areas are the Broca and the Wernicke Area.

Broca's aphasia

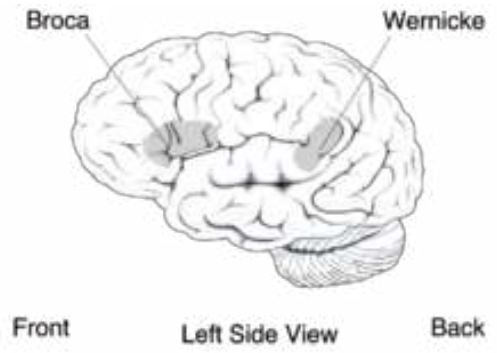


Figure **15.3.5**: Broca's and Wernicke's area (Image from Wikimedia Commons; File:BrocasAreaSmall.png; https://commons.wikimedia.org/wiki/F...sAreaSmall.png; **public domain** in the United States because it is a work prepared by an officer or employee of the United States Government as part of that person's official duties.

One of the most well-known aphasias is Broca's aphasia that causes patients to be unable to speak fluently. Moreover they have a great difficulty producing words. Comprehension, however, is relatively intact in those patients. Because these symptoms do not result from motoric problems of the vocal musculature, a region in the brain that is responsible for linguistic output must be lesioned. Broca discovered that the brain region causing fluent speech is responsible for linguistic output, must be located ventrally in the frontal lobe, anterior to the motor strip. Recent research suggested that Broca`s aphasia results also from subcortical tissue and white matter and not only cortical tissue.

Example of spontaneous Speech - Task: What do you see on this picture?
„O, yea. Det‘s a boy an‘ girl... an‘ ... a ... car ... house... light po‘ (pole). Dog an‘ a ... boat. ‚N det‘s a ... mm ... a ... coffee, an‘ reading. Det‘s a ... mm ... a ... det‘s a boy ... fishin‘.“ (Adapted from „Principles of Neuroscience“ 4th edition, 2000, p 1178)

Wernicke‘s aphasia

Another very famous aphasia, known as Wernicke`s aphasia, causes opposite syndromes. Patients suffering from Wernicke`s aphasia usually speak very fluently, words are pronounced correctly, but they are combined senselessly – “word salad” is the way it is most often described. Understanding what patients of Wernicke`s aphasia say is especially difficult, because they use paraphasias (substitution of a word in verbal paraphasia, of word with similar meaning in semantic paraphasia, and of a phoneme in phonemic paraphasia) and neologisms. With Wernicke`s aphasia the comprehension of simple sentences is a very difficult task. Moreover their ability to process auditory language input and also written language is impaired. With some knowledge about the brainstructure and their tasks one is able to conclude that the area that causes Wernicke`s aphasia, is situated at the joint of temporal, parietal and occipital regions, near Heschl`s gyrus (primary auditory area), because all the areas receiving and interpreting sensory information (posterior cortex), and those connecting the sensory information to meaning (parietal lobe) are likely to be involved.

Example of spontaneous Speech - Task: What do you see on this picture?
„Ah, yes, it‘s ah ... several things. It‘s a girl ... uncurl ... on a boat. A dog ... ‘S is another dog ... uh-oh ... long‘s ... on a boat. The lady, it‘s a young lady. An‘ a man a They were eatin‘. ‘S be place there. This ... a tree! A boat. No, this is a ... It‘s a house. Over in here ... a cake. An‘ it‘s, it‘s a lot of water. Ah, all right. I think I mentioned about that boat. I noticed a boat being there. I did mention that before ... Several things down, different things down ... a bat ... a cake ... you have a ...“ (adapted from „Principles of Neuroscience“ 4th edition, 2000, p 1178)

Conduction aphasia

Wernicke supposed that an aphasia between Broca‘s area and Wernicke‘s area, namely conduction aphasia, would lead to severe problems to repeat just heard sentences rather than having problems with the comprehension and production of speech. Indeed patients suffering from this kind of aphasia show an inability to reproduce sentences since they often make phonemic paraphasias, may substitute or leave out words, or might say nothing. Investigations determined that the "connection cable", namely the arcuate fasciculus between Wernicke‘s and Broca‘s area is almost invariably damaged in case of a conduction aphasia. That is why conduction aphasia is also regarded as a disconnection syndrome (the behavioural dysfunction because of a damage to the connection of two connected brain regions).

Example of the repetition of the sentence „The pastry-cook was elated“:
„The baker-er was /vaskerin/ ... uh ...“ (adapted from „Principles of Neuroscience“ 4th edition, 2000, p 1178)

Transcortical motor aphasia and global aphasia

Transcortical motor aphasia, another brain lesion caused by a connection disruption, is very similar to Broca`s aphasia, with the difference that the ability to repeat is kept. In fact people with a transcortical motor aphasia often suffer from echolalia, the need to repeat what they just heard. Usually patients` brain is damaged outside Broca`s area, sometimes more anterior and sometimes more superior. Individuals with transcortical sensory aphasia have similar symptoms as those suffering from Wernicke`s aphasia, except that they show signs of echolalia. Lesions in great parts of the left hemisphere lead to global aphasia, and thus to an inability of

both comprehending and producing language, because not only Broca`s or Wenicke`s area is damaged. (Barnich, 1997, pp. 276–282)

Type of Aphasia	Spontaneous Speech	Paraphasia	Comprehension	Repetition	Naming
• Broca`s • Wernicke`s • Conduction • Transcortical motor • Transcortical sensory • Global	• Nonfluent • Fluent • Fluent • Nonfluent • Fluent • Nonfluent	• Uncommon • Common (verbal) • Common (literal) • Uncommon • Common • Variable	• Good • Poor • Good • Good • Poor • Poor	• Poor • Poor • Poor • Good (echolalia) • Good (echolalia) • Poor	• Poor • Poor • Poor • Poor • Poor • Poor

Overview of the effects of aphasia from the neurological perspective

(Adapted from Benson, 1985,p. 32 as cited in Barnich, 1997, p. 287)

15.3.8.2 Psychological Perspective

Since the 1960‘s psychologists and psycholinguists tried to resolve how language is organized and represented inside the brain. Patients with aphasias gave good evidence for location and discrimination of the three main parts of language comprehension and production, namely phonology, syntax and semantics.

Phonology

Phonology deals with the processing of meaningful parts of speech resulting from the mere sound. More over there exists a differentiation between a phonemic representation of a speech sound which are the smallest units of sounds that leads to different meanings (e.g. the /b/ and /p/ in bet and pat) and phonetic representation. The latter means that a speech sound may be produced in a different manner at different situations. For instance the /p/ in pill sounds different than the /p/ in spill since the former /p/ is aspirated and the latter is not.

Examining which parts are responsible for phonetic representation, patients with Broca`s or Wernicke`s aphasia can be compared. As the speech characteristic for patients with Broca`s aphasia is non-fluent, i.e. they have problems producing the correct phonetic and phonemic representation of a sound, and people with Wernicke`s aphasia do not show any problems speaking fluently, but also have problems producing the right phoneme. This indicates that Broca`s area is mainly involved in phonological production and also, that phonemic and phonetic representation do not take place in the same part of the brain. Scientists examined on a more precise level the speech production, on the level of the distinctive features of phonemes, to see in which features patients with aphasia made mistakes.

A distinctive feature describes the different manners and places of articulation. /t/ (like in touch) and /s/ (like in such) for example are created at the same place but produced in different manner. /t/ and /d/ are created at the same place and in the same manner but they differ in voicing.

Results show that in fluent as well as in non-fluent aphasia patients usually mix up only one distinctive feature, not two. In general it can be said that errors connected to the place of articulation are more common than those linked to voicing. Interestingly some aphasia patients are well aware of the different features of two phonemes, yet they are unable to produce the right sound. This suggests that though patients have great difficulty pronouncing words correctly, their comprehension of words is still quite good. This is characteristic for patients with Broca`s aphasia, while those with Wernicke`s aphasia show contrary symptoms: they are able to pronounce words correctly, but cannot understand what the words mean. That is why they often utter phonologically correct words (neologisms) that are not real words with a meaning.

Syntax

Syntax describes the rules of how words must be arranged to result in meaningful sentences. Humans in general usually know the syntax of their mother tongue and thus slip their tongue if a word happens to be out of order in a sentence. People with aphasia,

however, often have problems with parsing of sentences, not only with respect to the production of language but also with respect to comprehension of sentences. Patients showing an inability of comprehension and production of sentences usually have some kind of anterior aphasia, also called agrammatical aphasia. This can be revealed in tests with sentences. These patients cannot distinguish between active and passive voice easily if both agent and object could play an active part. For example patients do not see a difference between “The boy chased the girl” and “The boy was chased by the girl”, but they do understand both “The boy saw the apple” and “The apple was seen by the boy”, because they can seek help of semantics and do not have to rely on syntax alone. Patients with posterior aphasia, like for example Wernicke`s aphasia, do not show these symptoms, as their speech is fluent. Comprehension by mere syntactic means would be possible as well, but the semantic aspect must be considered as well. This will be discussed in the next part.

Semantics

Semantics deals with the meaning of words and sentences. It has been shown that patients suffering from posterior aphasia have severe problems understanding simple texts, although their knowledge of syntax is intact. The semantic shortcoming is often examined by a Token Test, a test in which patients have to point to objects referred to in simple sentences. As might have been guessed, people with anterior aphasia have no problems with semantics, yet they might not be able to understand longer sentences because the knowledge of syntax then is involved as well.

	anterior Aphasia (e.g. Broca)	**posterior Aphasia (e.g. Wernicke)**
Phonology	phonetic and phonemic representation affected	phonemic representation affected
Syntax	affected	no effect
Syntax	no effect	affected

Overview of the effects of aphasia from the psychological perspective

In general studies with lesioned people have shown that anterior areas are needed for speech output and posterior regions for speech comprehension. As mentioned above anterior regions are also more important for syntactic processing, while posterior regions are involved in semantic processing. But such a strict division of the parts of the brain and their responsibilities is not possible, because posterior regions must be important for more than just sentence comprehension, as patients with lesions in this area can neither comprehend nor produce any speech. (Barnich, 1997, pp. 283–293)

15.3.8.3 Evidence from Advanced Neuroscience Methods

Measuring the functions of both normal and damaged brains has been possible since the 1970s, when the first brain imaging techniques were developed. With them, we are able to “watch the brain working” while the subject is e.g. listening to a joke. These methods (further described in chapter 4) show whether the earlier findings are correct and precise.

Generally, imaging shows that certain functional brain regions are much smaller than estimated in brain lesion studies, and that their boundaries are more distinct (cf. Banich p. 294). The exact location varies individually, therefore bringing the results of many brain lesion studies together caused too big estimated functional regions before. For example, stimulating brain tissue electrically (during epilepsy surgery) and observing the outcome (e.g. errors in naming tasks) led to a much better knowledge where language processing areas are located.

PET studies (Fiez & Petersen, 1993, as cited in Banich, p. 295) have shown that in fact both anterior and posterior regions were activated in language comprehension and processing, but with different strengths – in agreement with the lesion studies. The more active speech production is required in experiments, the more frontal is the main activation: For example, when the presented words must be repeated.

Another result (Raichle et al. 1994, as referred to in Banich, p. 295) was that the familiarity of the stimuli plays a big role. When the subjects were presented well-known stimuli sets in well-known experimental tasks and had to repeat them, anterior regions were activated. Those regions were known to cause conduction aphasia when damaged. But when the words were new ones, and/or the subjects never had to do a task like this before, the activation was recorded more posterior. That means, when you repeat an

unexpected word, the heaviest working brain tissue is about somewhere under your upper left earlap, but when you knew this word that would be the next to repeat before, it is a bit nearer to your left eye.

15.3.9 Visual Language Processing

The processing of written language is performed when we are reading or writing and is thought to happen in a distinct neural processing unit than auditory language processing. Reading and writing respectively rely on vision whereas spoken language is first mediated by the auditory system. Language systems responsible for written language processing have to interact with a sensory system different from the one involved in spoken language processing.

Visual language processing in general begins when the visual forms of letters (“c” or “C” or “c”) are mapped onto abstract letter identities. These are then mapped onto a word form and the corresponding semantic representation (the “meaning” of the word, i.e. the concept behind it). Observations of patients that lost a language ability due to a brain damage led to different disease patterns that indicated a difference between perception (reading) and production (writing) of visual language just like it is found in non-visual language processing.

Alexic patients possess the ability to write while not being able to read whereas patients with agraphia are able to read but cannot write. Though alexia and agraphia often occur together as a result of damage to the angular gyrus, there were patients found having alexia without agraphia (e.g. Greenblatt 1973, as cited in M. T. Banich, “Neuropsychology“, p. 296) or having agraphia without alexia (e.g. Hécaen & Kremin, 1976, as cited in M. T. Banich, “Neuropsychology“, p. 296). This is a double dissociation that suggests separate neural control systems for reading and writing.

Since double dissociations are also found in phonological and surface dyslexia, experimental results support the theory that language production and perception respectively are subdivided into separate neural circuits. The two route model shows how these two neural circuits are believed to provide pathways from written words to thoughts and from thoughts to written words.

15.3.9.1 **Two routes model**

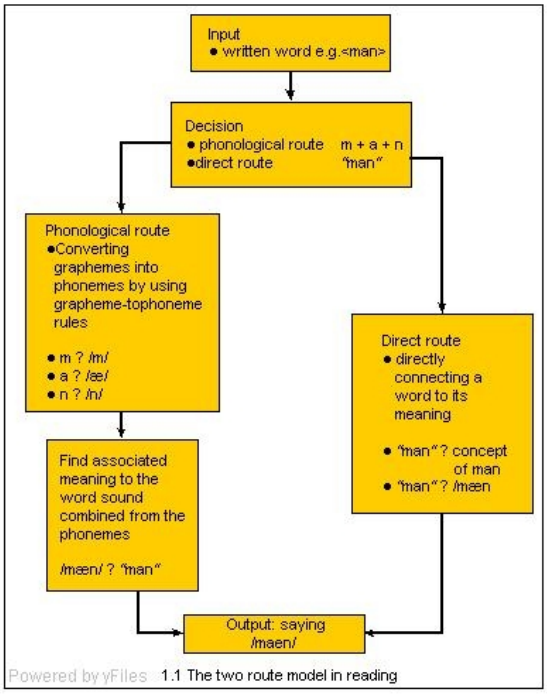


Figure **15.3.6**: Each route derives the meaning of a word or the word of a meaning in a different way, depending on word familiarity. (Image from Wikimedia Commons; File:1 1 twoRouteModelInReading.JPG; https://commons.wikimedia.org/wiki/F...lInReading.JPG; by TilmanATuos at en.wikibooks; licensed under *GNU Free Documentation License*.).

In essence, the two routes model contains two routes. Each of them derives the meaning of a word or the word of a meaning in a different way, depending on how familiar we are with the word.

Using the **phonological route** means having an intermediate step between perceiving and comprehending of written language. This intermediate step takes places when we are making use of grapheme-to-phoneme rules. Grapheme-to-phoneme rules are a way of determining the phonological representation for a given grapheme. A grapheme is the smallest written unit of a word (e.g. “sh” in “shore”) that represents a phoneme. A phoneme on the other hand is the smallest phonological unit of a word distinguishing it from another word that otherwise sounds the same (e.g. “bat” and “cat”). People learning to read or are encountering new words often use the phonological route to arrive at a meaning representation. They construct phonemes for each grapheme and then combine the individual phonemes to a sound pattern that is associated with a certain meaning (see 1.1).

The **direct route** is supposed to work without an intermediate phonological representation, so that print is directly associated with word-meaning. A situation in which the direct route has to be taken is when reading an irregular word like “colonel”. Application of grapheme-to-phoneme rules would lead to an incorrect phonological representation.

According to Taft (1982, as referred to in M. T. Banich,“Neuropsychology“, p. 297) and others the direct route is supposed to be faster than the phonological route since it does not make use of a “phonological detour” and is therefore said to be used for known words (see 1.1). However, this is just one point of view and others, like Chastain (1987, as referred to in M. T. Banich, “Neuropsychology“, p. 297), postulate a reliance on the phonological route even in skilled readers.

15.3.9.2 The processing of written language in reading

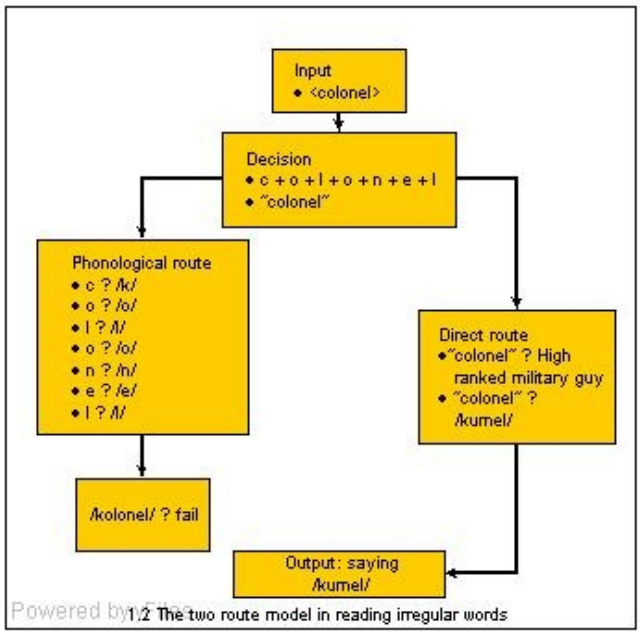


Figure 15.3.7: Regularity effects are common in cases of surface alexia. (Image from Wikimedia Commons; File:1 2 TwoRouteModelIrregularWords.JPG; https://commons.wikimedia.org/wiki/F...gularWords.JPG; by TilmanATuos at en.wikibooks; licensed under *GNU Free Documentation License*.).

Several kinds of alexia could be differentiated, often depending on whether the phonological or the direct route was impaired. Patients with brain lesions participated in experiments where they had to read out words and non-words as well as irregular words. Reading of non-words for example requires access to the phonological route since there cannot be a “stored” meaning or a sound representation for this combination of letters.

Patients with a lesion in temporal structures of the left hemisphere (the exact location varies) suffer from so called **surface alexia**. They show the following characteristic symptoms that suggest a strong reliance on the phonological route: Very common are regularity effects, that is a mispronunciation of words in which the spelling is irregular like "colonel" or "yacht" (see 1.2). These

words are pronounced according to grapheme-to-phoneme rules, although high-frequency irregularly spelled words may be preserved in some cases, the pronunciation according to the phonological route is just wrong.

Furthermore, the would-be pronunciation of a word is reflected in reading-comprehension errors. When asked to describe the meaning of the word “bear”, people suffering from surface alexia would answer something like “a beverage” because the resulting sound pattern of “bear” was the same for these people as that for “beer”. This characteristic goes along with a tendency to confuse homophones (words that sound the same but are spelled differently and have different meanings associated). However, these people are still able to read non-words with a regular spelling since they can apply grapheme-to-phoneme rules to them.

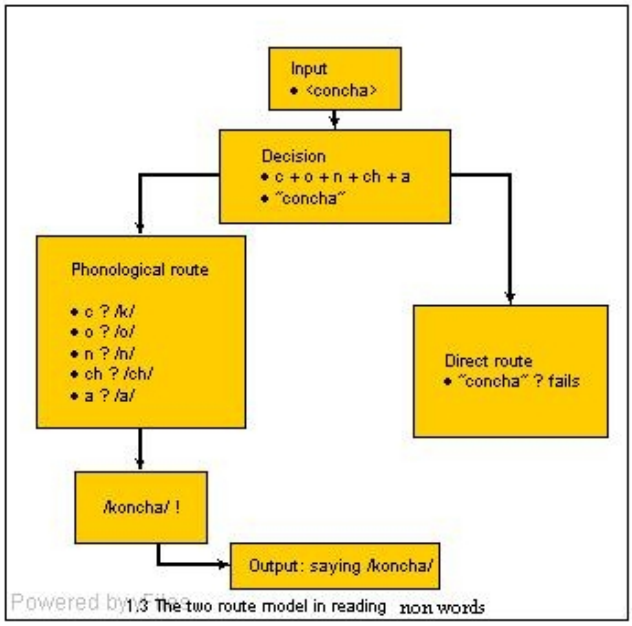


Figure 15.3.8: Patients with phonological alexia have to rely on the direct route. (Image from Wikimedia Commons; File:1 3 TwoRouteModelNonWords.JPG; https://commons.wikimedia.org/wiki/F...elNonWords.JPG; by TilmanATuos at en.wikibooks; licensed under *GNU Free Documentation License*.).

In contrast, **phonological alexia** is characterized by a disruption in the phonological route due to lesions in more posterior temporal structures of the left hemisphere. Patients can read familiar regular and irregular words by making use of stored information about the meaning associated with that particular visual form (so there is no regularity effect like in surface alexia). However, they are unable to process unknown words or non-words, since they have to rely on the direct route (see 1.3).

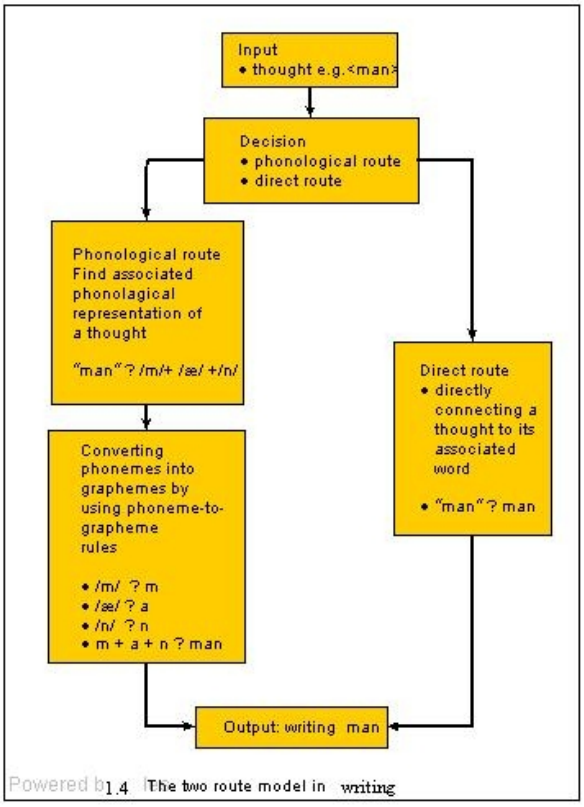
Word class effects and morphological errors are common, too. Nouns, for example, are read better than function words and sometimes even better than verbs. Affixes which do not change the grammatical class or meaning of a word (inflectional affixes) are often substituted (e.g. “farmer” instead of “farming”). Furthermore, concrete words are read with a lower error rate than abstract ones like “freedom” (concreteness effect).

Deep Alexia shares many symptomatic features with phonological alexia such as an inability to read out non-words. Just as in phonological alexia, patients make mistakes on word inflections as well as function words and show visually based errors on abstract words (“desire” → “desert”). In addition to that, people with deep alexia misread words as different words with a strongly related meaning (“woods” instead of “forest”), a phenomenon referred to as semantic paralexia. Coltheart (as referred to in the “Handbook of Neurolinguistics”, ch.41-3, p. 563) postulates that reading in deep dyslexia is mediated by the right hemisphere. He suggests that when large lesions affecting language abilities other than reading prevent access to the left hemisphere, the right-hemispheric language store is used. Lexical entries stored there are accessed and used as input to left-hemisphere output systems.

	Ability to read non-words	Ability to read irregular words	Ability to read regular words	Rely on	Disruption in	Other symptoms
Surface alexia	Ok	No	Ok	Phonological route	Direct route	-Confuse homophones - errors in word inflection
Phonological alexia	No	Ok	Ok	Direct route	Phonological route	-word class effects -morphological errors (e.g. farmer- farming) - concretness effect
Deep alexia	No	Ok	Ok	Direct route	Phonological route	-Confuse Homophones (e.g. pane -pain) -semantic paralexia (e.g. woods -forest)

Figure **15.3.9**: Overview of alexia. (Image from Wikimedia Commons; File:Overview alexia.JPG; https://commons.wikimedia.org/wiki/F...iew_alexia.JPG; by TilmanATuos at en.wikibooks; licensed under *GNU Free Documentation License*.).

15.3.9.3 The processing of written language in spelling



1.4 The two route model in writing

Figure **15.3.10**: The phonological route is supposed to make use of phoneme-to-grapheme rules while the direct route links thought to writing without an intermediary phonetic representation. (Image from Wikimedia Commons; File:1 4 TwoRoutesModelWritig.JPG; https://commons.wikimedia.org/wiki/F...odelWritig.JPG; by TilmanATuos at en.wikibooks; licensed under *GNU Free Documentation License*).

Just like in reading, two separate routes –a phonological and a direct route- are thought to exist. The phonological route is supposed to make use of phoneme-to-grapheme rules while the direct route links thought to writing without an intermediary phonetic representation (see 1.4).

It should be noted here that there is a difference between phoneme-to-grapheme rules (used for spelling) and grapheme-to-phoneme rules in that one is not simply the reverse of the other. In case of the grapheme “k” the most common phoneme for it is /k/. The most common grapheme for the phoneme /k/, however, is “c”. **Phonological agraphia** is caused by a **lesion in the left supramarginal gyrus**, which is located **in the parietal lobe above** the **posterior** section of the **Sylvian fissure** (M. T. Banich, “Neuropsychology“, p. 299). The ability to write regular and irregular words is preserved while the ability to write non-words is not. This, together with a poor retrieval of affixes (which are not stored lexically), indicates an inability to associate spoken words with their orthographic form via phoneme-to-grapheme rules. Patients rely on the direct route, which means that they use orthographic word-form representations that are stored in lexical memory. **Lesions at** the **conjunction of the posterior parietal lobe** and the parieto-occipital junction cause so called **lexical agraphia** that is sometimes also referred to as **surface agraphia**. As the name already indicates, it parallels surface alexia in that patients have difficulty to access lexical-orthographic representations of words. Lexical agraphia is characterized by a poor spelling of irregular words but good spelling for regular and non-words. When asked to spell irregular words, patients often commit regularization errors, so that the word is spelled phonologically correct (for example, “whisk” would be written as “wisque”).

	Ability to write non-words	Ability to write irregular words	Ability to write regular words	Rely on	Disruption in	Other symptoms
Phonological agraphia	No	Ok	Ok	Direct route	Phonological route	poor retrieval of affixes
Lexical agraphia	Ok	No	Ok	Phonological route	Direct route	

Figure **15.3.11**: Overview of agraphia. (Image from Wikimedia Commons; File:Overview agraphia.JPG; https://commons.wikimedia.org/wiki/F...w_agraphia.JPG; by TilmanATuos at en.wikibooks; licensed under *GNU Free Documentation License*).

15.3.9.4 Evidence from Advanced Neuroscience Methods

How can we find evidence for the theory of the two routes. Until now neuroscientific research is not able to ascertain that there are neural circuits representing a system like the one described above. The problem of finding evidence for visual language processing on two routes in contrast to one route (as stated by e.g. from Seidenberg & McClelland as referred to in M. T. Banich,"Neuropsychology", p. 308) is that it is not clear what characteristic brain activation would indicate that it is either happening on two or one routes. To investigate whether there are one or two systems, neuroimaging studies examine correlations between the activations of the **angular gyrus**, which is thought to be a crucial brain area in **written language processing**, and other brain regions. It was found out that during reading of non- words (which would strongly engage the phonological route) the activation is mostly correlated with brain regions which are involved in phonological processing e.g. **superior temporal regions** (BA 22) and **Broca’s area**. During reading of normal words (which would strongly engage the direct route) the highest activation was found in **occipital and ventral cortex**. That at least can imply that there are two distinct routes. However, these are conclusions drawn from highest correlations which do not ensure this suggestion. What neuroimaging studies do show is that the usage of a phonological and a direct route strongly overlap, which is rather unspectacular since it is quiet reasonable that fluent speakers mix both of the routes. Other studies additionally provide data in which the activated brain regions during reading of non-words and reading of normal words differ. ERP studies suggest that the left hemisphere possesses some sort of mechanism which responds to combinations of letters in a string, or to its orthography and/or to the phonological representation of the string. ERP waves differ, during early analysis of the visual form of the string, if the string represents a correct word or just pronounceable

nonsense (Posner & McCandliss, 1993 as referred in M.T. Banich, "Neuropsychology," p. 307-308). That indicates that this mechanism is sensitive to correct or incorrect words.

The opposite hemisphere, the **right hemisphere**, in contrast to the left hemisphere, is not involved in abstract mapping of word meaning but is rather responsible for **encoding word specific visual forms**. ERP and PET studies provide evidence that the right hemisphere responds in a stronger manner than the left hemisphere to letter-like strings. Moreover divided visual field studies reveal that the right hemisphere can better distinguish between different shapes of the same letter (e.g. in different handwritings) than the left hemisphere. The contribution of visual language processing by both hemispheres is that the **right hemisphere first recognizes** a **written word as letter sequences**, no matter how exactly they look like, **then the language network in the left hemisphere builds** up an **abstract representation** of the word, **which is the comprehension of the word**.

15.3.10 Other symbolic systems

Most neurolinguistic research is concerned with production and comprehension of English language, either written or spoken. However, looking at different language systems from a neuroscientific perspective can substantiate as well as differentiate acknowledged theories of language processing. The following section shows how neurological research of three symbolic systems, each different from English in some aspect, has made it possible to distinguish - at least to some extent - between brain regions that deal with the modality of the language (and therefore may vary from language to language, depending on whether the language in question is e.g. spoken or signed) from brain regions that seem to be necessary to language processing in general - regardless whether we are dealing with signed, spoken, or even musical language.

15.3.10.1 **Kana and Kanji**

Kana and Kanji are the two writing systems used parallel in the Japanese language. Since different approaches are used in them to represent words, studying Japanese patients with alexia is a great possibility to test the hypothesis about the existence of two different routes to meaning, discussed in the previous section.

The English writing system is phonological – each grapheme in written English roughly represents one speech sound – a consonant or a vowel. There are, however, other possible approaches to writing down a spoken language. In syllabic systems like the Japanese kana, one grapheme stands for one syllable. If written English were syllabic, it could e.g. (for example) include a symbol for the syllable “nut”, appearing both in the words “donut” and “peanut”. Syllabic systems are sound-based – since the graphemes represent units of spoken words rather than meaning directly, an auditory representation of the word has to be created in order to arrive at the meaning. Therefore, reading of syllabic systems should require an intact phonological route. In addition to kana, Japanese also use a logographic writing system called kanji, in which one grapheme represents a whole word or a concept. Different from phonological and syllabic systems, logographic systems don’t comprise systematical relationships between visual forms and the way they’re pronounced – instead, visual form is directly associated with the pronunciation and meaning of the corresponding word. Reading kanji should therefore require the direct route to meaning to be intact.

The hypothesis about the existence of two different routes to meaning has been confirmed by the fact that after brain damage, there can be a double dissociation between kana and kanji. Some Japanese patients can thus read kana but not kanji (surface alexia), whereas other can read kanji but not kana (phonological alexia). In addition, there is evidence that different brain regions of Japanese native speakers are active while reading kana and kanji, although like in the case of English native speakers, these regions also overlap.

Since the distinction between direct and phonological route also makes sense in case of Japanese, it may be a general principle common to all written languages that reading them relies on two independent (at least partially) systems, both using different

strategies to catch the meaning of a written word – either associating the visual form directly with the meaning (the direct route), or using the auditory representation as an intermediary between the visual form and the meaning of the word (the phonological route).



The Japanese Kana sign for the syllable "mu"



The Japanese Kanji sign for the concept "Book", "writing", or "calligraphy

Figure **15.3.12**: Two forms of Japanese writing elicit activity in different regions of the brain in native speakers.

15.3.10.2 Sign Language

From a linguistic perspective, sign languages share many features of spoken languages – there are many regionally bounded sign languages, each with a distinct grammar and lexicon. Since at the same time, sign languages differ from spoken languages in the way the words are “uttered”, i.e. in the modality, neuroscientific research in them can yield valuable insights into the question whether there are general neural mechanisms dealing with language, regardless of its modality.

Structure of SL

Sign languages are phonological languages - every meaningful sign consists of several phonemes (phonemes used to be called cheremes (Greek χερι: hand) until their cognitive equivalence to phonemes in spoken languages was realized) that carry no meaning as such, but are nevertheless important to distinguish the meaning of the sign. One distinctive feature of SL phonemes is the place of articulation – one hand shape can have different meanings depending on whether it’s produced at the eye-, nose-, or chin-level. Other features determining the meaning of a sign are hand shape, palm orientation, movement, and non-manual markers (e.g. facial expressions).

To express syntactic relationships, Sign Languages exploit the advantages of the visuo-spatial medium in which the signs are produced – the syntactic structure of sign languages therefore often differs from that of spoken languages. Two important features of most sign language's grammars (including American Sign Language (ASL), Deutsche Gebärdensprache (DGS) and several other major sign languages) are directionality and simultaneous encoding of elements of information:

- Directionality

The direction in which the sign is made often determines the subject and the object of a sentence. Nouns in SL can be 'linked' to a particular point in space, and later in the discourse they can be referred to by pointing to that same spot again (this is functionally related to pronouns in English). The object and the subject can then be switched by changing the direction in which the sign for a transitive verb is made.

- Simultaneous encoding of elements of information

The visual medium also makes it possible to encode several pieces of information simultaneously. Consider e.g. the sentence "The flight was long and I didn't enjoy it". In English, the information about the duration and unpleasantness of the flight have to be encoded sequentially by adding more words to the sentence. To enrich the utterance "The flight was long” with the information about the unpleasantness of the flight, another sentence (“I did not enjoy it") has to be added to the original one. So, in order to convey more information, the length of the original sentence must grow. In sign language, however, the increase of information in an utterance doesn’t necessarily increase the length of the utterance. To convey information about the unpleasantness of a long flight experienced in the past, one can just use the single sign for "flight" with the past tense marker, moved in a way that represents the attribute "long", combined with the facial expression of disaffection. Since all these features are signed simultaneously, no additional time is needed to utter "The flight was long" as compared to "The flight was long and I didn't enjoy it".

Neurology of SL

Since sentences in SL are encoded visually, and since its grammar is often based on visual rather than sequential relationships among different signs, it could be suggested that the processing of SL mainly depends on the right hemisphere, which is mainly concerned with the performance on visual and spatial tasks. However, there is evidence suggesting that processing of SL and spoken language might be equally dependent on the left hemisphere, i.e. that the same basic neural mechanism may be responsible for all language functioning, regardless of its modality (i.e. whether the language is spoken or signed).

The importance of the left hemisphere in SL processing indicated e.g. by the fact that signers with a damaged right hemisphere may not be aphasiacs, whereas as in case of hearing subjects, lesions in the left hemisphere of signers can result in subtle linguistic difficulties (Gordon, 2003). Furthermore, studies of aphasic native signers have shown that **damage to anterior portions of the left hemisphere (Broca's area) result in a syndrome similar to Broca's aphasia** – the patients lose fluency of communication, they aren't able to correctly use syntactic markers and inflect verbs, although the words they sign are semantically appropriate. In contrast, patients with **damages to posterior portions of the superior temporal gyrus (Wernicke's area)** can still properly inflect verbs, set up and retrieve nouns from a discourse locus, but **the sequences they sign have no meaning** (Poizner, Klima & Bellugi, 1987). So, **like in the case of spoken languages, anterior and posterior portions of the left hemisphere seem to be responsible for the syntax and semantics of the language respectively.** Hence, it's not essential for the "syntax processing mechanisms" of the brain whether the syntax is conveyed simultaneously through spatial markers or successively through word order and morphemes added to words - the **same underlying mechanisms might be responsible for syntax in both cases**.

Further evidence for the same underlying mechanisms for spoken and signed languages comes from studies in which fMRI has been used to compare the language processing of:

- 1. congenitally deaf native signers of British Sign Language,
- 2. hearing native signers of BSL (usually hearing children of deaf parents)
- 3. hearing signers who have learned BSL after puberty
- 4. non-signing subjects

Investigating language processing in these different groups allows making some distinctions between different factors influencing language organization in the brain - e.g. to what amount does deafness influences the organization of language in the brain as compared to just having SL as a first language(1 vs. 2), or to what amount does learning of SL as a first language differ from learning SL as native language (1,2 vs.3), or to what amount is language organized in speakers as compared to signers (1,2,3 vs.4).

These studies have shown that **typical areas in the left hemisphere** are activated in both native English speakers given written stimuli and native signers given signs as stimuli. Moreover, there are also areas that are equally activated both in case of deaf subjects processing sign language and hearing subjects processing spoken language – a finding which **suggests that these areas constitute the core language system regardless of the language modality** (Gordon, 2003).

Different from speakers, however, signers also show a strong activation of the right hemisphere. This is partly due to the necessity to process visuo-spatial information. Some of those areas, however (e.g. the **angular gyrus**) are only activated in native signers and not in hearing subjects that learned SL after puberty. This **suggests that the way of learning sign languages (and languages in general) changes with time:** Late learner's brains are unable to recruit certain brain regions specialized for processing this language (Newman et al., 1998).]

We have seen that evidence from aphasias as well as from neuroimaging suggest the **same underlying neural mechanisms to be responsible for sign and spoken languages.** It 's natural to ask whether these neural mechanisms are even more general, i.e. whether they are able to process any type of symbolic system underlying some syntax and semantics. One example of this kind of more general symbolic system is music.

15.3.10.3 Music

Like language, music is a human universal involving some combinatorial principles that govern the organizing of discrete elements (tones) into structures (phrases) that convey some meaning – music is a symbolic system with a special kind of **syntax and semantics**. It's therefore interesting to ask whether music and natural language share some neural mechanisms: whether processing of music is dependent on processing of language or the other way round, or whether the underlying mechanisms underlying them are completely separate. By investigating the neural mechanisms underlying music we might find out whether the

neural processes behind language are unique to the domain of natural language, i.e. whether language is modular. Up to now, research in the neurobiology of music has yielded contradicting evidence regarding these questions.

On the one hand, there is evidence that there is a double **dissociation of language and music abilities**. People suffering from **amusia** are unable to perceive harmony, to remember and to recognize even very simple melodies; at the same time they have **no problems in comprehending or producing speech**. There is even a case of a patient who developed amusia without aprosodia, i.e. although she couldn't recognize tone in musical sequences, she nevertheless could still make use of pitch, loudness, rate, or rhythm to convey meanings in spoken language (Pearce, 2005). This highly selective **problem in processing music (amusia)** can occur as a result of **brain damage**, or be inborn; in some cases it runs on families, suggesting a **genetic** component. The complement syndrome of amusia also exists – after suffering a **brain damage in the left hemisphere**, the Russian composer Shebalin **lost** his **speech** functions, **but** his **musical abilities remained** intact (Zatorre, McGill, 2005).

On the other hand, **neuroimaging** data suggest that **language and music** have a **common mechanism for processing syntactical structures**. The P600 ERP`s in the Broca area, measured as a response to ungrammatical sentences, is also elicited in subjects listening to musical chord sequences lacking harmony (Patel, 2003) – the expectation of typical sequences in music could therefore be mediated by the same neural mechanisms as the expectation of grammatical sequences in language.

A possible solution to this apparent contradiction is the dual system approach (Patel, 2003) according to which music and language share some procedural mechanisms (frontal brain areas) responsible for processing the general aspects of syntax, but in both cases these mechanisms operate on different representations (posterior brain areas) – notes in case of music and words in case of language.

15.3.11 Summary

Many questions are yet to be answered. It is still unclear whether there is a distinct language module (that you could cut out without affecting anything in other brain functions) or not. As Evely C. Ferstl points out in her review, the next step after exploring distinct small regions responsible for subtasks of language processing will be to find out how they work together and build up the language network.

15.3.12 References & Further Reading

Books

- Brigitte Stemmer, Harry A. Whitaker. Handbook of Neurolinguistics. Academic Press (1998). ISBN 0126660557
- Marie T. Banich: Neuropsychology. The neural bases of mental function (1997).
- Ewa Dąbrowska: Language, Mind and Brain. Edinburgh University press Ltd.(2004)
- *a review:* Evelyn C. Ferstl, The functional neuroanatomy of text comprehension. What's the story so far?" from: Schmalhofer, F. & Perfetti, C. A. (Eds.), Higher Level Language Processes in the Brain:Inference and Comprehension Processes. Lawrence Erlbaum. (2004)
- Müller,H.M.& Rickert,G. (Hrsg.): Neurokognition der Sprache. Stauffenberg Verlag (2003)
- Poizner, Klima & Bellugi: What the hands reveal about the brain. MIT Press (1987)
- N. Chomsky: Aspects of the Theory of Syntax. MIT Press (1965). ISBN 0262530074
- Neville & Bavelier: Variability in the effects of experience on the development of cerebral specializations: Insights from the study of deaf individuals. Washington, D.C.: US Government Printing Office (1998)
- Newman et al.: Effects of Age of Acquisition on Cortical Organization for American Sign Language: an fMRI Study. NeuroImage, 7(4), part 2 (1998)

Links - english

- Robert A. Mason and Marcel Adam Just: How the Brain Processes Causal Inferences in Text
- Neal J. Pearlmutter and Aurora Alma Mendelsohn: Serial versus Parallel Sentence Comprehension
- Brain Processes of Relating a Statement to a Previously Read Text: Memory Resonance and Situational Constructions
- Clahsen, Harald: Lexical Entries and Rules of Language: A Multidisciplinary Study of German Inflection.
- Cherney, Leora (2001): Right Hemisphere Brain Damage
- Grodzinsky, Yosef (2000): The neurology of syntax: Language use without Broca's area.

- Müller, H.M. & Kutas, M. (1996). What's in a name? Electrophysiological differences between spoken nouns, proper names and one's own name. NeuroReport 8:221-225.
- Müller, H. M., King, J. W. & Kutas, M. (1997). Event-related potentials elicited by spoken relative clauses Cognitive Brain Research 4:193-203.

15.3.13 Attributions

"An Alternative to the Wernicke-Geschwind Model" and "Other Models" adapted by Kenneth A. Koenigshofer, Ph.D., Chaffey College, from Broca's Area, Wernicke's Area, and other Language-Processing Areas in the Brain by Bruno Dubuc under a Copyleft license.

"Lateralization, Auditory and Visual Language Processing" adapted by Kenneth A. Koenigshofer, Ph.D., Chaffey College, from Cognitive Psychology and Cognitive Neuroscience, Wikibooks. Neuroscience of Text Comprehension; https://en.wikibooks.org/wiki/Cognit..._Comprehension; available under the Creative Commons Attribution-ShareAlike License.

This page titled 15.3: Other Brain Models of Spoken and Written Language is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

15.4: Handedness, Language, and Brain Lateralization

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe hemispheric asymmetries
2. Describe the role of testosterone and stress in brain lateralization and handedness
3. Describe sex differences in brain lateralization and suggest an explanation
4. Describe lateralization in other species and briefly discuss implications for possible origins of human brain lateralization

Handedness, testosterone, and lateralization for language appear to be related. Approximately 10% of people are left-handed and 2/3 of left-handers are male. 95% of right handed men and women have language lateralized to the left hemisphere. The greatest asymmetries are in the posterior language areas including the temporal planum and angular gyrus. If the left hemisphere is damaged or defective prenatally the right hemisphere can acquire language functions.

15.4.1 Handedness, Language, and Lateralization

The brain's anatomical asymmetry, its lateralization for language, and the phenomenon of handedness are all clearly interrelated, but their influences on one another are complex. Though about 90% of people are right-handed, and about 95% of right-handers have their language areas on the left side of their brains, that still leaves 5% of right-handers who are either right-lateralized for language or have their language areas distributed between their two hemispheres. And then there are the left-handers, among whom all of these patterns can be found , including left-lateralization.

Some scientists suggest that the left hemisphere's dominance for language evolved from this hemisphere's better control over the right hand. The circuits controlling this "skillful hand" may have evolved so as to take control over the motor circuits involved in language. Broca's area, in particular, is basically a premotor module of the neocortex and co-ordinates muscle contraction patterns that are related to other things besides language.

Brain-imaging studies have shown that several structures involved in language processing are larger in the left hemisphere than in the right. For instance, Broca's area in the left frontal lobe is larger than the homologous area in the right hemisphere. But the **greatest asymmetries** are found mainly in the **posterior language areas**, such as the **temporal planum and the angular gyrus**.

Two other notable asymmetries are the larger protrusions of the frontal lobe on the right side and the occipital lobe on the left. These protrusions might, however, be due to a slight rotation of the hemispheres (counterclockwise, as seen from above) rather than to a difference in the volume of these areas. These protrusions are known as the **right-frontal and left-occipital petalias** ("petalias" originally referred to the indentations that these protrusions make on the inside of of the skull).

The **structures involved in producing and understanding language** seem to be laid down in accordance with **genetic instructions** that come into play as **neuronal migration** proceeds in the **human embryo**. Nevertheless, the two hemispheres can remain just about equipotent until language acquisition occurs. Normally, the language specialization develops in the left hemisphere, which matures slightly earlier. The earlier, more intense activity of the neurons in the left hemisphere would then lead both to right-handedness and to the control of language functions by this hemisphere. But **if the left hemisphere is damaged or defective**, **language** can be **acquired by** the **right hemisphere**. An **excess** of **testosterone** in newborns due to stress at the time of birth might well be one of the most common causes of **slower development in the left hemisphere** resulting in **greater participation by the right.**

This hypothesis of a central role for testosterone is supported by experiments which showed that **in rats, cortical asymmetry** is **altered** if the rodents are injected with **testosterone at birth**. This hormonal hypothesis would also explain why **two-thirds of all left-handed persons are males**.

Interindividual variations, which are essential for natural selection, are expressed in various ways in the **human brain**. **Volume and weight can vary by a factor of two** or even more. The **brain's vascular structures are extremely variable**; the deficit

caused by an obstruction at a given point in the vascular system can vary greatly from one individual to another. At the macroscopic anatomical level, the **folds and grooves in the brain also vary tremendously from individual to individual**, especially in the areas associated with language . **Variability in** the language areas can also be observed at the microscopic level, for example, in the **synaptic structure** of the **neurons in Wernicke's area.**

Interindividual variability is also expressed in the brain's functional organization, and particularly in the phenomenon of hemispheric asymmetry. For instance, some data indicate that **language functions may be more bilateral in women than in men**. The percentage of **atypical lateralization for language** also varies with handedness: it is considerably **higher among left-handers** than among right-handers.

Lastly, as if all this were not enough, there is also such a thing as intraindividual variability. In the same individual, a given mental task can sometimes activate different neuronal assemblies in different circumstances—for instance, when the individual is performing this task for the first time, as opposed to when he or she has already performed it many times before.

Even in many species that are quite distant from humans in evolutionary terms (frogs, for example), the brain is left-lateralized for the vocalization function.

In chimpanzees, lateralization for the anatomical areas corresponding to Broca's and Wernicke's areas already exists, even though it does not yet correspond to the language function. And like the majority of humans, the majority of chimpanzees use their right hand in preference to their left.

These asymmetries in the other primates represent persuasive evidence of the **ancient phylogenetic origin of lateralization** in the human brain. The expansion of the prefrontal cortex in humans might in part reflect its role in the production of language.

An asymmetrical lateralized brain may have evolved in order to reduce redundancy of information processing functions in favor of greater processing capacity in a brain whose volume is limited by the size of the human female birth canal (Corballis, 2017). To increase processing capacity if brain volume is fixed, lateralization of function may more efficiently utilize the processing resources available.

Women have the reputation of being able to talk and listen while doing all sorts of things at the same time, whereas men supposedly prefer to talk or hear about various things in succession rather than simultaneously. Brain-imaging studies may now have revealed an anatomical substrate for this behavioural difference, by demonstrating that language functions tend to place more demands on both hemispheres in women while being more lateralized (and mainly left-lateralized) in men. Women also have more nerve fibers connecting the two hemispheres of their brains (via a broader corpus callosum), which also suggests that more information is exchanged between them.

15.4.2 Summary

Most people are right-handed and language in most people is lateralized to the left hemisphere. Lateralization of the brain's functions may have been one way to increase the brain's processing capacity by reducing redundant usage of available processing resources. Vocalization is left-lateralized in species ranging from frogs to chimpanzees, suggesting ancient evolutionary origins of brain lateralization in humans.

15.4.3 References

Corballis, M. C. (2017). The evolution of lateralized brain circuits. *Frontiers in psychology*, *8*, 1021.

15.4.4 Attributions

"Handedness, Language, and Lateralization" adapted by Kenneth A. Koenigshofer, Ph.D., Chaffey College, from Broca's Area, Wernicke's Area, and other Language-Processing Areas in the Brain by Bruno Dubuc in The Brain from Top to Bottom, under a Copyleft license.

This page titled 15.4: Handedness, Language, and Brain Lateralization is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

15.5: The Right Hemisphere's Contribution to Language

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe the cause and symptoms of hemineglect (unilateral neglect)
2. Describe the primary contributions of the right hemisphere to human language expression and comprehension
3. Identify and describe language pragmatics and its disorders
4. Describe theory of mind and how it is related to autism

15.5.1 Overview

Hemineglect, also known as unilateral neglect, following damage in the right parietal cortex is characterized by inability to attend to sensory inputs on the left side of the body leading to lack of awareness and emotional indifference to these inputs. Because of the disorder, such patients can "lose track" of the left side of their body and limbs because of the lack of attention to inputs from the left side of space including the left side of their own bodies. Some of these patients may fail to understand that the left side of their bodies belong to them. Right hemisphere damage can disrupt the emotional and contextual aspects of language use, suggesting that the right hemisphere is more emotional than the left and that normally the right hemisphere contributes the emotional aspects of human speech.

15.5.2 The Role of the Right Hemisphere in Language

To follow a conversation, a written document, or an exchange of witticisms, you must be able not only to understand the syntax of sentences and the meanings of words, but also to interrelate multiple elements and interpret them with respect to a given context. While various types of damage to the left hemisphere produce the many documented forms of aphasia, right hemisphere damage (RHD) causes a variety of communication deficits involving the interpretation of context. These deficits can be divided into two main categories.

The first category of RHD-induced deficits affect communication indirectly, by disrupting people's ability to interact effectively with their environment. One example of a deficit that can be caused by RHD is **hemineglect (unilateral neglect)**, in which an individual pays no attention to stimuli presented to the various sensory modalities on the left side of the body. The individual may also suffer from **anosognosia**: unawareness of such deficits. For instance, some people who have damage just posterior to the central sulcus in their **parietal lobe** in their **right** hemispheres cannot even recognize certain parts of their own bodies as being their own. Thus this type of RHD produces a kind of indifference that is the opposite of the minimum emotional investment required to establish harmonious communication.

The other major family of RHD-induced deficits affect communication and cognition directly. These deficits can be grouped under the heading of pragmatic communication disorders, pragmatics being the discipline that studies the relationships between language and the way that people use it in context. **Pragmatic disorders can be subdivided into disorders in prosody, discourse organization, and understanding of non-literal language.**

Prosody refers to the **intonation and stress** with which the phonemes of a language are **pronounced**. People with **aprosodia—RHD** that **impairs** their use of **prosody—cannot use intonation and stress to effectively express** the **emotions** they actually feel. As a result, they speak and behave in a way that seems flat and emotionless.

The **second category of pragmatic communication disorders** that can be caused by RHD affect the organization of discourse according to the rules that govern its construction. In some individuals, these disorders take the form of a **reduced ability to interpret** the signs that establish the **context** for a communication, or the nuances conveyed by certain words, **or the speaker's intentions or body language**, or the applicable **social conventions**. With regard to social conventions, for example, people generally do not address their boss the same way they would their brother, but people with certain kinds of RHD have difficulty in making this distinction.

Last but not least among the **types of pragmatic communication disorders caused by RHD** are **disorders** in the **understanding of non-literal language**. It is estimated that fewer than half of the sentences that we speak express our meaning literally, or at least they do not do so entirely. For instance, whenever we use **irony, or metaphors, or other forms of indirect language**, people's ability to understand our actual meaning depends on their ability to interpret our intentions.

To understand irony, for example, people must apply two levels of awareness, just as they must do to understand jokes. First, they must understand the speaker's state of mind, and second, they **must understand the speaker's intentions** as to how his or her words should be construed. Someone who is telling a joke wants these words not to be taken seriously, while someone who is speaking ironically wants the listener to perceive their actual meaning as the opposite of their literal one.

Metaphors too express an intention that belies a literal interpretation of the words concerned. If a student turns to a classmate and says "This prof is a real sleeping pill", the classmate will understand the implicit analogy between the pill and the prof and realize that the other student finds this prof boring. But someone with RHD that affects their understanding of non-literal language might not get this message.

Lastly, the various indirect ways that we commonly use language in everyday life can cause problems for people with RHD. In such cases, the speaker's actual intention underlies their oral statement as such. For example, someone who says "I wonder what the time is now " is indirectly asking for someone to tell them the time, but a person with RHD may not understand that.

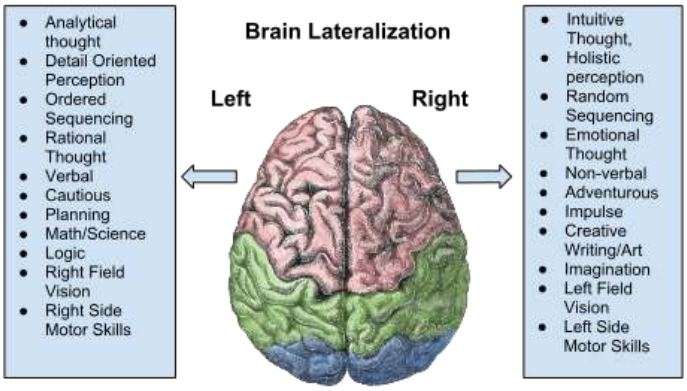


Figure 15.5.1: The oversimplification of lateralization in pop psychology. This belief was widely held even in the scientific community for some years. The left brain controls functions that have to do with logic and reason, while the right brain controls functions involving creativity and emotion. This simplified view of brain lateralization is no longer considered accurate by neuroscientists. Instead, new research using brain imaging shows that the two halves of the brain work together much more than earlier research had implied. (Image and first two sentences of caption from Wikimedia Commons, remainer of caption by Kenneth A. Koenigshofer, Ph.D.; File:Brain Lateralization.svg; https://commons.wikimedia.org/wiki/F...ralization.svg; by Chickensaresocute; licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license).

Though the **left hemisphere** is still regarded as the **dominant hemisphere for language**, the role of the **right hemisphere** in understanding the **context** in which language is used is now well established. We know that in the absence of the left hemisphere (for example, when Wada's test is performed which temporarily inactivates targeted brain tissue), the **right hemisphere can produce some rudimentary language.** But lesion studies have shown that the right hemisphere's role in language appears to be far wider—so much so that it is now more accurate to think of the **two hemispheres' language specializations** not as separate functions, but rather as a variety of abilities that **operate in parallel** and whose interaction makes human language in all its complexity possible.

Many theories have been offered to explain people's ability to adapt their use of language to the interpersonal context. One of these is the theory of mind. According to Premack and Woodruff (1978), the **theory of mind is the ability that lets people ascribe mental processes to other people**, to reason on the basis of these ascribed processes, and to understand the behaviors that arise from them. Premack and Woodruff were the first authors to use the term "theory of mind". They did so in a study on the ability of chimpanzees to ascribe beliefs and intentions to human beings. Since the time of this study, the theory of mind has been applied mainly in studies comparing the cognitive development of normal children and **autistic children**, because the latter represent a

population that is known to display **deficits in social reasoning** from the very earliest age. In normal children, theory of mind (ToM) develops between 3 and 4 years of age and is fully developed by 5 years of age (Roth & Dicke, 2012).

When experimental subjects are asked to **identify the emotional content of recorded sentences** that are played back into only one of their ears, they perform **better if these sentences are played into their left ear (which sends them to the right hemisphere)** than into their right (which sends them to the left hemisphere). These results confirm that the right hemisphere has a role in processing the emotional content of speech.

15.5.3 Summary

Hemineglect, also known as unilateral neglect, caused by damage to the right parietal lobe, is a disorder of attention to sensory inputs from the left side of space including the left side of one's own body. Language in patients with this disorder, reflects indifference to sensory inputs from the left half of space. The right hemisphere is involved in understanding the speaker's intentions and the emotional components of language. Language pragmatics including prosody and indirect language appear to be at least in part dependent upon processing by the right hemisphere. Some researchers suggest that damage to the right hemisphere may interfere with theory of mind, the ability to understand that others have minds, beliefs and intentions, and that persons with autism may have impairments in theory of mind.

15.5.4 References

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and brain sciences*, *1*(4), 515-526.

Roth, G., & Dicke, U. (2012). Evolution of the brain and intelligence in primates. *Progress in brain research*, *195*, 413-430.

15.5.5 Attributions

"The Right Hemisphere's Contribution to Language," adapted from Broca's Area, Wernicke's Area, and other Language-Processing Areas in the Brain by Bruno Dubuc, in The Brain from Top to Bottom, under a Copyleft license.

This page titled 15.5: The Right Hemisphere's Contribution to Language is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

CHAPTER OVERVIEW

16: Emotion and Stress

> This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

16.1: The Neurological Bases of Emotions
16.2: Valence of Emotions
16.3: Fight or Flight
16.4: Social Neuroscience
16.5: Stress and health

This page titled 16: Emotion and Stress is shared under a mixed license and was authored, remixed, and/or curated by ASCCC OERI & Bakhtawar Bhadha (ASCCC Open Educational Resources Initiative (OERI)) .

16.1: The Neurological Bases of Emotions

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Identify what general and specific brain parts are involved in emotions.
- Analyze the anatomical and chemical bases of basic emotions.
- Analyze the biopsychological methods of studying emotions.
- Evaluate the functions of emotions.

16.1.1 Overview

There is a strong connection between brain/body and emotions/affective states. Human and animal research findings have shown brain networks and associated neurotransmitters involved in basic emotion/affective systems.

16.1.2 **Affective Neuroscience: What is it?**

Affective neuroscience examines how the brain creates emotional responses. Emotions are psychological phenomena that involve changes to the body (e.g., facial expression), changes in autonomic nervous system activity, feeling states (subjective responses), and urges to act in specific ways (motivations; Izard, 2010). Affective **neuroscience** aims to understand how matter (brain structures and chemicals) creates one of the most fascinating aspects of the mind, the emotions. Affective neuroscience uses unbiased, observable measures that provide credible evidence to other sciences and laypersons on the importance of emotions. It also leads to biologically based treatments for affective disorders (e.g., depression).



Figure **16.1.1**: Although we all experience emotions all the time, they are very difficult to describe and study. Fortunately, technological advances are making this easier. (Image: Waag Society, https://goo.gl/F0KdnB, CC BY-NC-SA 2.0, goo.gl/iF4hmM)

The human brain and its responses, including emotions, are complex and flexible. In comparison, nonhuman animals possess simpler nervous systems and more basic emotional responses. Invasive neuroscience techniques, such as electrode implantation, lesioning, and hormone administration, can be more easily used in animals than in humans. Human neuroscience must rely primarily on noninvasive techniques such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI),

and on studies of individuals with brain lesions caused by accident or disease. Thus, animal research provides useful models for understanding affective processes in humans. Affective circuits found in other species, particularly social mammals such as rats, dogs, and monkeys, function similarly to human affective networks, although nonhuman animals' brains are more basic.

In humans, emotions and their associated neural systems have additional layers of complexity and flexibility. Compared to animals, humans experience a vast variety of nuanced and sometimes conflicting emotions. Humans also respond to these emotions in complex ways, such that conscious goals, values, and other cognitions influence behavior in addition to emotional responses.

Across species, emotional responses are organized around the organism's survival and reproductive needs. Emotions influence perception, cognition, and behavior to help organisms survive and thrive (Farb, Chapman, & Anderson, 2013). Networks of structures in the brain respond to different needs, with some overlap between different emotions. Specific emotions are not located in a single structure of the brain. Instead, emotional responses involve networks of activation, with many parts of the brain activated during any emotional process. In fact, the brain circuits involved in emotional reactions include nearly the entire brain (Berridge & Kringelbach, 2013). Brain circuits located deep within the brain below the cerebral cortex are primarily responsible for generating basic emotions (Berridge & Kringelbach, 2013; Panksepp & Biven, 2012). In the past, research attention was focused on specific brain structures that will be reviewed here, but future research may find that additional areas of the brain are also important in these processes.

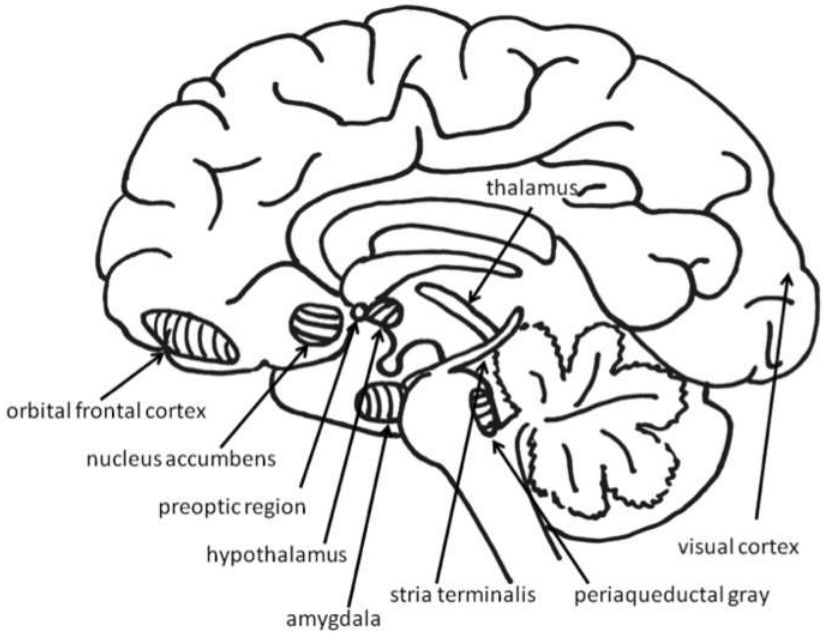


Figure **16.1.2**: Structures in the brain - orbital frontal cortex, nucleus accumbens, preoptic region, hypothalamus, amygdala, stria terminalis, PAG, visual cortex and thalamus play a role in emotion.

16.1.3 Basic Emotions

16.1.3.1 Desire: The neural systems of reward seeking

One of the most important affective neuronal systems relates to feelings of desire, or the appetite for rewards. Researchers refer to these appetitive processes using terms such as "wanting" (Berridge & Kringelbach, 2008), "seeking" (Panksepp & Biven, 2012), or "behavioural activation sensitivity" (Gray, 1987). When the appetitive system is aroused, the organism shows enthusiasm, interest, and curiosity. These neural circuits motivate the animal to move through its environment in search of rewards such as appetizing foods, attractive sex partners, and other pleasurable stimuli. When the appetitive system is underaroused, the organism appears depressed and helpless.

Much evidence for the structures involved in this system comes from animal research using direct brain stimulation. When an electrode is implanted in the lateral **hypothalamus** or in cortical or mesencephalic regions to which the hypothalamus is connected,

animals will press a lever to deliver electrical stimulation, suggesting that they find the stimulation pleasurable. The regions in the desire system also include the amygdala, nucleus accumbens, and **frontal cortex** (Panksepp & Biven, 2012). The neurotransmitter dopamine, produced in the mesolimbic and mesocortical dopamine circuits, activates these regions. It creates a sense of excitement, meaningfulness, and anticipation. These structures are also sensitive to drugs such as cocaine and amphetamines because they are dopamine agonists (Panksepp & Biven, 2012).



Figure **16.1.3**: Just looking at an image of appealing food should increase the activity in your left frontal cortex. Yum! (CC0 Public Domain, goo.gl/m25gce)

Research in both humans and nonhuman animals shows that the left frontal cortex (compared to the right frontal cortex) is more active during appetitive emotions such as desire and interest. Researchers first noted that persons who had suffered damage to the left frontal cortex developed depression, whereas those with damage to the right frontal cortex developed mania (Goldstein, 1939). The relationship between left frontal activation and approach-related emotions (that is those emotions that involve a movement toward the stimulus such as with happiness and also with anger) has been confirmed in healthy individuals using EEG and fMRI (Berkman & Lieberman, 2010). For example, increased left frontal activation occurs in 2- to 3-day-old infants when sucrose is placed on their tongues (Fox & Davidson, 1986), and in hungry adults as they view pictures of desirable desserts (Gable & Harmon-Jones, 2008). In addition, greater left frontal activity in appetitive situations has been found to relate to dopamine (Wacker, Mueller, Pizzagalli, Hennig, & Stemmler, 2013).

16.1.3.2 “Liking”: The Neural Circuits of Pleasure and Enjoyment

Surprisingly, the amount of desire an individual feels toward a reward need not correspond to how much he or she likes that reward. This is because the neural structures involved in the enjoyment of rewards are different from the structures involved in the desire for the rewards. “Liking” (e.g., enjoyment of a sweet liquid) can be measured in babies and nonhuman animals by measuring licking speed, tongue protrusions, and happy facial expressions, whereas “wanting” (desire) is shown by the willingness to work hard to obtain a reward (Berridge & Kringelbach, 2008). Liking has been distinguished from wanting in research on topics such as drug abuse. For example, drug addicts often desire drugs even when they know that the ones available will not provide pleasure (Stewart, de Wit, & Eikelboom, 1984).

Research on liking has focused on a small area within the **nucleus accumbens** and on the posterior half of the ventral pallidum. These brain regions are sensitive to opioids and endocannabinoids (endogenously produced substances that have effects similar to marijuana). Stimulation of other regions of the reward system increases wanting, but does not increase liking, and in some cases even decreases liking. The research on the distinction between desire and enjoyment contributes to the understanding of human addiction, particularly why individuals often continue to frantically pursue rewards such as cocaine, opiates, gambling, or sex, even when they no longer experience pleasure from obtaining these rewards due to habituation.

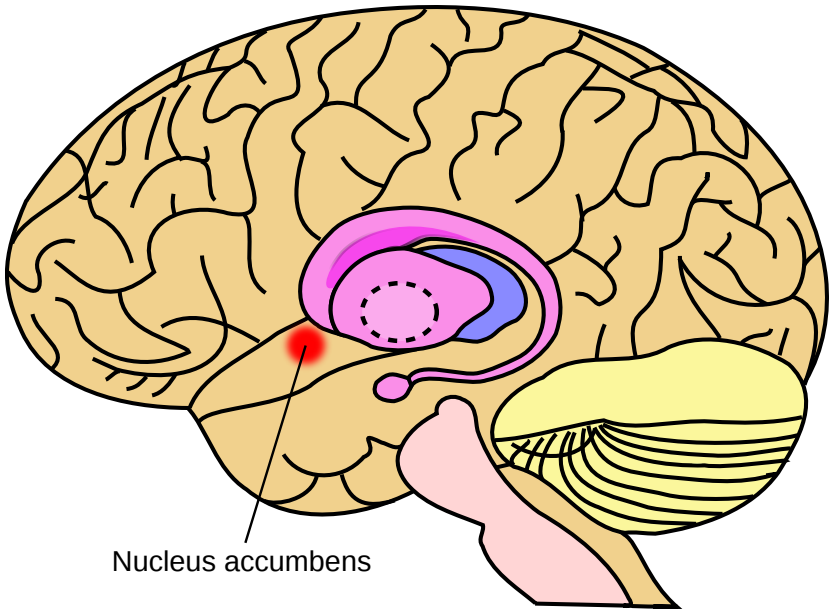


Figure **16.1.4**: The experience of pleasure also involves the orbitofrontal cortex. Neurons in this region and the nucleus accumbens particularly shown here in red fire when monkeys taste, or merely see pictures of, desirable foods. In humans, this region is activated by pleasant stimuli including money, pleasant smells, and attractive faces (Gottfried, O'Doherty & Dolan, 2002; O'Doherty, Deichmann, Critchley, & Dolan, 2002; O'Doherty, Kringelbach, Rolls, Hornak, & Andrews, 2001; O'Doherty, Winston, Critchley, Perrett, Burt, & Dolan, 2003). (File:BrainCaudatePutamen.svg: User:Leevanjackson Derivative work: User:SUM1, CC BY-SA 4.0, via Wikimedia Commons)

The experience of pleasure also involves the orbitofrontal cortex. Neurons in this region fire when monkeys taste, or merely see pictures of, desirable foods. In humans, this region is activated by pleasant stimuli including money, pleasant smells, and attractive faces (Gottfried, O'Doherty & Dolan, 2002; O'Doherty, Deichmann, Critchley, & Dolan, 2002; O'Doherty, Kringelbach, Rolls, Hornak, & Andrews, 2001; O'Doherty, Winston, Critchley, Perrett, Burt, & Dolan, 2003).

16.1.3.3 Fear: The Neural System of Freezing and Fleeing

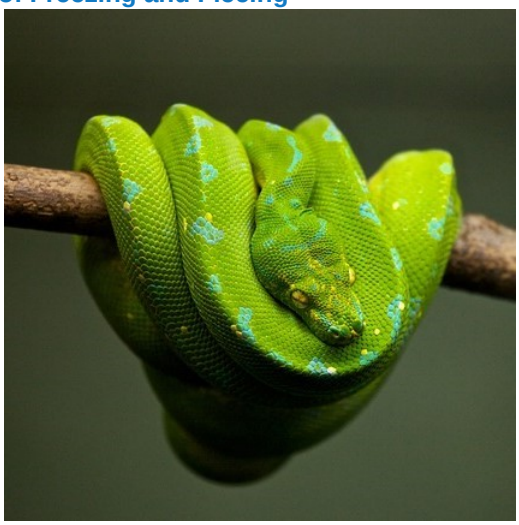


Figure **16.1.5**: Because fear is so important for our survival (i.e., fear informs us when something threatens us), our brains are able to "recognize" frightening stimuli before we ourselves are even consciously aware of them. (Copyright; CC0 Public Domain, goo.gl/m25gce)

Fear is an unpleasant emotion that motivates avoidance of potentially harmful situations. Slight stimulation of the fear-related areas in the brain causes animals to freeze, whereas intense stimulation causes them to flee. The fear circuit extends from the central amygdala to the **periaqueductal gray** in the midbrain. These structures are sensitive to glutamate, corticotrophin releasing factor, adreno-cortico-trophic hormone, cholecystokinin, and several different neuropeptides. Benzodiazepines and other tranquilizers inhibit activation in these areas (Panksepp & Biven, 2012). As is discussed in a later section, stress plays a role in releasing these

chemicals activating the fight/flight/freeze response. As discussed elsewhere (Chapter 6 and again in Chapter 16), benzodiazepines are used to lower anxiety or anticipation of a fearful situation by inhibiting the activity of these areas because they are GABA agonists.

The role of the **amygdala** in fear responses has been extensively studied. Perhaps because fear is so important to survival, two pathways send signals to the amygdala from the sensory organs. When an individual sees a snake, for example, the sensory information travels from the eye to the **thalamus** and then to the visual cortex. The **visual cortex** sends the information on to the amygdala, provoking a fear response. However, the thalamus also quickly sends the information straight to the amygdala, so that the organism can react before consciously perceiving the snake (LeDoux, Farb, & Ruggiero, 1990). The pathway from the thalamus to the amygdala is fast but less accurate than the slower pathway from the visual cortex. Damage to the amygdala or areas of the ventral hippocampus interferes with fear conditioning in both humans and nonhuman animals (LeDoux, 1996).

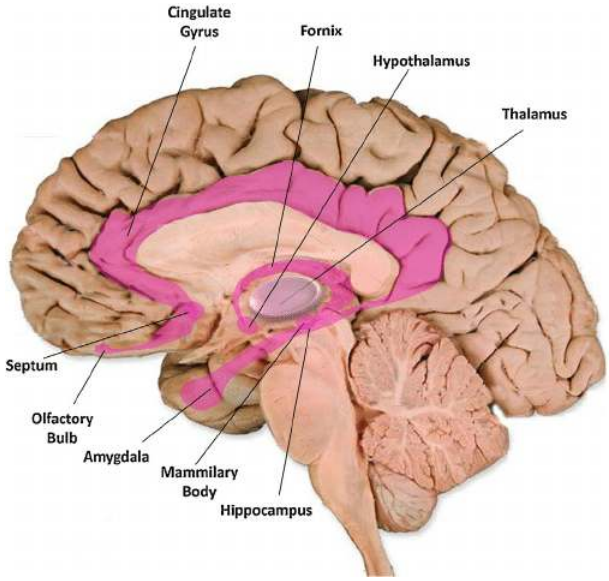


Figure **16.1.6**: The limbic system (highlighted here in pink), which includes the hypothalamus, thalamus, amygdala, and the hippocampus, is involved in mediating emotional response and memory. (Limbic system by SoniaM2020 licensed CC BY-SA 4.0)

16.1.3.4 Love: The Neural Systems of Care and Attachment



Figure **16.1.7**: Just as scientists today distinguish between types of love like “romantic” and “parental,” so did the ancient Greeks who used the terms “eros” and “storge.” [Image: Go-tea 郭天, https://goo.gl/KLFxJP, CC BY 2.0, goo.gl/zHmGV2]

For social animals such as humans, attachment to other members of the same species produces the positive emotions of attachment: love, warm feelings, and affection. The emotions that motivate nurturing behavior (e.g., maternal care) are distinguishable from those that motivate staying close to an attachment figure in order to receive care and protection (e.g., infant attachment). Important regions for maternal nurturing include the dorsal **preoptic area** (Numan & Insel, 2003) (see Figure) and the bed nucleus of the **stria terminalis**(Panksepp, 1998).

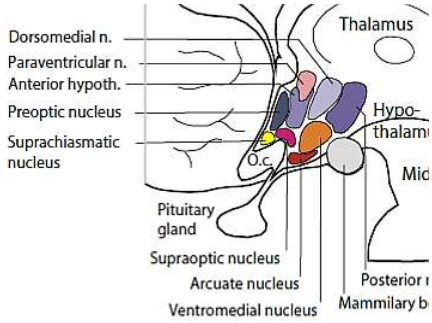


Figure **16.1.8**: Figure showing key hypothalamic nuclei including the preoptic nucleus in dark blue anterior to the anterior hypothalamus, behind the frontal cortex and above the SCN and supraoptic nucleus. Sanador2.0, CC BY-SA 3.0, via Wikimedia Commons

These regions overlap with the areas involved in sexual desire, and are sensitive to some of the same neurotransmitters, including oxytocin, arginine-vasopressin, and endogenous opioids (endorphins and enkephalins).

16.1.3.5 Grief: The Neural Networks of Loneliness and Panic

The neural networks involved in infant attachment are also sensitive to separation. These regions produce the painful emotions of grief, panic, and loneliness. When infant humans or other infant mammals are separated from their mothers, they produce distress vocalizations, or crying. The attachment circuits are those that cause organisms to produce distress vocalizations when electrically stimulated.

The attachment system begins in the midbrain periaqueductal gray, very close to the area that produces physical pain responses, suggesting that it may have originated from the pain circuits (Panksepp, 1998). Separation distress can also be evoked by stimulating the dorsomedial thalamus, ventral septum, dorsal preoptic region, and areas in the bed nucleus of stria terminalis (near sexual and maternal circuits). (Panksepp, Normansell, Herman, Bishop, & Crepeau, 1988)

These regions are sensitive to endogenous opiates, oxytocin, and prolactin. All of these neurotransmitters *prevent* separation distress. Opiate drugs such as morphine and heroin, as well as nicotine, artificially produce feelings of pleasure and gratification, similar to those normally produced during positive social interactions. This may explain why these drugs are addictive. Panic attacks appear to be an intense form of separation distress triggered by the attachment system, and panic can be effectively relieved by opiates. Testosterone also reduces separation distress, perhaps by reducing attachment needs. Consistent with this, panic attacks are more common in women than in men.

16.1.4 Knowledge Emotions

Paul Silvia (University of North Carolina, Greensboro) suggested that when people think of emotions they usually think of the obvious ones, such as happiness, fear, anger, and sadness. He instead looks at the knowledge emotions, a family of emotional states that foster learning, exploring, and reflecting. Surprise, interest, confusion, and awe come from events that are unexpected, complicated, and mentally challenging, and they motivate learning in its broadest sense, be it learning over the course of seconds (finding the source of a loud crash, as in surprise) or over a lifetime (engaging with hobbies, pastimes, and intellectual pursuits, as in interest). Of course there are several causes, consequences, and individual differences. As a group, the knowledge emotions motivate people to engage with new and puzzling things rather than avoid them. Over time, engaging with new things, ideas, and people broadens someone’s experiences and cultivates expertise. The knowledge emotions thus don’t gear up the body like fear, anger, and happiness do, but they do gear up the mind—a critical task for humans, who must learn essentially everything that they know.

So while emotions are something we think of in terms of feelings and the fight or flight response, or even as longer lasting sadness or love, Silvia's ideas point out the less noticed strong connection between emotion and cognition. It is clear that emotions can and do provide a strong basis for how and what we think about.

16.1.5 Plasticity: Experiences Can Alter the Brain

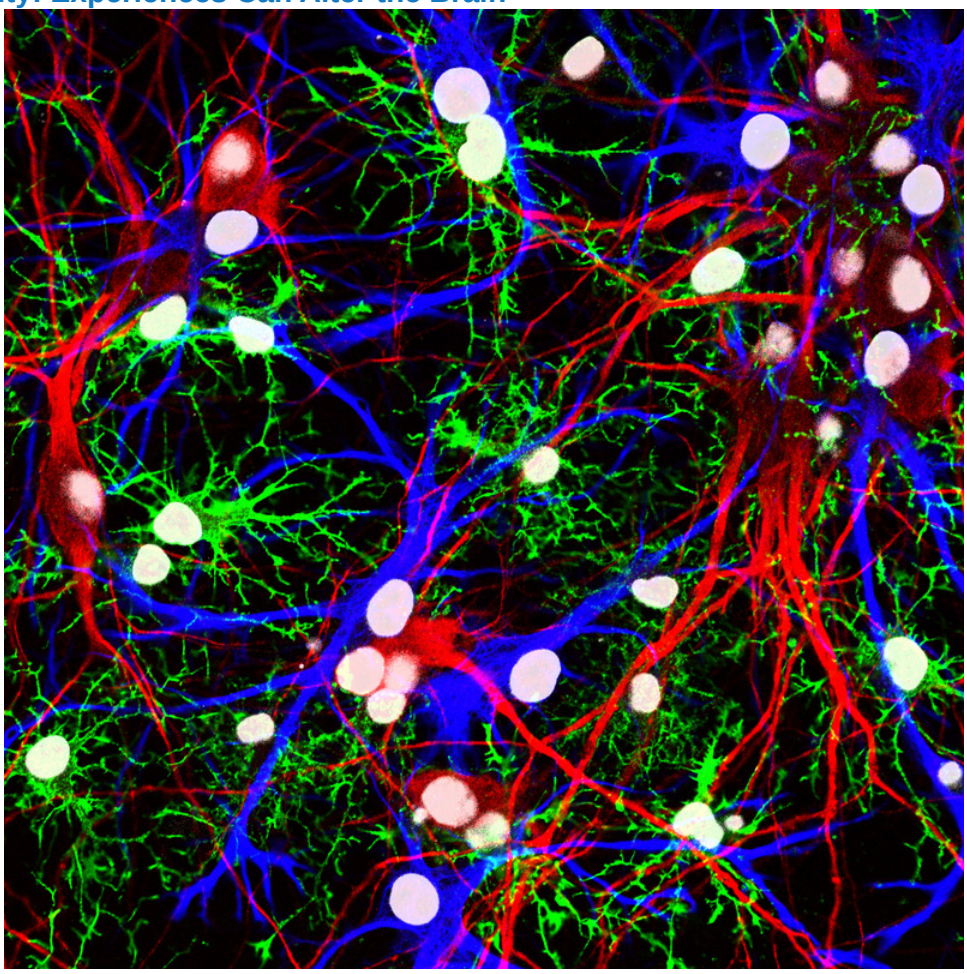


Figure **16.1.9**: Neural plasticity can be summed up in the phrase: "Neurons that fire together, wire together." Or in other words, when certain emotions are paired with certain contexts, we learn to associate the two together. Here we see the networks of neurons that are connected to each other. [Image: NICHD NIH, https://goo.gl/T3PUlg, CC BY 2.0, goo.gl/BRvSA7]

The responses of specific neural regions may be modified by experience. For example, the front shell of the nucleus accumbens is generally involved in appetitive behaviors, such as eating, and the back shell is generally involved in fearful defensive behaviors (Reynolds & Berridge, 2001, 2002). Research using human neuroimaging has also revealed this front–back distinction in the functions of the nucleus accumbens (Seymour, Daw, Dayan, Singer, & Dolan, 2007). However, when rats are exposed to stressful environments, their fear-generating regions expand toward the front, filling almost 90% of the nucleus accumbens shell. On the other hand, when rats are exposed to preferred home environments, their fear-generating regions shrink and the appetitive regions expand toward the back, filling approximately 90% of the shell (Reynolds & Berridge, 2008). Consider how this might be generalized to human experiences that are stressful versus those that are generally comforting and comfortable.

16.1.6 Brain Structures Have Multiple Functions

Although much affective neuroscience research has emphasized whole structures, such as the amygdala and nucleus accumbens, it is important to note that many of these structures are more accurately referred to as complexes. They include distinct groups of nuclei that perform different tasks. At present, human neuroimaging techniques such as fMRI are unable to examine the activity of individual nuclei in the way that invasive animal neuroscience can. For instance, the amygdala of the nonhuman primate can be divided into 13 nuclei and cortical areas (Freese & Amaral, 2009). These regions of the amygdala perform different functions. The central nucleus sends outputs involving brainstem areas that result in innate emotional expressions and associated physiological responses. The basal nucleus is connected with striatal areas that are involved with actions such as running toward safety. Furthermore, it is not possible to make one-to-one maps of emotions onto brain regions. For example, extensive research has

examined the involvement of the amygdala in fear, but research has also shown that the amygdala is active during uncertainty (Whalen, 1998) as well as positive emotions (Anderson et al., 2003; Schulkin, 1990).

16.1.7 Conclusion

Research in affective neuroscience has contributed to knowledge regarding emotional, motivational, and behavioral processes. The study of the basic emotional systems of nonhuman animals provides information about the organization and development of more complex human emotions. Although much still remains to be discovered, current findings in affective neuroscience have already influenced our understanding of drug use and abuse, psychological disorders such as panic disorder, and complex human emotions such as desire and enjoyment, grief and love.

16.1.8 Outside Resources

Video: A 1-hour interview with Jaak Panksepp, the father of affective neuroscience
Video: A 15-minute interview with Kent Berridge on pleasure in the brain
Video: A 5-minute interview with Joseph LeDoux on the amygdala and fear
Web: Brain anatomy interactive 3D model

16.1.9 References

- Anderson, A. K., Christoff, K., Stappen, I., Panitz, D., Ghahremani, D. G., Glover, G., . . . Sobel, N. (2003). Dissociated neural representations of intensity and valence in human olfaction. *Nature Neuroscience*, *6*, 196–202.
- Berkman, E. T., & Lieberman, M. D. (2010). Approaching the bad and avoiding the good: Lateral prefrontal cortical asymmetry distinguishes between action and valence. *Journal of Cognitive Neuroscience*, *22*(9), 1970–1979. doi: 10.1162/jocn.2009.21317
- Berridge, K. C., & Kringelbach, M. L. (2013). Neuroscience of affect: brain mechanisms of pleasure and displeasure. *Current Opinion in Neurobiology*, *23*, 294–303. doi.org/10.1016/j.conb.2013.01.017
- Berridge, K. C., & Kringelbach, M. L. (2008). Affective neuroscience of pleasure: Reward in humans and animals. *Psychopharmacology*, *199*, 457–480. doi: 10.1007/s00213-008-1099-6
- Blanchard, D. C., & Blanchard, R. J. (2003). What can animal aggression research tell us about human aggression? *Hormones and Behavior*, *44*, 171–177.
- Farb, N.A.S., Chapman, H. A., & Anderson, A. K. (2013). Emotions: Form follows function. *Current Opinion in Neurobiology, 23*, 393–398. http://dx.doi.org/10.1016/j.conb.2013.01.015
- Fox, N. A., & Davidson, R. J. (1986). Taste-elicited changes in facial signs of emotion and the asymmetry of brain electrical activity in human newborns. *Neuropsychologia*, *24*, 417–422.
- Freese, J. L., & Amaral, D. G. (2009). Neuroanatomy of the primate amygdala. In P. J. Whalen & E. A. Phelps (Eds.), *The human amygdala* (pp. 3–42). New York, NY: Guilford Press.
- Gable, P. A., & Harmon-Jones, E. (2008). Relative left frontal activation to appetitive stimuli: Considering the role of individual differences. *Psychophysiology*, *45*, 275-278.
- Goldstein, K. (1939). *The organism: An holistic approach to biology, derived from pathological data in man*. New York, NY: American Book.
- Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2002). Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *Journal of Neuroscience*, *22*, 10829–10837.
- Gray, J. A. (1987). *The psychology of fear and stress* (2nd ed.). Cambridge, England: Cambridge University Press.
- Harmon-Jones, E., Harmon-Jones, C., & Price, T. F. (2013). What is approach motivation? *Emotion Review, 5*, 291–295. doi: 10.1177/1754073913477509
- Heinrichs, M., von Dawans, B., & Domes, G. (2009). Oxytocin, vasopressin, and human social behavior. *Frontiers in Neuroendocrinology, 30*, 548–557.
- Izard, C. E. (2010). The many meanings/aspects of emotion: Definitions, functions, activation, and regulation. *Emotion Review, 2*, 363–370. doi: 10.1177/1754073910374661
- LeDoux, J. E. (1996). *The emotional brain: The mysterious underpinnings of emotional life*. New York, NY: Simon & Schuster.
- LeDoux, J. E., Farb, C. F., Ruggiero, D. A. (1990). Topographic organization of neurons in the acoustic thalamus that project to the amygdala. *Journal of Neuroscience*, *10*, 1043–1054.
- Numan, M., & Insel, T. R. (2003). *The neurobiology of parental behavior*. New York, NY: SpringerVerlag.

- O'Doherty J. P., Deichmann, R., Critchley, H. D., & Dolan, R. J. (2002). Neural responses during anticipation of a primary taste reward. *Neuron, 33*, 815–826.
- O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience*, *4*, 95–102.
- O'Doherty, J., Winston, J., Critchley, H., Perrett, D., Burt, D. M., & Dolan, R. J. (2003). Beauty in a smile: The role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia*, *41*, 147–155.
- Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. New York, NY: Oxford University Press.
- Panksepp, J., & Biven, L. (2012). *The archaeology of mind: Neuroevolutionary origins of human emotions*. New York, NY: Norton.
- Panksepp, J., Normansell, L., Herman, B., Bishop, P., & Crepeau, L. (1988). Neural and neurochemical control of the separation distress call. In J. D. Newman (Ed.), *The physiological control of mammalian vocalization* (pp. 263–299). New York, NY: Plenum.
- Peterson, C. K., & Harmon-Jones, E. (2012). Anger and testosterone: Evidence that situationally-induced anger relates to situationally-induced testosterone. *Emotion, 12*, 899–902. doi: 10.1037/a0025300
- Reynolds, S. M., & Berridge, K. C. (2008). Emotional environments retune the valence of appetitive versus fearful functions in nucleus accumbens. *Nature Neuroscience*, *11*, 423–425.
- Reynolds, S. M., & Berridge, K. C. (2002). Positive and negative motivation in nucleus accumbens shell: Bivalent rostrocaudal gradients for GABA-elicited eating, taste "liking"/"disliking" reactions, place preference/avoidance, and fear. *Journal of Neuroscience*, *22*, 7308–7320.
- Reynolds, S. M., & Berridge, K. C. (2001). Fear and feeding in the nucleus accumbens shell: Rostrocaudal segregation of GABA-elicited defensive behavior versus eating behavior. *Journal of Neuroscience*, *21*, 3261–3270.
- Schulkin, J. (1991). *Sodium hunger: The search for a salty taste*. New York, NY: Cambridge University Press.
- Seymour, B., Daw, N., Dayan, P., Singer, T., & Dolan, R. (2007). Differential encoding of losses and gains in the human striatum. *Journal of Neuroscience*, *27*, 4826–4831.
- Stewart, J., De Wit, H., & Eikelboom, R. (1984). Role of unconditioned and conditioned drug effects in the self-administration of opiates and stimulants. *Psychological Review, 91*, 251-268.
- Wacker, J., Mueller, E. M., Pizzagalli, D. A., Hennig, J., & Stemmler, G. (2013). Dopamine-D2-receptor blockade reverses the association between trait approach motivation and frontal asymmetry in an approach-motivation context. *Psychological Science*, *24*(4), 489–497. doi: 10.1177/0956797612458935
- Whalen, P. J. (1998). Fear, vigilance, and ambiguity: initial neuroimaging studies of the human amygdala. *Current Directions in Psychological Science*, *7*, 177–188.

Attributions:

Affective Neuroscience By Eddie Harmon-Jones and Cindy Harmon-Jones, University of New South Wales, Noba Project, licensed CC BY-NC-SA 4.0

File:BrainCaudatePutamen.svg: User:LeevanjacksonDerivative work: User:SUM1, CC BY-SA 4.0, via Wikimedia Commons

Images are generated by Life Science Databases(LSDB)., CC BY-SA 2.1 JP, via Wikimedia Commons

Sanador2.0, CC BY-SA 3.0, via Wikimedia Commons

Knowledge Emotions by Paul Silvia, University North Carolina, Greensboro, Noba Project, licensed CC BY-NC-SA 4.0

This page titled 16.1: The Neurological Bases of Emotions is shared under a mixed license and was authored, remixed, and/or curated by ASCCC OERI & Bakhtawar Bhadha (ASCCC Open Educational Resources Initiative (OERI)) .

16.2: Valence of Emotions

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Describe the functions of emotions and categories of positive and negative ones.
- Differentiate intensity and fluctuation of different emotions.
- Analyze the significance of context and external environment in emotional experiences.
- Evaluate how it is possible to feel more than one emotion at a time, and how that is represented in the brain.

16.2.1 Overview

Emotions don't just feel good or bad, they also contribute crucially to people's well-being and health. In general, experiencing positive emotions is good for us, whereas experiencing negative emotions is bad for us. However, research on emotions and well-being suggests this simple conclusion is incomplete and sometimes even wrong. Taking a closer look at this research, this section provides a more complex relationship between emotion and well-being. At least three aspects of the emotional experience appear to affect how a given emotion is linked with well-being: the intensity of the emotion experienced, the fluctuation of the emotion experienced, and the context in which the emotion is experienced. While it is generally good to experience more positive emotion and less negative emotion, this is not always the guide to the good life.

Given how we feel adds much of the flavor to life's highest—and lowest—moments, can you think of an important moment in your life that didn't involve strong feelings? In fact, it might be hard to recall *any* times when you had no feeling at all. Given how full life is with feelings and given how profoundly feelings affect us, it is not surprising that much theorizing and research has been devoted to uncovering how we can optimize our feelings, or "emotion experiences," as they are referred to in psychological research.

16.2.2 Feelings Contribute to Well-Being

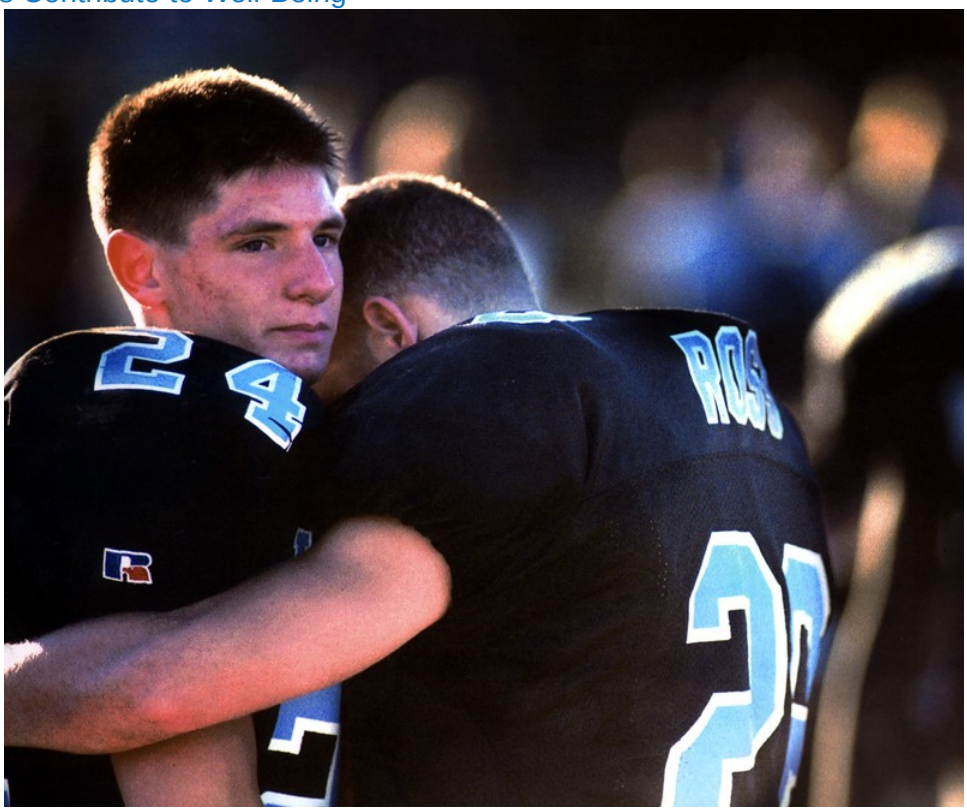


Figure **16.2.1**: Although we tend to think we should always strive for "positive" or "feel-good" emotions, in some situations a negative emotion might be best. For example feeling disappointment when experiencing a failure might help motivate you for future success. [Image: CC0 Public Domain, goo.gl/m25gce]

So, which **emotions** are the "best" ones to feel? Take a moment to think about how you might answer this question. At first glance, the answer might seem obvious. Of course, we should experience as much positive emotion and as little negative emotion as possible! Why? Because it is pleasant to experience positive emotions and it is unpleasant to experience negative emotions (Russell & Barrett, 1999). The conclusion that positive feelings are good and negative feelings are bad might seem so obvious as not to even warrant the question, much less bother with psychological research. In fact, the very labels of "positive" and "negative" imply the answer to this question. However, for the purposes of this section, it may be helpful to think of "positive" and "negative" as descriptive terms used to discuss two different types of experiences, rather than a true value judgment. Thus, whether positive or negative emotions are good or bad for us needs to be tested in particular situations (See **16.2.1**).

As it turns out, this empirical question has been on the minds of theorists and researchers for many years. Many psychologists began asking whether the effects of feelings could go beyond the obvious momentary pleasure or displeasure. In other words, can emotions do more for us than simply make us feel good or bad? This is not necessarily a new question; variants of it have appeared in the texts of thinkers such as Charles Darwin (1872) and Aristotle (1999). However, modern psychological research has provided empirical evidence that feelings are not just inconsequential byproducts. Rather, each emotional experience, however fleeting, has effects on cognition, behavior, and the people around us. For example, feeling happy is not only pleasant, but is also useful to feel when in social situations because it helps us be friendly and collaborative, thus promoting our positive relationships. Over time, the argument goes, these effects add up to have tangible effects on people's **well-being** (good mental and physical health).

A variety of research has been inspired by the notion that our emotions are involved in, and maybe even causally contribute to, our well-being. This research has shown that people who experience more frequent positive emotions and less frequent negative emotions have higher well-being (e.g., Fredrickson, 1998; Lyubomirksy, King, & Diener, 2005), including increased life satisfaction (Diener, Sandvik, & Pavot, 1991), increased physical health (Tugade, Fredrickson, & Barrett, 2004; Veenhoven, 2008), greater resilience to stress (Folkman & Moskowitz, 2000; Tugade & Fredrickson, 2004), better social connection with others (Fredrickson, 1998), and even longer lives (Veenhoven, 2008). Notably, the effect of positive emotion on longevity is about as powerful as the effect of smoking on reducing lifespan! Perhaps most importantly, some research directly supports that emotional

experiences *cause* these various outcomes rather than being just a consequence of them (Fredrickson, Cohn, Coffey, Pek, & Finkel, 2008; Lyubomirsky et al., 2005).



Figure **16.2.2**: Not only do the emotions we feel vary by the context, but also the emotions we should feel depend on the circumstances, too. [Image: puppywind, https://goo.gl/BQKhKK, CC BY-NC-ND 2.0, goo.gl/h9gK3o]

At this point, you might be tempted to conclude that you should always strive to experience as much positive emotion and as little negative emotion as possible. However, there is evidence to suggest that this conclusion may be premature. This is because this conclusion neglects three central aspects of the emotion experience. First, it neglects the intensity of the emotion: Positive and negative emotions might not have the same effect on well-being at all intensities. Second, it neglects how emotions fluctuate over time: Stable emotion experiences might have quite different effects from experiences that change a lot. Third, it neglects the context in which the emotion is experienced: The context in which we experience an emotion might profoundly affect whether the emotion is good or bad for us (See Figure **16.2.2**). So, to address the question "Which emotions should we feel?" we must answer, "It depends!" We next consider each of the three aspects of feelings, and how they influence the link between feelings and well-being.

16.2.3 The Intensity of the Emotion Matters

Experiencing more *frequent* positive emotions is generally beneficial. But does this mean that we should strive to feel as *intense* positive emotion as possible? Recent research suggests that this unqualified conclusion might be wrong.

In fact, experiencing very high levels of positive emotion may be harmful (Gruber, 2011; Oishi, Diener, & Lucas, 2007). For instance, experiencing *very* high levels of positive emotion makes individuals more likely to engage in risky behaviors, such as binge eating and drug use (Cyders & Smith, 2008; Martin et al., 2002). Furthermore, intense positive emotion is associated with the experience of mania (Gruber et al., 2009; Johnson, 2005). It appears that the experience of positive emotions follows an inverted U-shaped curve in relation to well-being: more positive emotion is linked with increased well-being, but only up to a point, after which even more positive emotion is linked with decreased well-being (Grant & Schwartz, 2011). These empirical findings underscore the sentiment put forth long ago by the philosopher Aristotle: Moderation is key to leading a good life (1999).

Too much positive emotion may pose a problem for well-being. Might too little negative emotion similarly be cause for concern? Although there is limited empirical research on this subject, initial research supports this idea. For example, people who aim *not* to feel negative emotion are at risk for worse well-being and adaptive functioning, including lower life satisfaction, lower social

support, worse college grades, and feelings of worse physical health (Tamir & Ford, 2012a). Similarly, feeling too little embarrassment in response to a social faux pas may damage someone's social connections if they aren't motivated by their embarrassment to make amends (Keltner & Buswell, 1997). Low levels of negative emotion also seem to be involved in some forms of psychopathology. For instance, blunted sadness in response to a sad situation is a characteristic of major depressive disorder (Rottenberg, Gross, & Gotlib, 2005) and feeling too little fear is a hallmark of psychopathy (Marsh et al., 2008; Patrick, 1994).

In sum, this first section suggests that the conclusion "Of course we should experience as much positive emotions and as little negative emotions as possible" is sometimes wrong. As it turns out, there can be too much of a good thing and too little of a bad thing.

16.2.4 The Fluctuation of the Emotion Matters



Figure **16.2.3**: Throughout our lives, our emotions vary such that we experience great highs and great lows. However, if you average those emotions over time, we tend to end up somewhere in the middle. [Image: Moazzam Brohi, https://goo.gl/acZniv, CC BY 2.0, goo.gl/BRvSA7]

Emotions naturally vary—or fluctuate—over time (Davidson, 1998) (See Figure **16.2.3**). We probably all know someone whose emotions seem to fly everywhere—one minute they're ecstatic, the next they're upset. We might also know a person who is pretty even-keeled, moderately happy, with only modest fluctuations across time. When looking only at *average* emotion experience, say across a month, both of these people might appear identical: moderately happy. However, underlying these identical averages are two very different patterns of fluctuation across time. Might these **emotion fluctuations** across time—beyond average intensity—have implications for well-being?

Overall, the available research suggests that how much emotions fluctuate does indeed matter. In general, greater fluctuations are associated with worse well-being. For example, higher fluctuation of positive emotions—measured either within a single day or across two weeks—was linked with lower well-being and greater depression (Gruber, Kogan, Quoidbach, & Mauss, 2013). Fluctuation in negative emotions, in turn, has been linked with increased depressive symptoms (Peeters, Berkhof, Delespaul, Rottenberg, & Nicolson, 2003), borderline personality disorder (Trull et al., 2008), and neuroticism (Eid & Diener, 1999). These associations tend to hold even when controlling for average levels of positive or negative emotion, which means that beyond the overall intensity of positive or negative emotion, the fluctuation of one's emotions across time is associated with well-being. While

it is not entirely clear why fluctuations are linked to worse well-being, one explanation is that strong fluctuations are indicative of emotional instability (Kuppens, Oravecz, & Tuerlinckx, 2010).

Of course, this should not be taken to mean that we should rigidly feel the exact same way every minute of every day, regardless of context. After all, psychological flexibility—or the ability to adapt to changing situational demands and experience emotions accordingly—has generally demonstrated beneficial links with well-being (Bonanno, Papa, Lalande, Westphal, & Coifman, 2004; Kashdan, & Rottenberg, 2010). The question remains, however: what exact amount of emotional fluctuation constitutes unhealthy instability and what amount of emotional fluctuation constitutes healthy flexibility.

Again, then, we must qualify the conclusion that it is always better to experience more positive emotions and less negative emotions. The degree to which emotions fluctuate across time plays an important role. Overall, relative stability (but not rigidity) in emotion experience appears to be optimal for well-being.

16.2.5 The Context of the Emotion Experience Matters

This section has already discussed two features of emotion experiences that affect how they relate to well-being: the intensity of the emotion and the fluctuation of the emotion over time. However, neither of these features takes into account the context in which the emotion is experienced. At least three different contexts may critically affect the links between emotion and well-being: (1) the external environment in which the emotion is being experienced, (2) the other emotional responses (e.g., physiology, facial behavior) that are currently activated, and (3) the other emotions that are currently being experienced.

16.2.6 The External Environment



Figure **16.2.4**: Feeling an emotion that matches the person's context (e.g., experiencing happiness at a birthday party) is typically the most functional or beneficial emotion to feel. [Image: OakleyOriginals, https://goo.gl/IxfIsq, CC BY 2.0, goo.gl/BRvSA7]

Emotions don't occur within a vacuum. Instead, they are usually elicited by and experienced within specific situations that come in many shapes and sizes —from birthday parties to funerals, job interviews to mundane movie nights. The situation in which an

emotion is experienced has strong implications for whether a given emotion is the “best” emotion to feel. Take happiness, for example. Feeling happiness at a birthday party may be a great idea (See Figure **16.2.4**). However, having the exact same experience of happiness at a funeral would likely not bode well for your well-being.

When considering how the environment influences the link between emotion and well-being, it is important to understand that each emotion has its own function. For example, although fear is a negative emotion, fear helps us notice and avoid threats to our safety (öhman & Mineka, 2001), and may thus be the “best” emotion to feel in dangerous situations. Happiness can help people cooperate with others, and may thus be the best emotion to feel when we need to collaborate (e.g., Van Kleef, van Dijk, Steinel, & van Beest, 2008). Anger can energize people to compete or fight with others, and may thus be advantageous to experience it in confrontations (e.g., Tamir & Ford, 2012b; Van Kleef et al., 2008). It might be disadvantageous to experience happiness (a positive emotion) when we need to fight with someone; in this situation, it might be better to experience anger (a negative emotion). This suggests that emotions’ implications for well-being are not determined only by whether they are positive or negative but also by whether they are well-matched to their context.

In support of this general idea, people who experience emotions that fit the context at hand are more likely to recover from depression and trauma (Bonanno et al., 2004; Rottenberg, Kasch, Gross, & Gotlib, 2002). Research has also found that participants who want to feel emotions that match the context at hand (e.g., anger when confronting someone)—even if that emotion was negative—are more likely to experience greater well-being (Tamir & Ford, 2012a). Conversely, people who pursue emotions without regard to context—even if those emotions are positive, like happiness—are more likely to experience lower subjective well-being, more depression, greater loneliness, and even worse grades (Ford & Tamir, 2012; Mauss et al., 2012; Mauss, Tamir, Anderson, & Savino; 2011; Tamir & Ford, 2012a).

In sum, this research demonstrates that regardless of whether an emotion is positive or negative, the context in which it is experienced critically influences whether the emotion helps or hinders well-being.

16.2.7 Other Emotional Responses



Figure **16.2.5**: If we experience the emotion of amusement (from seeing something funny), we often have the physiological response to laugh. This is an example of emotion coherence, where we express a particular behavior associated with a particular emotion. [Image: Ed Schipul, https://goo.gl/7NUYmR, CC BY-SA 2.0, goo.gl/rxiUsF]

The subjective experience of an emotion—what an emotion *feels* like—is only one aspect of an emotion. Other aspects include behaviors, facial expressions, and physiological activation (Levenson, 1992). For example, if you feel excited about having made a

new friend, you might want to be near that person, you might smile, and your heart might be beating faster as you do so. Often, these different responses travel together, meaning that when we feel an emotion we *typically* have corresponding behaviors and physiological responses (e.g.,Ekman, 1972; Levenson, 1992) (See Figure **16.2.5**) . The degree to which responses travel together has sometimes been referred to as **emotion coherence** (Mauss, Levenson, McCarter, Wilhelm, & Gross, 2005). However, these different responses do not co-occur in all instances and for all people (Bradley & Lang, 2000; Mauss et al., 2005; for review, see Fridlund, Ekman, & Oster, 1987). For example, some people may choose not to express an emotion they are feeling internally (English & John, 2013), which would result in lower coherence.

Does coherence—above and beyond emotion experience per se—matter for people's well-being? To examine this question, one study measured participants' emotion coherence by showing them a funny film clip of stand-up comedy while recording their experience of positive emotion as well as their behavioral displays of positive emotion (Mauss, Shallcross, et al., 2011). As predicted, participants differed quite a bit in their coherence. Some showed almost perfect coherence between their behavior and experience, whereas others' behavior and experience corresponded not much at all. Interestingly, the more that participants' behavior and experience cohered in the laboratory session, the lower levels of depressive symptoms and the higher levels of well-being they experienced 6 months later. This effect was found when statistically controlling for overall intensity of positive emotions experienced. In other words, experiencing high levels of positive emotion aided well-being *only if it was accompanied by corresponding positive facial expressions.*

But *why* would coherence of different emotional responses predict well-being? One of the key functions of an emotion is social communication (Keltner & Haidt, 1999), and arguably, successful social communication depends on whether an individual's emotions are being *accurately* communicated to others. When someone's emotional behavior doesn't match their experience, it may disrupt communication because it could make the individual appear confusing or inauthentic to others. In support of this theory, the above study found that lower coherence was associated with worse well-being because people with lower coherence felt less socially connected to others (Mauss, Shallcross, et al., 2011). These findings are also consistent with a large body of research examining the extent to which people mask the outward display of an emotional experience, or *suppression*. This research has demonstrated that people who habitually use suppression not only experience worse well being (Gross & John, 2003), but they also seem to be particularly worse off with regard to their social relationships (Srivastava, Tamir, McGonigal, John, & Gross, 2009).

These findings underscore the importance of examining whether an individual's experience is traveling together with his or her emotional responses, above and beyond overall levels of subjective experience. Thus, to understand how emotion experiences predict well-being, it is important not only to consider the experience of an emotion, but also the other emotional responses currently activated.

16.2.8 Other Emotions

Up until now, we have treated emotional experiences as though people can only experience one emotion at a time. However, it should be kept in mind that positive and negative emotions are not simply the opposite of one another. Instead, they tend to be independent of one another, which means that a person can feel positive and negative emotions at the same time (Larsen, McGraw, Mellers, & Cacioppo, 2004). For example, how does it feel to win a prize when you expected a greater prize? Given "what might have been," situations like this can elicit both happiness and sadness. Or, take "schadenfreude" (a German term for deriving pleasure from someone else's misfortune), or "aviman" (an Indian term for prideful, loving anger), or nostaligia (an English term for affectionate sadness about something from the past): these terms capture the notion that people can feel both positively and negatively within the same emotional experience. And as it turns out, the other emotions that someone feels (e.g., sadness) during the experience of an emotion (e.g., happiness) influence whether that emotion experience has a positive or negative effect on well-being.



Figure **16.2.6**: What experiences can you recall where you felt mixed emotions? Happiness and disappointment? Hope and fear? Admiration and envy? [Image: Ron Cogswell, https://goo.gl/JKuzmU, CC BY 2.0, goo.gl/BRvSA7]

Importantly, the extent to which someone experiences different emotions at the same time—or *mixed emotions*—may be beneficial for their well-being (See Figure **16.2.6**). Early support for this theory was provided by a study of bereaved spouses. In the study, participants were asked to talk about their recently deceased spouse, which undoubtedly elicited strong negative emotions. However, some participants expressed positive emotions in addition to the negative ones, and it was those participants who recovered more quickly from their loss (Bonanno & Keltner, 1997). A recent study provides additional support for the benefits of mixed emotions, finding that adults who experienced more mixed emotions over a span of 10 years were physically healthier than adults whose experience of mixed emotions did not increase over time (Hershfield, Scheibe, Sims & Carstensen, 2013). Indeed, individuals who can experience positive emotions even in the face of negative emotions are more likely to cope successfully with stressful situations (Larsen, Hemenover, Norris, & Cacioppo, 2003).

Why would mixed emotions be beneficial for well-being? Stressful situations often elicit negative emotions, and recall that negative emotions have some benefits, as we outlined above. However, so do positive emotions, and thus having the ability to "take the good with the bad" might be another key component of well-being. Again, experiencing more positive emotion and less negative emotion may not always be optimal. Sometimes, a combination of both may be best.

16.2.9 Conclusion

Are emotions just fleeting experiences with no consequence beyond our momentary comfort or discomfort? A variety of research answers a firm "no"—emotions are integral predictors of our well-being. This module examined how, exactly, emotion experience might be linked to well-being. The obvious answer to this question is: *of course*, experiencing as much positive emotions and as little negative emotions as possible is good for us. But although this is true in general, recent research suggests that this obvious answer is incomplete and sometimes even wrong. As philosopher Robert Solomon said, "Living well is not just maximizing the good feelings and minimizing the bad. (…) A happy life is not necessarily filled with happy moments" (2007, p. 86).

16.2.10 Outside Resources

Journal: If you are interested in direct access to research on emotion, take a look at the journal *Emotion*

http://www.apa.org/pubs/journals/emo/index.aspx

Video: Check out videos of expert emotion researchers discussing their work

http://www.youtube.com/playlist?list...n43G_Y5otqKzJA

16.2.11 Discussion Questions

1. Much research confirms the relative benefits of positive emotions and relative costs of negative emotions. Could positive emotions be detrimental, or could negative emotions be beneficial? Why or why not?
2. We described some contexts that influence the effects of emotional experiences on well-being. What other contexts might influence the links between emotions and well-being? Age? Gender? Culture? How so?
3. How could you design an experiment that tests…(A) When and why it is beneficial to feel a negative emotion such as sadness? (B) How is the coherence of emotion behavior and emotion experience linked to well-being? (C) How likely a person is to feel mixed (as compared to simple) emotions?

16.2.12 Vocabulary

Emotion

An experiential, physiological, and behavioral response to a personally meaningful stimulus.

Emotion coherence

The degree to which emotional responses (subjective experience, behavior, physiology, etc.) converge with one another.

Emotion fluctuation

The degree to which emotions vary or change in intensity over time.

Well-being

The experience of mental and physical health and the absence of disorder.

16.2.13 References

- Aristotle. (1999). *Nicomachean ethics* (M. Ostwald, Trans.). Upper Saddle River, NJ: Prentice Hall.
- Bonanno, G. A., & Keltner, D. (1997). Facial expressions of emotion and the course of conjugal bereavement. *Journal of Abnormal Psychology, 106*, 126–137.
- Bonanno, G. A., Papa, A., Lalande, K., Westphal M., & Coifman, K. (2004). The importance of being flexible: The ability to both enhance and suppress emotional expression predicts long-term adjustment. *Psychological Science, 15*, 482–487.
- Bradley, M. M., & Lang, O. J. (2000). Measuring emotion: Behavior, feeling, and physiology. In R. D. Lane & L. Nadel (Eds.), *Cognitive neuroscience of emotion* (pp. 242–276). New York, NY: Oxford University Press.
- Cyders, M. A., & Smith, G. T. (2008). Emotion-based dispositions to rash action: Positive and negative urgency. *Psychological Bulletin, 134*, 807–828.
- Darwin, C. (1872). *The expression of emotions in man and animals*. New York, NY: Philosophical Library.
- Davidson, R. J. (1998). Affective style and affective disorders: Perspectives from affective neuroscience. *Cognition & Emotion, 12*, 307–330.
- Diener, E., Sandvik, E., & Pavot, W. (1991). Happiness is the frequency, not the intensity, of positive versus negative affect. *Subjective well-being: An interdisciplinary perspective, 21*, 119–139.
- Eid, M., & Diener, E. (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of Personality and Social Psychology, 76*, 662–676.
- Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska symposium on motivation* (pp. 207–282). Lincoln: University of Nebraska Press.
- English, T., & John, O. P. (2013). Understanding the social effects of emotion regulation: The mediating role of authenticity for individual differences in suppression. *Emotion, 13*, 314.
- Folkman, S., & Moskowitz, J. T. (2000). Positive affect and the other side of coping. *American Psychologist, 55*, 647–654.

- Ford, B. Q., & Tamir, M. (2012). When getting angry is smart: Emotional preferences and emotional intelligence. *Emotion, 12*, 685.
- Fredrickson, B. L. (1998). What good are positive emotions? *Review of General Psychology, 2*, 300–319.
- Fredrickson, B. L., Cohn, M. A., Coffey, K. A., Pek, J., & Finkel, S. M. (2008). Open hearts build lives: Positive emotions, induced through loving-kindness meditation, build consequential personal resources. *Journal of Personality and Social Psychology, 95*, 1045-1062.
- Fredrickson, B. L., Cohn, M. A., Coffey, K. A., Pek, J., & Finkel, S. M. (2008). Open hearts build lives: Positive emotions, induced through loving-kindness meditation, build consequential personal resources. *Journal of Personality and Social Psychology, 95*, 1045-1062.
- Fridlund, A. J., Ekman, P., & Oster, H. (1987). Facial expressions of emotion. In A. W. Siegman & S. Feldstein (Eds.), *Nonverbal behavior and communication* (2nd ed., pp. 143–223). Hillsdale, NJ: Erlbaum.
- Grant, A. M., & Schwartz, B. (2011). Too much of a good thing: The challenge and opportunity of the inverted U. *Perspectives on Psychological Science, 6*, 61–76.
- Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *Journal of personality and social psychology, 85*, 348-362.
- Gruber, J. (2011). When feeling good can be bad: Positive emotion persistence (PEP) in bipolar disorder. *Current Directions in Psychological Science, 20*, 217–221.
- Gruber, J., Culver, J. L., Johnson, S. L., Nam, J., Keller, K. L., & Ketter, T. K. (2009). Do positive emotions predict symptomatic change in bipolar disorder? *Bipolar Disorders, 11*, 330–336.
- Gruber, J., Kogan, A., Quoidbach, J., & Mauss, I. B. (2013). Happiness is best kept stable: Positive emotion variability is associated with poorer psychological health. *Emotion, 13*, 1–6.
- Hershfield, H. E., Scheibe, S., Sims, T. L., & Carstensen, L. L. (2013). When feeling bad can be good: Mixed emotions benefit physical health across adulthood. *Social Psychological and Personality Science, 4*, 54–61.
- Johnson, S. L. (2005). Mania and dysregulation in goal pursuit. *Clinical Psychology Review, 25*, 241–262.
- Kashdan, T. B., & Rottenberg, J. (2010). Psychological flexibility as a fundamental aspect of health. *Clinical Psychology Review, 30*, 865–878.
- Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion, 13*, 505-521.
- Keltner, D. & Buswell, B. N. (1997). Embarrassment: Its distinct form and appeasement functions. *Psychological Bulletin, 122*, 250–270.
- Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology, 99*, 1042–1060.
- Larsen, J. T., Hemenover, S. H., Norris, C. J., & Cacioppo, J. T. (2003). Turning adversity to advantage: On the virtues of the coactivation of positive and negative emotions. In L. G. Aspinwall & U. M. Staudinger (Eds.), *A psychology of human strengths: Fundamental questions and future directions for a positive psychology* (pp. 211–225).
- Larsen, J. T., McGraw, A. P., Mellers, B. A., & Cacioppo, J. T. (2004). The agony of victory and thrill of defeat: Mixed emotional reactions to disappointing wins and relieving losses. *Psychological Science, 15*, 325–330.
- Levenson, R. W. (1992). Autonomic nervous system differences among emotions. *Psychological Science, 3*, 23–27.
- Lyubomirsky, S., King, L. A., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success? *Psychological Bulletin, 131*, 803–855.
- Marsh, A. A., Finger, E. C., Mitchell, D. G. V., Reid, M. E., Sims, C., Kosson, D. S., ... Blair, R. J. R. (2008). Reduced amygdala response to fearful expressions in children and adolescents with callous-unemotional traits and disruptive behavior disorders. *American Journal of Psychiatry, 165*, 712–720.
- Martin, L. R., Friedman, H. S., Tucker, J. S., Tomlinson-Keasey, C., Criqui, M. H., & Schwartz, J. E. (2002). A life course perspective on childhood cheerfulness and its relation to mortality risk. *Personality and Social Psychology Bulletin, 28*, 1155–1165.
- Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H., & Gross, J. J. (2005). The tie that binds? Coherence among emotion experience, behavior, and autonomic physiology. *Emotion, 5*, 175–190.
- Mauss, I. B., Savino, N. S., Anderson, C. L., Weisbuch, M., Tamir, M., & Laudenslager, M. L. (2012a). The pursuit of happiness can be lonely. *Emotion, 12*, 908–912.

- Mauss, I. B., Shallcross, A. J., Troy, A. S., John, O. P., Ferrer, E., Wilhelm, F. H., & Gross, J. J. (2011). Don't hide your happiness! Positive emotion dissociation, social connectedness, and psychological functioning. *Journal of Personality and Social Psychology, 100*, 738–748.
- Mauss, I. B., Tamir, M., Anderson, C. L., & Savino, N. S. (2011). Can seeking happiness make people unhappy? Paradoxical effects of valuing happiness. *Emotion, 11*, 807–815.
- Oishi, S., Diener, E., & Lucas, R.E. (2007). Optimal level of well-being: Can people be too happy? *Perspectives on Psychological Science, 2*, 346–360.
- Patrick, C. J. (1994). Emotion and psychopathy: Startling new insights. *Psychophysiology, 31*,319–330.
- Peeters, F., Berkhof, J., Delespaul, P., Rottenberg, J., & Nicolson, N. A. (2003). Diurnal mood variation in major depressive disorder. *Emotion, 6*, 383–391.
- Rottenberg, J., Gross, J. J., & Gotlib, I. H. (2005). Emotion context insensitivity in major depressive disorder. *Journal of Abnormal Psychology, 114*, 627–639.
- Rottenberg, J., Kasch, K. L., Gross, J. J., & Gotlib, I. H. (2002). Sadness and amusement reactivity differentially predict concurrent and prospective functioning in major depressive disorder. *Emotion, 2*, 135–146.
- Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology, 76*, 805-819.
- Solomon, B. C. (2007). *True to our feelings: What our emotions are really telling us*. New York, NY: Oxford University Press.
- Srivastava, S., Tamir, M., McGonigal, K. M., John, O. P., & Gross, J. J. (2009). The social costs of emotional suppression: A prospective study of the transition to college. *Journal of Personality and Social Psychology, 96*, 883–897.
- Tamir, M., & Ford, B. Q. (2012a). Should people pursue feelings that feel good or feelings that do good? Emotional preferences and well-being. *Emotion, 12*, 1061–1070.
- Tamir, M., & Ford, B. Q. (2012b). When feeling bad is expected to be good: Emotion regulation and outcome expectancies in social conflicts. *Emotion, 12*, 807–816.
- Trull, T. J., Solhan, M. B., Tragesser, S. L., Jahng, S., Wood, P. K., Piasecki, T. M., & Watson, D. (2008). Affective instability: Measuring a core feature of borderline personality disorder with ecological momentary assessment. *Journal of Abnormal Psychology, 117*, 647–661.
- Tugade, M., Fredrickson, B. L., & Barrett, L. F. (2004). Psychological resilience and positive emotional granularity: Examining the benefits of positive emotions on coping and health. *Journal of Personality, 72*, 1161–1190.
- Tugade, M. M., & Fredrickson, L. L. (2004). Resilient individuals use positive emotions to bounce back from negative emotional experiences. *Journal of Personality and Social Psychology, 86*, 320–333.
- Veenhoven, R. (2008). Healthy happiness: Effects of happiness on physical consequences for preventive health care. *Journal of Happiness Studies, 9*, 449–469.
- van Kleef, E., van Dijk, G., Steinel, W., & van Beest, I. (2008). A social functional approach to emotions in bargaining: When communicating anger pays and when it backfires. *Journal of Personality and Social Psychology, 94*, 600–614.
- Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108*, 483–522.

16.2.14 Attributions

Emotion Experience and Well-Being by Brett Ford, University of Toronto, and Iris Mauss, University of California, Berkeley, through Noba Project, licensed CC BY-NC-SA 4.0

This page titled 16.2: Valence of Emotions is shared under a mixed license and was authored, remixed, and/or curated by ASCCC OERI & Bakhtawar Bhadha (ASCCC Open Educational Resources Initiative (OERI)) .

16.3: Fight or Flight

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Compare different theories about the generation of emotions.
- Describe the anatomical and chemical bases of anger and fear.
- Analyze how epigenetic research has enhanced our understanding of the nature-nurture debate surrounding anger and fear.

16.3.1 Overview

An important function of emotions in humans, and other animals that live in groups of any kind, is to communicate with other members of one's own species. The emotions that are particularly important to communicate to other members are those related to fear, and those related to anger. When we think of the physiological arousal associated with emotions, we often talk about the "fight or flight" response. These correspond nicely with the two major negative emotions - anger and fear. Multiple animal paradigm studies and human neuroscientific research, including studies with psychopathological conditions, have served to examine the nature of these emotions. Some of these theories and findings will be discussed.

16.3.2 Theories of Emotional Distinctions

Can emotions be better described as *qualitatively* distinct, for example, as discrete “basic emotions” or “natural kinds” (Ekman et al., 1983; Izard, 1992; Panksepp, 2005) or as *quantitatively* distinct, for example, as points along a circumplex defined by dimensions like arousal and valence (Russell and Barrett, 1999; Barrett and Wager, 2006)? Recent years have seen a protracted debate in the literature about how to most accurately capture the nature of emotion (Barrett et al., 2007; Izard, 2007; Panksepp, 2007; Tracy and Randles, 2011), with proposed models of emotion including not only basic emotion and dimensional models, but also those that focus upon goal-relevant appraisals of emotional stimuli (Moors et al., 2013), emotions as coping responses (Roseman, 2013), and emotions as survival circuits (LeDoux, 2012). An extended conversation about the strengths and weaknesses of these various views will not be reviewed in full here, rather, the focus will be on the basic consideration of whether different emotions (e.g., fear, anger) are best viewed as qualitatively or quantitatively distinct.

16.3.2.1 Qualitative Ideas



Figure **16.3.1**: Individual emotions as they develop - first we only have basic emotions (fear, anger, disgust, contempt, joy, sadness and interest). Then self conscious (guilt, pride, embarrassment, shame, triumph) ones devleop. And then finally more cognitively complex (envy, gratitude, disappointment, regret, hope, schadenfreude, empathy, compassion) ones appear.(Copyright; Individual emotions by U3161650, CC BY-SA 4.0, via Wikimedia Commons)

Models that posit emotions to be qualitatively distinct, such as “basic emotion” models, holds that a limited number of emotions like fear, anger, and positive excitement emerge from dissociable neurophysiological processes (Ekman et al., 1983; Izard, 1992; Panksepp, 2005; Lench et al., 2011). In Figure **16.3.1** we see one representation of this idea that initially we have basic emotions including fear, anger, disgust, contempt, joy, sadness and interest, which then grow to include self conscious emotions (guilt, pride, embarrassment, shame, triumph) and finally cognitively complex emotions (envy, gratitude, disappointment, regret, hope, schadenfreude, empathy, compassion). These neurophysiological processes are generally linked to activity in the evolutionarily ancient subcortical structures of the midbrain, striatum, and limbic system most commonly linked to emotion (Panksepp, 2005;

Vytal and Hamann, 2010). So, for example, the generation of positive excitement is linked to activation in a striatal circuit centered on dopaminergic neurons in the nucleus accumbens (Ikemoto and Panksepp, 1999), whereas the generation of fear is associated with activity in a circuit involving the periaqueductal gray, anterior and medial hypothalamus, and amygdala (LeDoux, 2000). In this view, finer gradations of experience result when basic emotions are modulated or elaborated by higher-level cognitive processes controlled by the cerebral cortex, but the emergence of qualitatively distinct emotions is not dependent on these cortically-controlled processes (Panksepp, 2005).

16.3.2.2 Quantitative Ideas

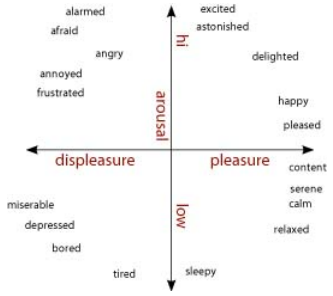


Figure **16.3.2**: Model showing two orthogonal dimensions of valence (on the horizonal axis) and arousal (on the vertical axis) with examples of emotions and where they might fall on the graph. For example, delight falls in the top right quadrant because it indicates high arousal and high pleasure while boredom falls in the diagonally opposite quadrant, and relaxed in the bottom right (Copyright; Psychologist Russell's model of arousal and valence by http://imagine-it.org/gamessurvey/, licensed CC BY 3.0, via Wikimedia Commons)

Models that posit emotions to be quantitatively distinct hold that emotions like fear, anger, and happiness are best described as points on one or more core dimensions. Core dimensions typically proposed to distinguish among emotions are physiological arousal or activation (low—high) and valence (bad—good) (Bradley et al., 2001). [Some have proposed a withdrawal—approach dimension as a substitute or supplement to the valence axis (Wager et al., 2003; Christie and Friedman, 2004; van Honk and Schutter, 2006)]. As shown in figure **16.3.2**, arranged orthogonally, these dimensions form a circumplex upon which emotions can be plotted and quantitatively compared (Barrett and Russell, 1999; Russell and Barrett, 1999; Colibazzi et al., 2010). Positive excitement is plotted as high in arousal and positive in valence, and sadness is low in arousal and negative in valence. Fear is typically plotted as high arousal and strongly negative, as is anger (Russell and Barrett, 1999). Further distinctions among emotions are thought to reflect differences in cognitive construals of the events surrounding the basic changes in arousal and valence. Thus, whether an individual experiences anger or fear (which are similar in terms of arousal or valence) may be shaped by interpretations of neurophysiological changes in valence and arousal in light of the eliciting stimulus and the individual's idiosyncratic stores of semantic knowledge, memories, and behavioral responses that shape the subjectively experienced state (Russell, 2003). Under this view, distinctions among experienced emotional states are highly dependent on these cognitively complex processes, which are subserved by a distributed network of regions of the cerebral cortex (Lindquist et al., 2012).

These models generate distinct predictions to the question of whether a disorder or lesion could result in a single emotion being disabled without affecting the experience of other emotions. The discrete emotions view would argue that a disorder or lesion that resulted in dysfunction in the specific structures subserving a particular emotion could affect the experience of one emotion while leaving others intact. In contrast, the dimensional view would require either that other emotions that are dimensionally similar to the affected emotion also be affected, or that deficits in a particular emotion would reflect dysfunction in cortically-driven higher-level cognitive processes.

16.3.2.3 Psychopathy Supports the Qualitative Model

The case of psychopathy lends clear support to notion that fear is qualitatively distinct from other emotions. In psychopathy, the bulk of the clinical and empirical evidence points toward the conclusion that fear responding is uniquely disabled, with other high-arousal (positive excitement, anger) and negatively valenced (anger, disgust) emotions remaining intact. The dimensional view cannot easily explain why in psychopaths the high arousal, negatively valenced state of anger is easily (perhaps too easily) generated, whereas the high arousal, negatively valenced state of fear is not. The problem cannot lie in a failure to fully engage neurocognitive systems underlying either the arousal or valence dimension, because psychopaths experience other high-arousal emotions (positive excitement) as well as other negatively valenced emotions (disgust). It also cannot result from some difficulty arising at the interaction of these axes, because anger and fear are highly similar in terms of both dimensions. Models that

substitute a withdrawal—approach axis for a negative—positive axis are no more successful; the two most strongly withdrawal-linked emotions are disgust and fear, and there is no evidence for disgust-based impairments in psychopathy. Individuals with psychopathy also fail to recognize and therefore have no empathic response to others' fear.

On the whole, the empirical data support the idea that the amygdala, along with its efferent projections, is an essential structure for the generation of conditioned fear responses, which account for the majority of experienced fear (Davis, 1992, 1997). Extensive early evidence demonstrated that the amygdala plays a crucial role in the creation of conditioned fear in rodents. For example, lesions to the amygdala prevent rats from developing a conditioned fear response, like freezing in response to a stimulus that predicts shock (Blanchard and Blanchard, 1972). Later studies clarified the roles of the various subnuclei of the amygdala, demonstrating that the lateral nucleus is primarily involved in the acquisition of the fear response whereas the central nucleus is involved in both the acquisition and the expression of conditioned fear responses (Davis, 1992; Wilensky et al., 2006). The amygdala's many efferent projections coordinate autonomic and behavioral responses to fear eliciting stimuli. Projections from the central nucleus of the amygdala to the lateral hypothalamus are involved in activating autonomic sympathetic nervous system responses, and projections to the ventrolateral periaqueductal gray direct the expression of behavior responses, such as defensive freezing (Davis, 1992; LeDoux, 2012). The amygdala's central role in coordinated fear responding can be demonstrated by electrical stimulation studies showing that complex patterns of behavioral and autonomic changes associated with fear responses result from stimulation of the relevant regions of the amygdala (Davis, 1992). Heavy reliance on animal models is justified in the study of fear responding and the amygdala given how strongly conserved the amygdala nuclei involved in responding to conditioned threats are across species ranging from reptiles to birds to rodents to primates (LeDoux, 2012).

16.3.3 Research Suggesting Connections Between Anger and Fear

Neumann et al. (2010) hypothesize that aggressive behaviors, that are of two basic types - reactive and proactive, are mediated by anxiety-based neurological bases. Based on animal models, they suggest the following: "Male aggression is necessary for the acquisition and maintenance of nutrition, territory, and mating partners. Species-specific rules have to be strictly obeyed to guarantee effective and harmless communication. Thus, adaptive offensive aggression is comprised primarily of harmless threat behaviors allowing the opponent to escape or to switch to submissive behaviors in order to avoid direct physical confrontation. In rodents, such signs of offensive aggression include piloerection (intimidation of the opponent by larger appearance) and lateral threat (arched back and exposure of the flank). In case of an offensive attack, less vulnerable body parts of the opponent, such as those covered with muscles and a thick layer of skin, are targeted to avoid serious injuries (Blanchard and Blanchard, 1977 ; Blanchard et al., 2003). While offensive aggression is usually expressed during a fight for territory or exclusive mating, defensive aggression is mainly displayed in life-threatening situations and is linked to increased fear (Blanchard and Blanchard, 1981). As opposed to offensive aggression, defensive aggression is less or not signaled in advance, and attack targets include more vulnerable body parts (such as the head, belly, and genitals) (Blanchard and Blanchard, 1977 ; Blanchard et al., 2003)." The resident-intruder animal paradigms have been used to measure these ideas (see Figure **16.3.3**).



Figure **16.3.3**: Resident mouse with aggressive behavior and intruder mouse. Image by DataBase Center for Life Science (DBCLS) licensed CC-BY 4.0 via Wikimedia commons.

They also claim: "Anxiety may be interpreted as an emotional anticipation of an aversive situation and is reflected by species-specific behavioural fear responses to stressful and threatening stimuli characteristic for individual trait anxiety. Fear is not seen as basal state (as is anxiety), but as a complex behavioural response, such as startle or freezing. Further, in addition to factors which determine innate (trait) anxiety, several environmental or pharmacological factors may interact with the genetic background and determine the individual level of state anxiety and the final behavioural phenotype. Emotionality, often used as synonym for anxiety as well as fearfulness, may be seen in a broader sense, comprising both trait and state anxiety and stimulus-related fear. Emotionality is one of the major components underlying the ability of an organism to assess stressful stimuli and scenarios, and to adequately cope with them." (Neumann et al., 2010)

Neumann et al. (2010) show that several clinical and laboratory studies with humans and rodents have shown that there are complicated mechanisms by which anxiety and aggression are co-regulated. There are several neurochemicals found to be involved in complex ways as follows:

Glucocorticoids - The regulation of the HPA (Hypothalamic-Pituitary-Adrenal) axis seems very much connected to the experience of anxiety and aggression. However, the direction is not straightforward in that both high and low glucocorticoids are related to high aggression.

Vasopressin - Vasopressin is produced in the hypothalamus, and testosterone is produced in the gonads, and both modulate aggression as well as anxiety in males. Vasopressin is also involved in reducing anxiety as well as pair-bonding.

Testosterone - Along with aggression, testosterone rises in puberty, castration reduces both (through reduction in Vasopressin also). There are some environmental factors that have a moderating effect, but in general, testosterone is clearly related to modulating aggression in males.

Serotonin - lower 5-HT seems linked to greater aggression and violence in female rhesus. Genetic factors and early life stressors tend to mediate these effects.

GABA - when anxiety levels are increased, aggression increased depending on genetic predispositions to it, and activity also increased in "brain regions including the central amygdala, BST (Bed nucleus of the stria terminalis), lateral septum, and PVN (paraventricular nucleus) that are associated with stress-, fear- and aggression-related behaviour."

"The striking evidence for an overlap in neuroendocrine and neurochemical systems regulating aggression as well as anxiety suggests a strong correlation between these two behaviours. Thus, aggression and anxiety are not always co-regulated, but, under some circumstances, these behaviours may come under the control of the same genes and neuroactive substances including sexual steroids, neuropeptides and neuroamines within specific brain circuitries. Such a view is in agreement with clinical findings. On the one hand, excessive and violent behaviours are seen in humans exposed to adverse early life experiences and in patients with

depression- and anxiety-related disorders, or PTSD. On the other hand, rather conflicting data exist on the effects of anxiolytic drugs on anti-social and aggressive behaviours. In future studies that focus on the neurobiological mechanisms of (co-)regulation of aggression and anxiety, epigenetic modifications need to be considered in addition to the neuronal and neuroendocrine parameters discussed above." (Neumann et al, 2010).

16.3.4 Nature v. Nurture

Both genetic factors and experiences play a role in aggressive behaviors. The genetic regulation of serotonin release appears to play a role in aggression in many mammals including humans. For example, Peeters et al.'s (2020) study concluded that the short version of the serotonin transporter gene (S-allele of the 5-HTTLPR) seems to be linked to greater reactive aggression, and those individuals with higher avoidance tendencies toward angry facial expressions. Their findings indicate that evaluative impulses in response to social cues play an important role in mediating the genetic predisposition of the 5-HTTLPR polymorphism to increased expression of reactive aggression. Because of the chief role of MAOs in the metabolism of key neurotransmitters involved in aggressive behavior, most notably serotonin, it is not surprising that a substantial body of research has linked aggressive phenotypes with the MAO system. Many results inspired both the scientific community and media outlets to refer to *MAO-A* as the *warrior* or *criminal* gene. Although initial pharmacological studies supported a role for MAOIs in the reduction of aggressive phenotypes, data from these studies were hard to interpret because of the side effects of MAOIs, and because of their impact on a myriad of unrelated behaviors. Substantive evidence supporting the role of the *MAO* genes in aggression comes from studies in knockout (KO) mice. Selective knockout models for the *MAO-A* gene, for instance, exhibited increased aggressiveness compared to their wild-type counterparts. Genetic studies in humans provide evidence that *MAO-A* is linked to aggression but only when other environmental factors are also present during development (e.g., abuse, stressors). Other lines of research have demonstrated that the gene product of *MAO-A* (rather than the gene per se) influences violent traits. For instance, cortical and subcortical MAO-A activity in vivo -measured with positron emission tomography (PET)- was negatively associated with trait aggression (attribution: Mentis, Dardiotis, Katsouni & Chrousos, 2021).

Over the last two decades, the study of the relationship between *nature* and *nurture* in shaping human behavior has encountered a renewed interest. Behavioral genetics showed that distinct polymorphisms of genes that code for proteins that control neurotransmitter metabolic and synaptic function are associated with individual vulnerability to aversive experiences, such as stressful and traumatic life events, and may result in an increased risk of developing psychopathologies associated with violence. On the other hand, recent studies indicate that experiencing aversive events modulates gene expression by introducing stable changes to DNA without modifying its sequence, a mechanism known as “epigenetics”. For example, experiencing adversities during periods of maximal sensitivity to the environment, such as prenatal life, infancy and early adolescence, may introduce lasting epigenetic marks in genes that affect maturational processes in brain, thus favoring the emergence of dysfunctional behaviors, including exaggerate aggression in adulthood.

Adverse environmental influences during critical periods of development have been correlated with epigenetic markers that affect glucocorticoid receptor function. As seen in many studies, low cortisol release is correlated with low self control and higher aggression. Production of oxytocin is related to higher social functioning and attachment formation. Oxytocin secretion is stimulated by early maternal care, whereas adverse experiences in prenatal and early infancy development are related to lower oxytocin receptor methylation (a process involved in epigenetics). Serotonin is related to regulation of aggression. Genes responsible for controlling the reuptake of serotonin from the synaptic cleft are affected by early stressors (including trauma as well as conditions generated by poverty like poor medical care, housing quality and exposure to violent neighborhoods) that alter brain anatomy and function such as cortical thickness and amygdala reactivity. (See Figure **16.3.4** for the multiple ways in which stress and traumas lead to epigenetic modifications in the oxytocin, serotonin and HPA axis systems that affect aggressive behavior: reactive and proactive aggression, hostility, delinquency externalizing problems, violence, conduct disorders, physical and verbal aggression and callous unemotional traits.)

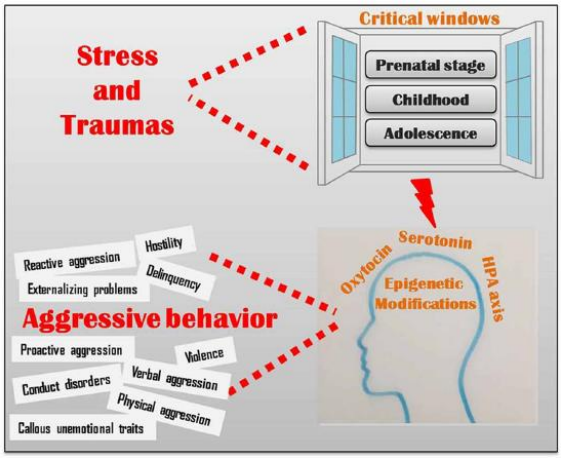


Figure **16.3.4**: The epigenetic control of biological pathways, including the neuroendocrine, serotonergic and oxytocinergic pathways, significantly mediates the behavioral responses to the environment (Copyright; Genes and Aggressive Behavior: Epigenetic Mechanisms Underlying Individual Susceptibility to Aversive Environments by Sara Palumbo, Veronica Mariotti, Caterina Ioffreda & Silvia Pellegrini in Frontiers in Behavioral Neuroscience (2018) licensed CC BY 4.0)

Palumbo, Mariotti, Ioffreda and Pellegrini (2018) conclude that epigenetics is shedding a new light on the fine interaction between *nature* and *nurture*, by providing a novel tool to understand the molecular events that underlie the relationship among genes, brain, environment and behavior. Altogether, the results of the studies that we briefly discussed in the present article, clearly indicate that, when it comes to (human) behavior, *nature* and *nurture* are not to be regarded as two distinct and separate factors, contrary to the alternating predominance of either one that has been proposed in different historic phases (Levitt, 2013; Moore, 2016). Indeed, distinct genetic backgrounds differentially modulate the individual susceptibility to the environment and at the same time various environmental conditions differentially affect gene expression, in an intimate and fascinating manner that scientists have now begun to disentangle. The findings from this research pave the way to a novel approach to the understanding of human behavior, with important implications also for social sciences, including philosophy, ethics and law.

16.3.5 Conclusions

Anger and fear are considered basic emotions. While they are primarily negative, they serve very important survival functions. They appear to have clear genetic and biological bases that have been discussed in the research above.

16.3.6 References

- Barrett, L. F., Lindquist, K. A., Bliss-Moreau, E., Duncan, S., Gendron, M., Mize, J., et al. (2007). Of mice and men: natural kinds of emotions in the mammalian brain? A response to Panksepp and Izard. Perspect. Psychol. Sci. 2, 297–312. doi: 10.1126/science.7652558
- Barrett, L. F., and Russell, J. A. (1999). The structure of current affect. Curr. Dir. Psychol. Sci. 8, 10.
- Barrett, L. F., and Wager, T. D. (2006). The structure of emotion: Evidence from neuroimaging studies. Curr. Dir. Psychol. Sci. 5, 79–83.
- Blanchard, D. C., and Blanchard, R. J. (1972). Innate and conditioned reactions to threat in rats with amygdaloid lesions. J. Comp. Physiol. Psychol. 81, 281–290.
- Bradley, M. M., Codispoti, M., Cuthbert, B. N., and Lang, P. J. (2001). Emotion and motivation I: defensive and appetitive reactions in picture processing. Emotion 1, 276–298. https://doi.org/10.1037/1528-3542.1.3.276
- Buckholtz, J. W. & Meyer-Lindenberg, A. (2008). MAOA and the neurogenetic architecture of human aggression. Trends Neurosci. https://doi.org/10.1016/j.tins.2007.12.006
- Cases, O. et al. (1995). Aggressive behavior and altered amounts of brain serotonin and norepinephrine in mice lacking MAOA. Science (80-) https://doi.org/10.1126/science.7792602
- Christie, I. C., and Friedman, B. H. (2004). Autonomic specificity of discrete emotion and dimensions of affective space: a multivariate approach. Int. J. Psychophysiol. 51, 143–153. https://doi.org/10.1016/j.ijpsycho.2003.08.002
- Colibazzi, T., Posner, J., Wang, Z., Gorman, D., Gerber, A., Yu, S., et al. (2010). Neural systems subserving valence and arousal during the experience of induced emotions. Emotion 10, 377–389. https://doi.org/10.1037/a0018484

- Davis, M. (1992). The role of the amygdala in fear and anxiety. Annu. Rev. Neurosci. 15, 353–375. doi/10.1146/annurev.ne.15.030192.002033
- Davis, M. (1997). Neurobiology of fear responses: the role of the amygdala. J. Neuropsychiatry Clin. Neurosci. 9, 382–402. https://doi.org/10.1176/jnp.9.3.382
- Ekman, P., Levenson, R. W., and Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. Science 221, 1208–1210. doi: 10.1126/science.6612338
- Feinstein, J. S., Buzza, C., Hurlemann, R., Follmer, R. L., Dahdaleh, N. S., Coryell, W. H., et al. (2013). Fear and panic in humans with bilateral amygdala damage. Nat. Neurosci. 16, 270–222. https://doi.org/10.1016/j.cub.2010.11.042
- Ficks, C. A. & Waldman, I. D. (2014). Candidate genes for aggression and antisocial behavior: a meta-analysis of association studies of the 5HTTLPR and MAOA-uVNTR. Behav. Genet. https://doi.org/10.1007/s10519-014-9661-y
- Godar, S. C., Fite, P. J., McFarlin, K. M. & Bortolato, M. (2016). The role of monoamine oxidase A in aggression: current translational developments and future challenges. Prog. Neuro-Psychopharmacol. Biol. Psychiatry https://doi.org/10.1016/j.pnpbp.2016.01.001
- Ikemoto, S., and Panksepp, J. (1999). The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. Brain Res. Brain Res. Rev. 31, 6–41. https://doi.org/10.1016/S0165-0173(99)00023-5
- Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. Psychol. Rev. 99, 561–565. https://doi.org/10.1037/0033-295X.99.3.561
- Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. Perspect. Psychol. Sci. 2, 260–280.
- Johnson, P. L., Fitz, S. D., Hollis, J. H., Moratalla, R., Lightman, S. L., Shekhar, A., et al. (2011). Induction of c-Fos in “panic/defence”-related brain circuits following brief hypercarbic gas exposure. J. Psychopharmacol. 25, 26–36. https://doi.org/10.1177/0269881109353464
- LeDoux, J. E. (2000). Emotion circuits in the brain. Annu. Rev. Neurosci. 23, 155–184. https://doi.org/10.1146/annurev.neuro.23.1.155
- LeDoux, J. (2012). Rethinking the emotional brain. Neuron 73, 653–676. https://doi.org/10.1016/j.neuron.2012.02.004
- Lench, H. C., Flores, S. A., and Bench, S. W. (2011). Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: a meta-analysis of experimental emotion elicitations. Psychol. Bull. 137, 834–855. https://doi.org/10.1037/a0024244
- Levitt, M. (2013). Perceptions of nature, nurture and behaviour. Life Sci. Soc. Policy, 9(13). doi: 10.1186/2195-7819-9-13
- Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., and Barrett, L. F. (2012). The brain basis of emotion: a meta-analytic review. Behav. Brain. Sci. 35, 121–143. doi:10.1017/S0140525X11000446
- Moore, D. S. (2016). Behavioral epigenetics. Wiley Interdiscip. Rev. Syst. Biol. Med. 9:e1333. doi: 10.1002/wsbm.1333
- Moors, A., Ellsworth, P. C., Scherer, K. R., and Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. Emot. Rev. 5, 119–124.
- Panksepp, J. (2005). Affective consciousness: core emotional feelings in animals and humans. Conscious. Cogn. 14, 30–80. https://doi.org/10.1016/j.concog.2004.10.004
- Neumann I.D., Veenema, A.H. & Beiderbeck, D.I. (2010) Aggression and anxiety: social context and neurogiological links. *Front. Behav. Neurosci*. *4*(12). doi: 10.3389/fnbeh.2010.00012
- Panksepp, J. (2007). Neurologizing the psychology of affects. Perspect. Psychol. Sci. 2, 281–296.
- Peeters, D. G. A., Lange, W.-G., von Borries A. K. L., Franke, B., Volman I., Homberg J. R., Verkes R.-J., Roelofs, K. (2020). Threat-avoidance tendencies moderate the link between serotonin transporter genetic variation and reactive aggression. Frontiers in Behavioral Neuroscience, 14, 172. doi:10.3389/fnbeh.2020.562098
- Popova, N. K. (2006). From genes to aggressive behavior: the role of serotonergic system. BioEssays https://doi.org/10.1002/bies.20412
- Roseman, I. (2013). Appraisal in the emotion system: coherence in strategies for coping. Emot. Rev. 5, 141–149.
- Ruisch, I. H., Dietrich, A., Glennon, J. C., Buitelaar, J. K. & Hoekstra, P. J. (2019). Interplay between genome-wide implicated genetic variants and environmental factors related to childhood antisocial behavior in the UK ALSPAC cohort. Eur. Arch. Psychiatry Clin. Neurosci. https://doi.org/10.1007/s00406-018-0964-5
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. Psychol. Rev. 110, 145–172. https://doi.org/10.1037/0033-295X.110.1.145

- Russell, J. A., and Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. J. Pers. Soc. Psychol. 76, 805–819. https://doi.org/10.1037/0022-3514.76.5.805
- Shih, J. C. & Chen, K. (1999). MAO-A and -B gene knock-out mice exhibit distinctly different behavior. Neurobiology, 7, 235–246 (1999).
- Sohrabi, S. (2015). The criminal gene: the link between MAOA and aggression (REVIEW). BMC Proc. https://doi.org/10.1186/1753-6561-9-s1-a49
- Takahashi, A., Quadros, I. M., de Almeida, R. M. M. & Miczek, K. A. (2012). Behavioral and pharmacogenetics of aggressive behavior. Curr. Top. Behav. Neurosci. https://doi.org/10.1007/7854_2011_191 Sohrabi, S. (2015). The criminal gene: the link between MAOA and aggression (REVIEW). BMC Proc. https://doi.org/10.1186/1753-6561-9-s1-a49
- Takahashi, A., Quadros, I. M., de Almeida, R. M. M. & Miczek, K. A. (2012). Behavioral and pharmacogenetics of aggressive behavior. Curr. Top. Behav. Neurosci. https://doi.org/10.1007/7854_2011_191
- Tracy, J. L., and Randles, D. (2011). Four models of basic emotions: a review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. Emot. Rev. 3, 397–405.
- van Honk, J., and Schutter, D. J. L. G. (2006). From affective valence to motivational direction The frontal asymmetry of emotion revised. Psychol. Sci. 17, 963–965. https://doi.org/10.1111/j.1467-9280.2006.01813.x
- Vassos, E., Collier, D. A. & Fazel, S. (2014). Systematic meta-analyses and field synopsis of genetic association studies of violence and aggression. Mol. Psychiatry https://doi.org/10.1038/mp.2013.31
- Vytal, K., and Hamann, S. (2010). Neuroimaging support for discrete neural correlates of basic emotions: a voxel-based meta-analysis. J. Cogn. Neurosci. 22, 2864–2885. https://doi.org/10.1162/jocn.2009.21366
- Wager, T. D., Phan, K. L., Liberzon, I., and Taylor, S. F. (2003). Valence, gender, and lateralization of functional brain anatomy in emotion: a meta-analysis of findings from neuroimaging. Neuroimage 19, 513–531. https://doi.org/10.1016/S1053-8119(03)00078-8
- Waltes, R., Chiocchetti, A. G. & Freitag, C. M. (2016). The neurobiological basis of human aggression: a review on genetic and epigenetic mechanisms. Am. J. Med. Genet. Part B Neuropsychiatr.Genet. https://doi.org/10.1002/ajmg.b.32388
- Wilensky, A. E., Schafe, G. E., Kristensen, M. P., and LeDoux, J. E. (2006). Rethinking the fear circuit: the central nucleus of the amygdala is required for the acquisition, consolidation, and expression of Pavlovian fear conditioning. J. Neurosci. 26, 12387–12396. https://doi.org/10.1523/JNEUROSCI.4316-06.2006
- Youdim, M. B. H., Edmondson, D. & Tipton, K. F. (2006). The therapeutic potential of monoamine oxidase inhibitors. Nat. Rev. Neurosci. https://doi.org/10.1038/nrn1883

16.3.7 Attributions

What can we learn about emotions by studying psychopathy by Abigail Marsh, in Frontiers in Human Neuroscience licensed CC BY 3.0

Psychologist Russell's model of arousal and valence by http://imagine-it.org/gamessurvey/, licensed CC BY 3.0, via Wikimedia Commons

Individual emotions by U3161650, CC BY-SA 4.0, via Wikimedia Commons

Image by DataBase Center for Life Science (DBCLS) licensed CC-BY 4.0 via Wikimedia commons.

From *warrior genes* to translational solutions: novel insights into monoamine oxidases (MAOs) and aggression. by Alexios-Fotios A. Mentis, Efthemios Dardiotis, Eleni Katsouni, George P. Chrousos. in Transl Psychiatry licensed CC-BY 4.0

Genes and Aggressive Behavior: Epigenetic Mechanisms Underlying Individual Susceptibility to Aversive Environments by Sara Palumbo, Veronica Mariotti, Caterina Ioffreda & Silvia Pellegrini in Frontiers in Behavioral Neuroscience (2018) licensed CC BY 4.0

This page titled 16.3: Fight or Flight is shared under a mixed license and was authored, remixed, and/or curated by ASCCC OERI & Bakhtawar Bhadha (ASCCC Open Educational Resources Initiative (OERI)) .

16.4: Social Neuroscience

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Analyze the neural and physiological bases of social behavior including brain areas that are associated with social tasks.
- Evaluate how social behaviors and phenomena affect biological systems.

16.4.1 Overview

The current section provides an overview of the new field of social neuroscience, which combines the use of neuroscience methods and theories to understand how other people influence our thoughts, feelings, and behavior. The module reviews research measuring neural and hormonal responses to understand how we make judgments about other people and react to stress. Through these examples, it illustrates how social neuroscience addresses three different questions: (1) how our understanding of social behavior can be expanded when we consider neural and physiological responses, (2) what the actual biological systems are that implement social behavior (e.g., what specific brain areas are associated with specific social tasks), and (3) how biological systems are impacted by social processes.

Psychology has a long tradition of using our brains and body to better understand how we think and act. For example, in 1939 Heinrich Kluver and Paul Bucy removed (i.e. lesioned) the temporal lobes in some rhesus monkeys and observed the effect on behavior. Included in these **lesions** was a subcortical area of the brain called the **amygdala**. After surgery, the monkeys experienced profound behavioral changes, including loss of fear. These results provided initial evidence that the amygdala plays a role in emotional responses, a finding that has since been confirmed by subsequent studies (Phelps & LeDoux, 2005; Whalen & Phelps, 2009).

16.4.2 What Is Social Neuroscience?

Social neuroscience similarly uses the brain and body to understand how we think and act, with a focus on how we think about and act toward other people. More specifically, we can think of social neuroscience as an interdisciplinary field that uses a range of neuroscience measures to understand how other people influence our thoughts, feelings, and behavior. As such, social neuroscience studies the same topics as social psychology, but does so from a multilevel perspective that includes the study of the brain and body. Figure **16.4.1** shows the scope of social neuroscience with respect to the older fields of social psychology and neuroscience. Although the field is relatively new – the term first appeared in 1992 (Cacioppo & Berntson, 1992) – it has grown rapidly, thanks to technological advances making measures of the brain and body cheaper and more powerful than ever before, and to the recognition that neural and physiological information are critical to understanding how we interact with other people.

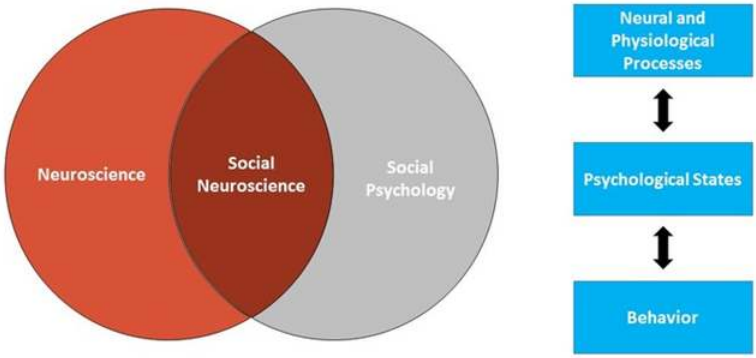


Figure 16.4.1: Social neuroscience is the intersection of social psychology and neuroscience. Under this multilevel approach, neural/physiological processes and behavior are two things we can measure or observe. Psychological states cannot be directly observed, but understanding them is the goal. Social neuroscientists use the observable neural/physiological processes and behavioral responses to make inferences about unobservable psychological states. The bidirectional arrows show that all levels of analysis are assumed to influence each other (e.g., psychological states can influence neural responses, and neural responses can influence psychological states). (Copyright; Tiffany Ito & Jennifer Kubota, via Noba Project, licensed CC BY-NC-SA 4.0)

Social neuroscience can be thought of as both a methodological approach (using measures of the brain and body to study social processes) and a theoretical orientation (seeing the benefits of integrating neuroscience into the study of social psychology). The overall approach in social neuroscience is to understand the psychological processes that underlie our social behavior. Because those psychological processes are intrapsychic phenomena that cannot be directly observed, social neuroscientists rely on a combination of measurable or observable neural and physiological responses as well as actual overt behavior to make inferences about psychological states (see Figure 16.4.1). Using this approach, social neuroscientists have been able to pursue three different types of questions: (1) What more can we learn about social behavior when we consider neural and physiological responses? (2) What are the actual biological systems that implement social behavior (e.g., what specific brain areas are associated with specific social tasks)? and (3) How are biological systems impacted by social processes?

16.4.3 How Automatically Do We Judge Other People?

Social categorization is the act of mentally classifying someone as belonging in a group. Why do we do this? It is an effective mental shortcut. Rather than effortfully thinking about every detail of every person we encounter, social categorization allows us to rely on information we already know about the person's group. For example, by classifying your restaurant server as a man, you can quickly activate all the information you have stored about men and use it to guide your behavior. But this shortcut comes with potentially high costs. The stored group beliefs might not be very accurate, and even when they do accurately describe some group members, they are unlikely to be true for every member you encounter. In addition, many beliefs we associate with groups – called **stereotypes** – are negative. This means that relying on social categorization can often lead people to make negative assumptions about others.

The potential costs of social categorization make it important to understand how social categorization occurs. Is it rare or does it occur often? Is it something we can easily stop, or is it hard to override? One difficulty answering these questions is that people are not always consciously aware of what they are doing. In this case, we might not always realize when we are categorizing someone. Another concern is that even when people are aware of their behavior, they can be reluctant to accurately report it to an experimenter. In the case of social categorization, subjects might worry they will look bad if they accurately report classifying someone into a group associated with negative stereotypes. For instance, many racial groups are associated with some negative stereotypes, and subjects may worry that admitting to classifying someone into one of those groups means they believe and use those negative stereotypes.



Figure 16.4.2: This man is wearing an elastic electrode cap into which individual electrodes (inside the white circles) are sewn into standardized locations. [Image: Hans, CC0 Public Domain, goo.gl/m25gce]

Social neuroscience has been useful for studying how social categorization occurs without having to rely on self-report measures, instead measuring brain activity differences that occur when people encounter members of different social groups. Much of this work has been recorded using the **electroencephalogram**, or **EEG**. EEG is a measure of electrical activity generated by the brain's neurons. Comparing this electrical activity at a given point in time against what a person is thinking and doing at that same time allows us to make inferences about brain activity associated with specific psychological states. One particularly nice feature of EEG is that it provides very precise timing information about when brain activity occurs. EEG is measured non-invasively with small electrodes that rest on the surface of the scalp. This is often done with a stretchy elastic cap, like the one shown in Figure 16.4.2, into which the small electrodes are sewn. Researchers simply pull the cap onto the subject's head to get the electrodes into place; wearing it is similar to wearing a swim cap. The subject can then be asked to think about different topics or engage in different tasks as brain activity is measured.

To study social categorization, subjects have been shown pictures of people who belong to different social groups. Brain activity recorded from many individual trials (e.g., looking at lots of different Black individuals) is then averaged together to get an overall idea of how the brain responds when viewing individuals who belong to a particular social group. These studies suggest that social categorization is an **automatic process** – something that happens with little conscious awareness or control – especially for dimensions like gender, race, and age (Ito & Urland, 2003; Mouchetant-Rostaing & Giard, 2003). The studies specifically show that brain activity differs when subjects view members of different social groups (e.g., men versus women, Blacks versus Whites), suggesting that the group differences are being encoded and processed by the perceiver. One interesting finding is that these brain changes occur both when subjects are purposely asked to categorize the people into social groups (e.g., to judge whether the person is Black or White), and also when they are asked to do something that draws attention away from group classifications (e.g., making a personality judgment about the person) (Ito & Urland, 2005). This tells us that we do not have to intend to make group classifications in order for them to happen. It is also very interesting to consider how quickly the changes in brain responses occur. Brain activity is altered by viewing members of different groups within 200 milliseconds of seeing a person's face. That is just two-tenths of a second. Such a fast response lends further support to the idea that social categorization occurs automatically and may not depend on conscious intention.

Overall, this research suggests that we engage in social categorization very frequently. In fact, it appears to happen automatically (i.e., without us consciously intending for it to happen) in most situations for dimensions like gender, age, and race. Since classifying someone into a group is the first step to activating a group stereotype, this research provides important information about how easily stereotypes can be activated. And because it is hard for people to accurately report on things that happen so quickly, this issue has been difficult to study using more traditional self-report measures. Using EEGs has, therefore, been helpful in providing interesting new insights into social behavior.

16.4.4 Do We Use Our Own Behavior to Help Us Understand Others?

Classifying someone into a social group then activating the associated stereotype is one way to make inferences about others. However, it is not the only method. Another strategy is to imagine what our own thoughts, feelings, and behaviors would be in a similar situation. Then we can use our simulated reaction as a best guess about how someone else will respond (Goldman, 2005).

After all, we are experts in our own feelings, thoughts, and tendencies. It might be hard to know what other people are feeling and thinking, but we can always ask ourselves how we would feel and act if we were in their shoes.

There has been some debate about whether simulation should/could be used to get into the mind of others or is an effective way to get into the minds of others (Carruthers & Smith, 1996; Gallese & Goldman, 1998). Social neuroscience research has addressed this question by looking at the brain areas used when people think about themselves and others. If the same brain areas are active for the two types of judgments, it lends support to the idea that the self may be used to make inferences about others via simulation.

We know that an area in the prefrontal cortex called the **medial prefrontal cortex** (mPFC) – located in the middle of the frontal lobe – is active when people think about themselves (Kelley, Macrae, Wyland, Caglar, Inati, & Heatherton, 2002). This conclusion comes from studies using **functional magnetic resonance imaging**, or fMRI. While EEG measures the brain's electrical activity, fMRI measures changes in the oxygenation of blood flowing in the brain. Remember as discussed in chapter 2, when neurons become more active, blood flow to the area increases to bring more oxygen and glucose to the active cells. fMRI allows us to image these changes in oxygenation by placing people in an fMRI machine or scanner (Figure **16.4.3**), which consists of large magnets that create strong magnetic fields. The magnets affect the alignment of the oxygen molecules within the blood (i.e., how they are tilted). As the oxygen molecules move in and out of alignment with the magnetic fields, their nuclei produce energy that can be detected with special sensors placed close to the head. Recording fMRI involves having the subject lay on a small bed that is then rolled into the scanner. While fMRI does require subjects to lie still within the small scanner and the large magnets involved are noisy, the scanning itself is safe and painless. Like EEG, the subject can then be asked to think about different topics or engage in different tasks as brain activity is measured. If we know what a person is thinking or doing when fMRI detects a blood flow increase to a particular brain area, we can infer that part of the brain is involved with the thought or action. fMRI is particularly useful for identifying which particular brain areas are active at a given point in time.



Figure **16.4.3**: Functional magnetic resonance imaging (fMRI) scanner used to image the brain while people perform tasks. The scanner allows researchers to view the changes in blood oxygenation in specific locations in the brain during a task. Images are collected using powerful magnets and radio waves that shift the position of atoms in oxygenated blood that rushes to areas involved in performing the task. MRIs are non-invasive and there are no known risks from exposure to the magnetic fields or radio waves. Subjects lay on the while bed with their head inside the head coil, then the bed is moved into the scanner. [Image: Janne Moren, https://goo.gl/MKb2jn, CC BY-NC-SA 2.0, goo.gl/Toc0ZF]

The conclusion that the mPFC is associated with the self comes from studies measuring fMRI while subjects think about themselves (e.g., saying whether traits are descriptive of themselves). Using this knowledge, other researchers have looked at whether the same brain area is active when people make inferences about others. Mitchell, Neil Macrae, and Banaji (2005) showed subjects pictures of strangers and had them judge either how pleased the person was to have his or her picture taken or how symmetrical the face appeared. Judging whether someone is pleased about being photographed requires making an inference about someone's internal feelings – we call this **mentalizing**. By contrast, facial symmetry judgments are based solely on physical appearances and do not involve mentalizing. A comparison of brain activity during the two types of judgments shows more activity in the mPFC when making the mental versus physical judgments, suggesting this brain area is involved when inferring the internal beliefs of others.

There are two other notable aspects of this study. First, mentalizing about others also increased activity in a variety of regions important for many aspects of social processing, including a region important in representing biological motion (**superior temporal sulcus** or STS), an area critical for emotional processing (amygdala), and a region also involved in thinking about the beliefs of others (**temporal parietal junction**, TPJ) (Gobbini & Haxby, 2007; Schultz, Imamizu, Kawato, & Frith, 2004) (See Figure **16.4.4**). This finding shows that a distributed and interacting set of brain areas is likely to be involved in social processing.

Second, activity in the most ventral part of the mPFC (the part closer to the belly rather than toward the top of the head), which has been most consistently associated with thinking about the self, was particularly active when subjects mentalized about people they rated as similar to themselves. Simulation is thought to be most likely for similar others (as in other people who one assumes to be similar to oneself), so this finding lends support to the conclusion that we use simulation to mentalize about others. After all, if you encounter someone who has the same musical taste as you, you will probably assume you have other things in common with him. By contrast, if you learn that someone loves music that you hate, you might expect him to differ from you in other ways (Srivastava, Guglielmo, & Beer, 2010). Using a simulation of our own feelings and thoughts will be most accurate if we have reason to think the person's internal experiences are like our own. Thus, we may be most likely to use simulation to make inferences about others if we think they are similar to us.

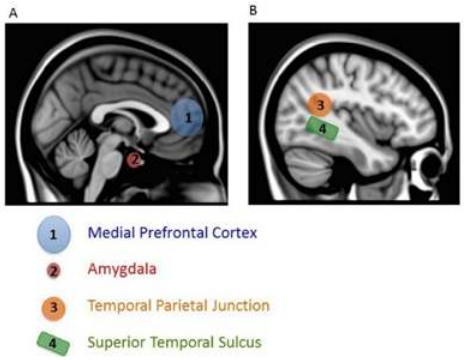


Figure **16.4.4**: The areas of the brain most commonly associated with processing of self and others. Panel A is a medial sagittal view, Panel B is a lateral sagittal view. The medial prefrontal cortex (mPFC) (Structure 1) is commonly activated when thinking about one's self and when thinking about similar others. Additionally, the amygdala (Structure 2) is important for learning about and detecting important things in our environments and plays an important role in fear learning and expression. The temporal parietal junction (TPJ) (Structure 3), located at the intersection of the parietal and temporal lobes, is activated when people think about the beliefs of others. Activity in the superior temporal sulcus (STS) (Structure 4) is commonly observed when people view biological motion. The common activation of this network of regions when people think about the feelings, thoughts, and intentions of others indicates that the processing of others involves a number of complex psychological processes.

This research is a good example of how social neuroscience is revealing the **functional neuroanatomy** of social behavior. That is, it tells us which brain areas are involved with social behavior. The mPFC (as well as other areas such as the STS, amygdala, and TPJ) is involved in making judgments about the self and others. This research also provides new information about how inferences are made about others. Whereas some have doubted the widespread use of simulation as a means for making inferences about others, the activation of the mPFC when mentalizing about others, and the sensitivity of this activation to similarity between self and other, provides evidence that simulation occurs.

16.4.5 What Is the Cost of Social Stress?

Stress is an unfortunately frequent experience for many of us. **Stress** – which can be broadly defined as a threat or challenge to our well-being – can result from everyday events like a course exam or more extreme events such as experiencing a natural disaster. When faced with a stressor, **sympathetic nervous system** activity increases in order to prepare our body to respond to the challenge. This produces what Selye (1950) called a **fight or flight response**. The release of **hormones**, which act as messengers from one part of an organism (e.g., a cell or gland) to another part of the organism, is part of the stress response. This is further discussed in the next section.

A small amount of stress can actually help us stay alert and active. In comparison, sustained stressors, or chronic stress, detrimentally affect our health and impair performance (Al'Absi, Hugdahl, & Lovallo, 2002; Black, 2002; Lazarus, 1974). This happens in part through the chronic secretion of stress-related hormones (e.g., Davidson, Pizzagalli, Nitschke, & Putnam, 2002; Dickerson, Gable, Irwin, Aziz, & Kemeny, 2009). In particular, stress activates the **hypothalamic-pituitary-adrenal (HPA) axis** to release **cortisol** (see Figure **16.4.5**). Chronic stress, by way of increases in cortisol, impairs attention, memory, and self-control (Arnsten, 2009). Cortisol levels can be measured non-invasively in bodily fluids, including blood and saliva. Researchers often collect a cortisol sample before and after a potentially stressful task. In one common collection method, subjects place polymer swabs under their tongue for 1 to 2 minutes to soak up saliva. The saliva samples are then stored and analyzed later to determine the level of cortisol present at each time point.

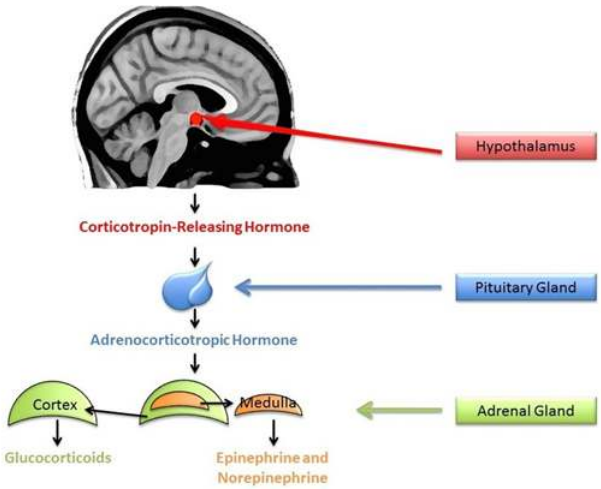


Figure **16.4.5**: The Hypothalamic-pituitary-adrenal (HPA) axis. Black arrows represent the stress response pathway starting in the brain at the hypothalamus (an area within the brain). Stress triggers neurons in the hypothalamus to release corticotrophin-releasing hormone (CRH). The CRH is transported to the pituitary gland, another area in the brain, that activates the secretion of andrenocorticotropic hormone (ACTH). In turn, ACTH stimulates the adrenal glands that sit on top of the kidneys. The adrenal glands are composed of the outer adrenal cortex and inner adrenal medulla. The adrenal cortex secretes glucorcorticoids (including cortisol) and the medulla secretes epinephrine and norepinephrine. Stress, both psychological and physical, activates the HPA axis and results in the systemic release of cortisol, epinephrine, and norepinephrine.

Whereas early stress researchers studied the effects of physical stressors like loud noises, social neuroscientists have been instrumental in studying how our interactions with other people can cause stress. This question has been addressed through **neuroendocrinology**, or the study of how the brain and hormones act in concert to coordinate the physiology of the body. One contribution of this work has been in understanding the conditions under which other people can cause stress. In one study, Dickerson, Mycek, and Zaldivar (2008) asked undergraduates to deliver a speech either alone or to two other people. When the students gave the speech in front of others, there was a marked increase in cortisol compared with when they were asked to give a speech alone. This suggests that like chronic physical stress, everyday social stressors, like having your performance judged by others, induces a stress response. Interestingly, simply giving a speech in the same room with someone who is doing something else did not induce a stress response. This suggests that the mere presence of others is not stressful, but rather it is the potential for them to judge us that induces stress.

Worrying about what other people think of us is not the only source of social stress in our lives. Other research has shown that interacting with people who belong to different social groups than us – what social psychologists call **outgroup** members – can increase physiological stress responses. For example, cardiovascular responses associated with stress like contractility of the heart ventricles and the amount of blood pumped by the heart (what is called cardiac output) are increased when interacting with outgroup as compared with **ingroup** members (i.e., people who belong to the same social group we do) (Mendes, Blascovich, Likel, & Hunter, 2002). This stress may derive from the expectation that interactions with dissimilar others will be uncomfortable (Stephan & Stephan, 1985) or concern about being judged as unfriendly and prejudiced if the interaction goes poorly (Plant & Devine, 2003).

The research just reviewed shows that events in our social lives can be stressful, but are social interactions always bad for us? No. In fact, while others can be the source of much stress, they are also a major buffer against stress. Research on **social support** shows that relying on a network of individuals in tough times gives us tools for dealing with stress and can ward off loneliness (Cacioppo & Patrick, 2008). For instance, people who report greater social support show a smaller increase in cortisol when performing a speech in front of two evaluators (Eisenberger, Taylor, Gable, Hilmert, & Lieberman, 2007).

What determines whether others will increase or decrease stress? What matters is the context of the social interaction. When it has potential to reflect badly on the self, social interaction can be stressful, but when it provides support and comfort, social interaction can protect us from the negative effects of stress. Using neuroendocrinology by measuring hormonal changes in the body has helped researchers better understand how social factors impact our body and ultimately our health.

16.4.6 Conclusions

Human beings are intensely social creatures – our lives are intertwined with other people and our health and well-being depend on others. Social neuroscience helps us to understand the critical function of how we make sense of and interact with other people. This module provides an introduction to what social neuroscience is and what we have already learned from it, but there is much still to understand. As we move forward, one exciting future direction will be to better understand how different parts of the brain and body interact to produce the numerous and complex patterns of social behavior that humans display. We hinted at some of this complexity when we reviewed research showing that while the mPFC is involved in mentalizing, other areas such as the STS, amygdala, and TPJ are as well. There are likely additional brain areas involved as well, interacting in ways we do not yet fully understand. These brain areas in turn control other aspects of the body to coordinate our responses during social interactions. Social neuroscience will continue to investigate these questions, revealing new information about how social processes occur, while also increasing our understanding of basic neural and physiological processes.

16.4.7 References

- Al'Absi, M., Hugdahl, K., & Lovallo, W. (2002). Adrenocortical stress responses and altered working memory performance. *Psychophysiology, 39*(1), 95–99.
- Arnsten, A. F. T. (2009). Stress signaling pathways that impair prefrontal cortex structure and function. *Nature Neuroscience Reviews, 10*(6), 410–422.
- Black, P. (2002). Stress and the inflammatory response: A review of neurogenic inflammation. *Brain, Behavior, & Immunity, 16*, 622–653.
- Cacioppo, J. T., & Berntson, G. G. (1992). Social psychological contributions to the decade of the brain: Doctrine of multilevel analysis. *American Psychologist, 47*, 1019–1028.
- Cacioppo, J. T., & Patrick, B. (2008). *Loneliness: Human nature and the need for social connection*. New York, NY: W. W. Norton & Company.
- Carruthers, P. and Smith, P. (1996). *Theories of Theories of Mind*. New York, NY: Cambridge University Press.
- Davidson, R. J., Pizzagalli, D., Nitschke, J. B., & Putnam, K. (2002). Depression: Perspectives from affective neuroscience. *Annual Review of Psychology, 53*, 545–574.
- Dickerson, S. S., Gable, S. L., Irwin, M. R., Aziz, N., & Kemeny, M. E. (2009). Social-evaluative threat and proinflammatory cytokine regulation an experimental laboratory investigation. *Psychological Science, 20*, 1237–1244.
- Dickerson, S. S., Mycek, P. J., & Zaldivar, F. (2008). Negative social evaluation, but not mere social presence, elicits cortisol responses to a laboratory stressor task. *Health Psychology, 27*(1), 116–121.
- Eisenberger, N. I., Taylor, S. E., Gable, S. L., Hilmert, C. J., & Lieberman, M. D. (2007). Neural pathways link social support to attenuated neuroendocrine stress responses. *Neuroimage, 35*(4), 1601–1612.
- Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2*, 493–501.
- Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia, 45*(1), 32–41.
- Goldman, A. I. (2005). Imitation, mind reading, and simulation. In S. Hurley & N. Chater (Eds.),*Perspectives on imitation: From neuroscience to social science (Vol. 2: Imitation, human development, and culture*, pp. 79–93). Cambridge, MA: MIT Press.
- Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology, 85*, 616–626.
- Ito, T.A., & Urland, G.R. (2005). The influence of processing objectives on the perception of faces: An ERP study of race and gender perception. *Cognitive, Affective, and Behavioral Neuroscience, 5*, 21–36.
- Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience, 14*, 785–794.
- Lazarus, R. S., (1974). Psychological stress and coping in adaptation and illness. *International Journal of Psychiatry in Medicine, 5*, 321–333.
- Mendes, W. B., Blascovich, J., Lickel, B., & Hunter, S. (2002). Challenge and threat during social interactions with White and Black men. *Personality and Social Psychology Bulletin, 28*, 939–952.

- Mitchell, J. P., Neil Macrae, C., & Banaji, M. R. (2005). Forming impressions of people versus inanimate objects: social-cognitive processing in the medial prefrontal cortex. *Neuroimage*, *26*(1), 251–257.
- Mouchetant-Rostaing, Y., & Giard, M. H. (2003). Electrophysiological correlates of age and gender perception on human faces. *Journal of Cognitive Neuroscience*, *15*, 900–910.
- Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron*, *48*, 175.
- Plant, E. A., & Devine, P. G. (2003). The antecedents and implications of interracial anxiety. *Personality and Social Psychology Bulletin, 29*, 790–801.
- Schultz, J., Imamizu, H., Kawato, M., & Frith, C. D. (2004). Activation of the human superior temporal gyrus during observation of goal attribution by intentional objects. *Journal of Cognitive Neuroscience*, *16*, 1695–1705.
- Selye, H. (1950). *The physiology and pathology of exposure to stress*. Montreal: Acta Inc.
- Srivastava, S., Guglielmo, S., & Beer, J. S. (2010). Perceiving others' personalities: Examining the dimensionality, assumed similarity to the self, and stability of perceiver effects. *Journal of Personality and Social Psychology*, *98*, 520.
- Stephan, W. G., & Stephan, C. W. (1985). Intergroup anxiety. *Journal of Social Issues*, *41*(3), 157–175.
- Whalen, P. J., & Phelps, E. A. (2009). *The human amygdala*. New York, NY: The Guilford Press.

16.4.8 Attributions

Social Neuroscience by Tiffany Ito & Jennifer Kubota, University of Colorado, Boulder, New York University, Noba Project, licensed CC BY-NC-SA 4.0

This page titled 16.4: Social Neuroscience is shared under a mixed license and was authored, remixed, and/or curated by ASCCC OERI & Bakhtawar Bhadha (ASCCC Open Educational Resources Initiative (OERI)) .

16.5: Stress and health

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Describe basic terminology used in the field of health psychology.
- Explain theoretical models of health, as well as the role of psychological stress in the development of disease.
- Describe psychological factors that contribute to resilience and improved health.
- Defend the relevance and importance of psychology to the field of medicine.

16.5.1 Overview

Our emotions, thoughts, and behaviors play an important role in our health. Not only do they influence our day-to-day health practices, but they can also influence how our body functions. Clearly, there is a connection between stress and health - both physical and mental. This section provides an overview of health psychology, which is a field devoted to understanding the connections between psychology and health. Discussed here are examples of topics a health psychologist might study, including stress, psychosocial factors related to health and disease, how to use psychology to improve health, and the role of psychology in medicine.

16.5.2 What Is Health Psychology?

Today, we face more **chronic disease** than ever before because we are living longer lives while also frequently behaving in unhealthy ways. One example of a chronic disease is coronary heart disease (CHD): It is the number one cause of death worldwide (World Health Organization, 2013). CHD develops slowly over time and typically appears midlife, but related heart problems can persist for years after the original diagnosis or cardiovascular event. In managing illnesses that persist over time (other examples might include cancer, diabetes, and long-term disability) many psychological factors will determine the progression of the ailment. For example, do patients seek help when appropriate? Do they follow doctor recommendations? Do they develop negative psychological symptoms due to lasting illness (e.g., depression)? Also important is that psychological factors can play a significant role in *who* develops these diseases, the prognosis, and the nature of the symptoms related to the illness. Health psychology is a relatively new, interdisciplinary field of study that focuses on these very issues, or more specifically, the role of psychology in maintaining health, as well as preventing and treating illness.



Figure **16.5.1**: Health psychologists are helping people to adapt behaviors to avoid disease, reduce stress, and improve overall health. [Image: Adelphi Lab Center, https://goo.gl/N9wXon, CC BY 2.0, goo.gl/BRvSA7]

Consideration of how psychological and social factors influence health is especially important today because many of the leading causes of illness in developed countries are often attributed to psychological and behavioral factors. In the case of CHD, discussed above, psychosocial factors, such as excessive stress, smoking, unhealthy eating habits, and some personality traits can also lead to increased risk of disease and worse health outcomes. That being said, many of these factors can be adjusted using psychological techniques. For example, clinical health psychologists can improve health practices like poor dietary choices and smoking, they can teach important stress reduction techniques, and they can help treat psychological disorders tied to poor health. Health psychology considers how the choices we make, the behaviors we engage in, and even the emotions that we feel, can play an important role in our overall health (Cohen & Herbert, 1996; Taylor, 2012).

Health psychology relies on the **Biopsychosocial Model of Health**. This model posits that biology, psychology, and social factors are just as important in the development of disease as biological causes (e.g., germs, viruses), which is consistent with the World Health Organization (1946) definition of **health**. This model replaces the older **Biomedical Model of Health**, which primarily considers the physical, or pathogenic, factors contributing to illness. Thanks to advances in medical technology, there is a growing understanding of the physiology underlying the **mind–body connection**, and in particular, the role that different feelings can have on our body's function (See Figure **16.5.1**) . Health psychology researchers working in the fields of **psychosomatic medicine** and **psychoneuroimmunology**, for example, are interested in understanding how psychological factors can "get under the skin" and influence our physiology in order to better understand how factors like stress can make us sick.

16.5.3 What Is stress?

It is clear that stress plays a major role in our mental and physical health, but what exactly is it? The term **stress** was originally derived from the field of mechanics where it is used to describe materials under pressure. The word was first used in a

psychological manner by researcher Hans Selye. He was examining the effect of an ovarian hormone that he thought caused sickness in a sample of rats. Surprisingly, he noticed that almost any injected hormone produced this same sickness. He smartly realized that it was not the hormone under investigation that was causing these problems, but instead, the unpleasant experience of being handled and injected by researchers that led to high physiological arousal and, eventually, to health problems like ulcers. Selye (1946) coined the term "stressor" to label a stimulus that had this effect on the body and developed a model of the stress response called the **General Adaptation Syndrome**. Since then, psychologists have studied stress in a myriad of ways, including stress as caused by negative events (e.g., natural disasters or major life changes like dropping out of school), as caused by chronically difficult situations (e.g., taking care of a loved one with Alzheimer's), as caused by short-term hassles, as a biological fight-or-flight response, and even as clinical illness like post-traumatic stress disorder (PTSD). It continues to be one of the most important and well-studied psychological correlates of illness, because excessive stress causes potentially damaging wear and tear on the body and can influence almost any imaginable disease process.

16.5.4 Stress and Health

You probably know exactly what it's like to feel stress, but what you may not know is that it can objectively influence your health. Answers to questions like, "How stressed do you feel?" or "How overwhelmed do you feel?" can predict your likelihood of developing both minor illnesses as well as serious problems like future heart attack (Cohen, Janicki-Deverts, & Miller, 2007). To understand how health psychologists study these types of associations, we will describe one famous example of a stress and health study. Imagine that you are a research subject for a moment. After you check into a hotel room as part of the study, the researchers ask you to report your general levels of stress. Not too surprising; however, what happens next is that you receive droplets of *cold virus* into your nose! The researchers intentionally try to make you sick by exposing you to an infectious illness. After they expose you to the virus, the researchers will then evaluate you for several days by asking you questions about your symptoms, monitoring how much mucus you are producing by weighing your used tissues, and taking body fluid samples—all to see if you are objectively ill with a cold. Now, the interesting thing is that not everyone who has drops of cold virus put in their nose develops the illness. (Clearly this study was done long before COVID-19 because I can't see people checking into a hotel, or being intentionally exposed to a virus - even the common cold one - after the pandemic!) Studies like this one find that people who are less stressed and those who are more positive at the beginning of the study are at a decreased risk of developing a cold (Cohen, Tyrrell, & Smith, 1991; Cohen, Alper, Doyle, Treanor, & Turner, 2006) (see Figure 16.5.2 for an example).

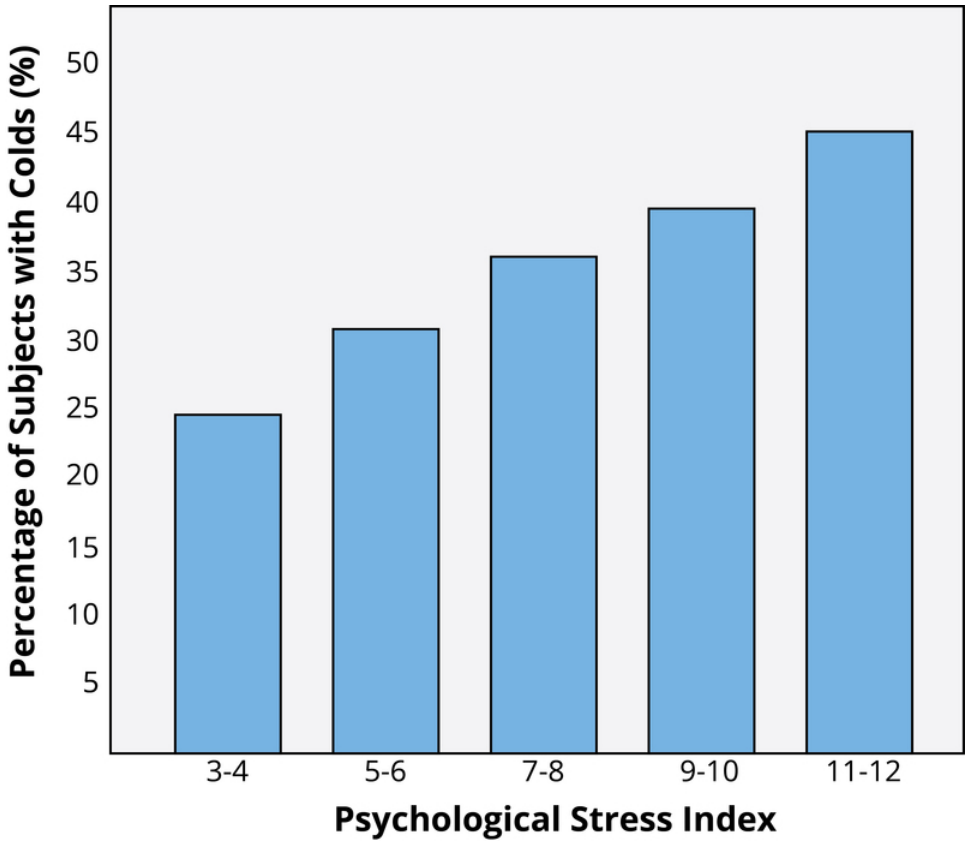


Figure **16.5.2**: Adapted from Cohen et al. 1991. This graph indicates that the greater the psychological stress index, the greater the percent of subjects with colds. In fact, there is a 25% increase in percentage of subjects with a cold from those who report a psychological stress index of 3-4 versus those reporting a psychological stress index of 11-12.

Importantly, it is not just major life **stressors** (e.g., a family death, a natural disaster) that increase the likelihood of getting sick. Even small **daily hassles** like getting stuck in traffic or fighting with your girlfriend/boyfriend or partner can raise your blood pressure, alter your stress hormones, and even suppress your immune system function (DeLongis, Folkman, & Lazarus, 1988; Twisk, Snel, Kemper, & van Machelen, 1999).

16.5.5 Immune System

The immune system is a host defense system. It comprises many biological structures —ranging from individual white blood cells to entire organs — as well as many complex biological processes. The function of the immune system is to protect the host from pathogens (something that causes a disease) and other causes of disease such as tumor cells. To function properly, the immune system must be able to detect a wide variety of pathogens. It also must be able to distinguish the cells of pathogens from the host's own cells and also to distinguish cancerous or damaged host cells from healthy cells. In humans and most other vertebrates, the immune system consists of layered defenses that have increased specificity for particular pathogens or tumor cells. The layered defenses of the human immune system are usually classified into two subsystems called the innate immune system and the adaptive immune system.

16.5.5.1 Innate Immune System

Any discussion of the innate immune response usually begins with the physical barriers that prevent pathogens (which are bacteria, viruses or other microorganisms that case diseases) from entering the body, destroy them after they enter, or flush them out before they can establish themselves in the hospitable environment of the body's soft tissues. Barrier defenses are part of the body's most basic defense mechanisms. The barrier defenses are not a response to infections, but they are continuously working to protect against a broad range of pathogens.

The phagocytes are the body's fast acting first line of immunological defense against organisms that have breached barrier defenses and have entered the vulnerable tissues of the body. For example, certain leukocytes (white blood cells) engulf and destroy

pathogens they encounter in the process called phagocytosis. The body's response against a pathogen's breach is also called inflammation. Inflammation is discussed in detail again a few paragraphs later.

16.5.5.2 Adaptive Immune System

The adaptive immune system is activated if pathogens successfully enter the body and manage to evade the general defenses of the innate immune system. An adaptive response is specific to the particular type of pathogen that has invaded the body or to cancerous cells. It takes longer to launch a specific attack, but once it is underway, its specificity makes it very effective. An adaptive immune response is set in motion by antigens that the immune system recognizes as foreign. Unlike an innate immune response, an adaptive immune response is highly specific to a particular pathogen (or its antigen).

An important function of the adaptive immune system that is not shared by the innate immune system is the creation of immunological memory or immunity. This occurs after the initial response to a specific pathogen. It allows a faster, stronger response on subsequent encounters with the same pathogen, usually before the pathogen can cause symptoms of illness. An adaptive response also usually leads to immunity. This is a state of resistance to a specific pathogen due to the ability of the adaptive immune system to “remember” the pathogen and immediately mount a strong attack tailored to that particular pathogen if it invades again in the future.

16.5.5.2.1 Lymphatic System

The lymphatic system is a human organ system that is a vital part of the adaptive immune system. It is also part of the cardiovascular system and plays a major role in the digestive system

The primary function of the lymphatic system is to host defense as part of the immune system. This function of the lymphatic system is centered on the production, maturation, and circulation of lymphocytes. Lymphocytes are leukocytes that are involved in the adaptive immune system. They are responsible for the recognition of, and tailored defense against, specific pathogens or tumor cells. Lymphocytes may also create a lasting memory of pathogens so they can be attacked quickly and strongly if they ever invade the body again. In this way, lymphocytes bring about long-lasting immunity to specific pathogens.

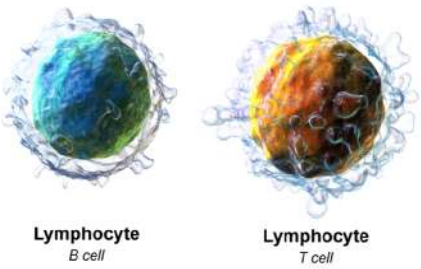


Figure **16.5.3**: These are Lymphocytes - B nd T cells - Notice the large nucleus in the center of these relatively round cells.

.

Lymphocytes are leukocytes that arise and mature in organs of the lymphatic system, including the bone marrow and thymus. There are two main types of lymphocytes involved in adaptive immune responses, called T cells and B cells which are illustrated in Figure **16.5.3**. Both B cells and T cells are involved in the adaptive immune response, but they play different roles. T cells destroy infected cells or release chemicals that regulate immune responses. B cells secrete antibodies that bind with antigens of pathogens so they can be removed by other immune cells or processes.

T Cells

There are multiple types of T cells or T lymphocytes. Major types are killer (or cytotoxic) T cells and helper T cells. Both types develop from immature T cells that become activated by exposure to an antigen. T cells must be activated. After the pathogen is phagocytized and digested by macrophages, a part of the pathogen is displayed on the surface of the macrophage. Helper T cells are more easily activated than killer T cells. Activation of killer T cells is strongly regulated and may require additional stimulation from helper T cells.

B Cells and B Cell Activation

B cells, or B lymphocytes, are the major cells involved in the creation of antibodies that circulate in blood plasma and lymph. Antibodies are large, Y-shaped proteins used by the immune system to identify and neutralize foreign invaders. Besides producing antibodies, B cells may also function as antigen-presenting cells or secrete cytokines that help control other immune cells and responses.

16.5.5.3 Self vs. Non-Self

Both innate and adaptive immune responses depend on the ability of the immune system to distinguish between self and non-self molecules. Self molecules are those components of an organism's body that can be distinguished from foreign substances by the immune system.

16.5.5.3.1 Antigens and Antibodies

Many non-self molecules comprise a class of compounds called antigens. Antigens, which are usually proteins, bind to specific receptors on immune system cells and elicit an adaptive immune response. Some adaptive immune system cells (B cells) respond to foreign antigens by producing antibodies. An antibody is a molecule that precisely matches and binds to a specific antigen. This may target the antigen (and the pathogen displaying it) for destruction by other immune cells.

Antigens on the surface of pathogens are how the adaptive immune system recognizes specific pathogens. Antigen specificity allows for the generation of responses tailored to the specific pathogen. It is also how the adaptive immune system "remembers" the same pathogen in the future.

16.5.5.4 Immune Surveillance

Another important role of the immune system is to identify and eliminate tumor cells. This is called immune surveillance. The transformed cells of tumors express antigens that are not found on normal body cells. The main response of the immune system to tumor cells is to destroy them. This is carried out primarily by aptly named killer T cells of the adaptive immune system.

16.5.5.5 Inflammation and Fever

The inflammatory response, or inflammation, is triggered by a cascade of chemical mediators and cellular responses that may occur when cells are damaged and stressed or when pathogens successfully breach the physical barriers of the innate immune system. Although inflammation is typically associated with negative consequences of injury or disease, it is a necessary process insofar as it allows for recruitment of the cellular defenses needed to eliminate pathogens, remove damaged and dead cells, and initiate repair mechanisms. Excessive inflammation, however, can result in local tissue damage and, in severe cases, may even become deadly.

A fever is an inflammatory response that extends beyond the site of infection and affects the entire body, resulting in an overall increase in body temperature. Body temperature is normally regulated and maintained by the hypothalamus, an anatomical section of the brain that functions to maintain homeostasis in the body. However, certain bacterial or viral infections can result in the production of pyrogens, chemicals that effectively alter the "thermostat setting" of the hypothalamus to elevate body temperature and cause fever. Pyrogens may be exogenous or endogenous. For example, the endotoxin lipopolysaccharide (LPS), produced by gram-negative bacteria, is an exogenous pyrogen that may induce the leukocytes to release endogenous pyrogens such as interleukin-1 (IL-1), IL-6, interferon-γ (IFN-γ), and tumor necrosis factor (TNF). In a cascading effect, these molecules can then lead to the release of prostaglandin E2 (PGE_2) from other cells, resetting the hypothalamus to initiate fever (Figure 16.5.4:).

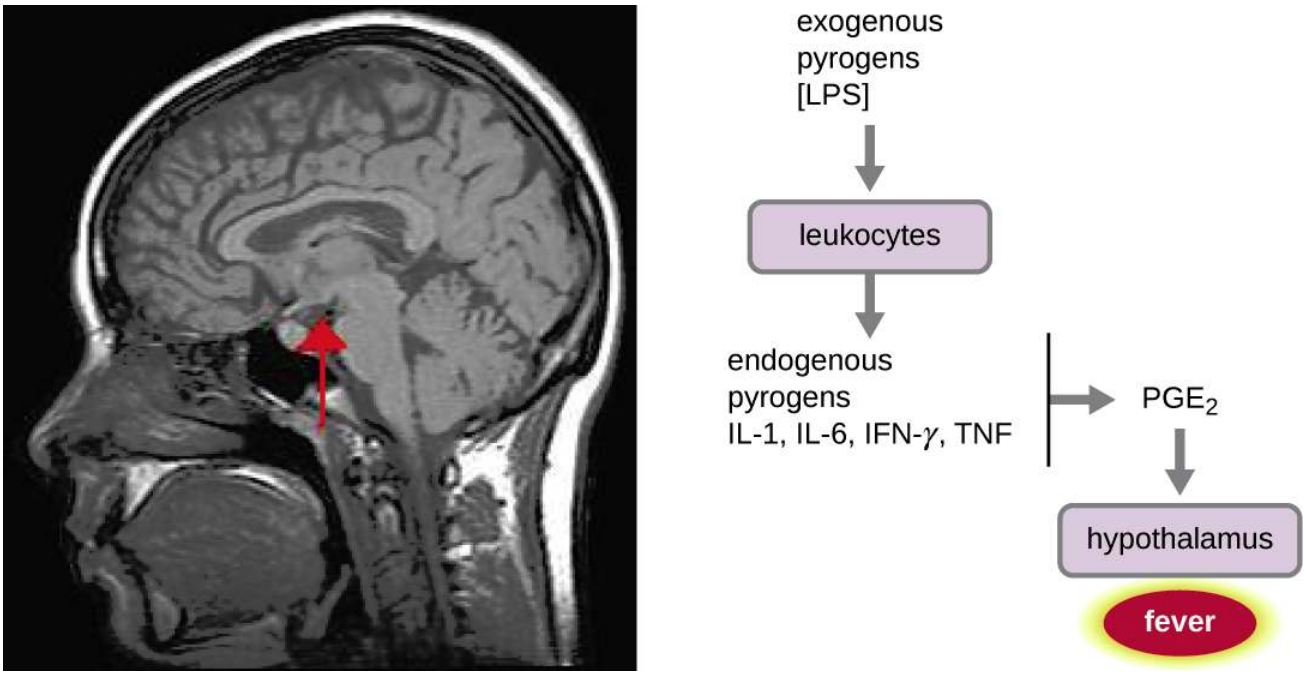


Figure 16.5.4:Macrophages recognize pathogens in an area and release cytokines that trigger inflammation. The cytokines also send a signal up the vagus nerve to the hypothalamus.

Inflammation and Fever by Openstax, licensed CC BY 4.0

Like other forms of inflammation, a fever enhances the innate immune defenses by stimulating leukocytes to kill pathogens. The rise in body temperature also may inhibit the growth of many pathogens since human pathogens are mesophiles with optimum growth occurring around 35 °C (95 °F). In addition, some studies suggest that fever may also stimulate release of iron-sequestering compounds from the liver, thereby starving out microbes that rely on iron for growth.

During fever, the skin may appear pale due to vasoconstriction of the blood vessels in the skin, which is mediated by the hypothalamus to divert blood flow away from extremities, minimizing the loss of heat and raising the core temperature. The hypothalamus will also stimulate shivering of muscles, another effective mechanism of generating heat and raising the core temperature.

The crisis phase occurs when the fever breaks. The hypothalamus stimulates vasodilation, resulting in a return of blood flow to the skin and a subsequent release of heat from the body. The hypothalamus also stimulates sweating, which cools the skin as the sweat evaporates.

Although a low-level fever may help an individual overcome an illness, in some instances, this immune response can be too strong, causing tissue and organ damage and, in severe cases, even death. The inflammatory response to bacterial superantigens is one scenario in which a life-threatening fever may develop. Superantigens are bacterial or viral proteins that can cause an excessive activation of T cells from the specific adaptive immune defense, as well as an excessive release of cytokines that overstimulates the inflammatory response. For example, *Staphylococcus aureus* and *Streptococcus pyogenes* are capable of producing superantigens that cause toxic shock syndrome and scarlet fever, respectively. Both of these conditions can be associated with very high, life-threatening fevers in excess of 42 °C (108 °F).

16.5.6 Stress and Immune Function

It is widely believed that stress suppresses immune function and increases susceptibility to infections and cancer. Paradoxically, stress is also known to exacerbate allergic, autoimmune, and inflammatory diseases. These observations suggest that stress may have bidirectional effects on immune function, being immunosuppressive (deterring the immune system) in some instances and immunoenhancing (enhancing the immune system) in others. It has recently been shown that in contrast to chronic stress that suppresses or dysregulates immune function, acute stress can be immunoenhancing. Acute stress enhances dendritic cell, neutrophil, macrophage, and lymphocyte trafficking, maturation, and function and has been shown to augment innate and adaptive

immune responses. Acute stress experienced prior to novel antigen exposure enhances innate immunity and memory T-cell formation and results in a significant and long-lasting immunoenhancement. Acute stress experienced during antigen reexposure enhances secondary/adaptive immune responses. Therefore, depending on the conditions of immune activation and the immunizing antigen, acute stress may enhance the acquisition and expression of immunoprotection or immunopathology. In contrast, chronic stress dysregulates innate and adaptive immune responses by changing the type 1-type 2 cytokine balance and suppresses immunity by decreasing leukocyte numbers, trafficking, and function. Chronic stress also increases susceptibility to skin cancer by suppressing type 1 cytokines and protective T cells while increasing suppressor T-cell function. The adaptive purpose of a physiological stress response may be to promote survival, with stress hormones and neurotransmitters serving as beacons that prepare the immune system for potential challenges (eg, wounding or infection) perceived by the brain (eg, detection of an attacker). However, this system may exacerbate immunopathology (diseases/malfunction of the immune system) if the enhanced immune response is directed against innocuous or self-antigens or dysregulated following prolonged activation, as seen during chronic stress (see Figure **16.5.5**). In view of the ubiquitous nature of stress and its significant effects on immunoprotection and immunopathology, it is important to further elucidate the mechanisms mediating stress-immune interactions and to meaningfully translate findings from laboratory bench to hospital bedside.

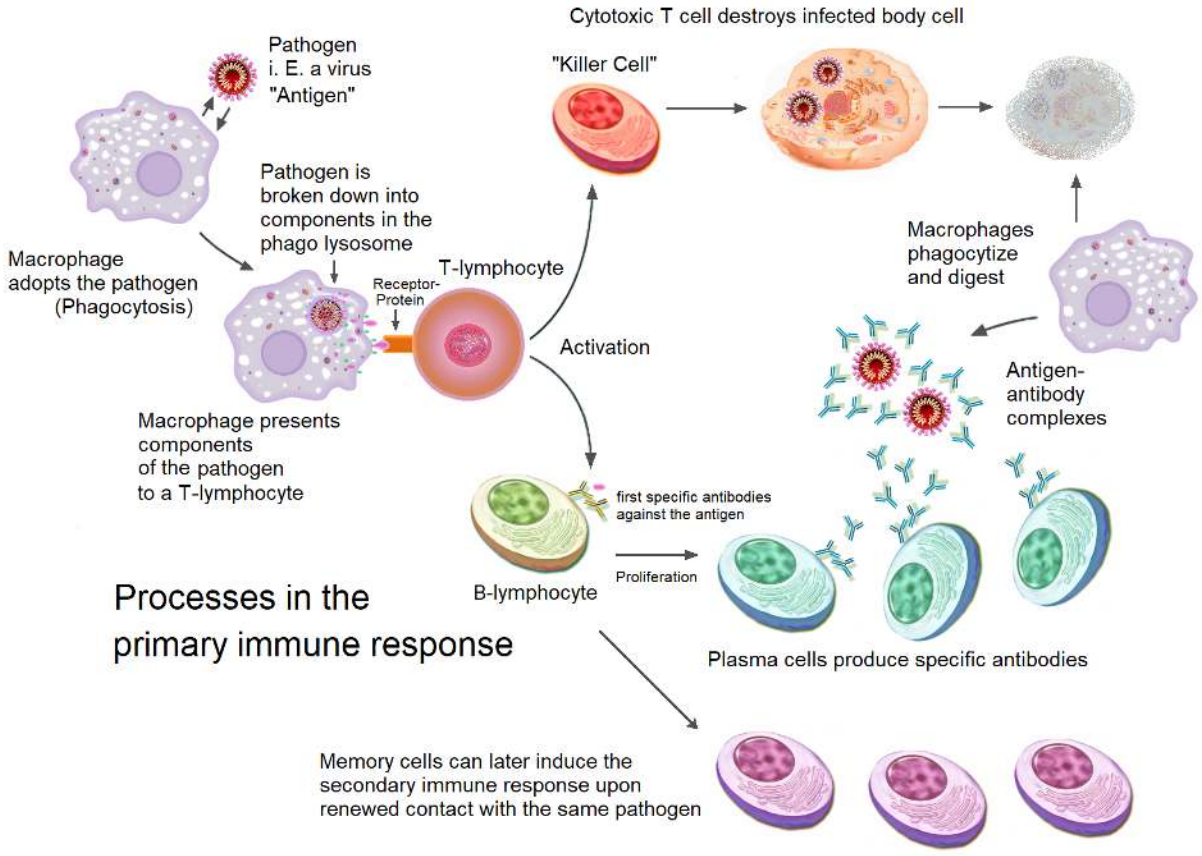


Figure **16.5.5**: Macrophages break down a pathogen (virus) and present the components to a T-cell which activates killer cells that destroy the infected body cell, and flag the antigens, and also trigger B cells which create antibodies for those antigens as well as memory cells for future reference. (Copyright;Processes in the primary immune response by: Sciencia58 an the makers of the single images Domdomegg, [1], Fæ, Petr94, Manu5, CC BY-SA 4.0, via Wikimedia Commons)

An important function of physiological mediators released under conditions of acute psychological stress may be to ensure that appropriate leukocytes are present in the right place and at the right time to respond to an immune challenge that might be initiated by the stress-inducing agent (eg, attack by a predator, invasion by a pathogen, etc.). The modulation of immune cell distribution by acute stress may be an adaptive response designed to enhance surveillance and increase the capacity of the immune system to

respond to challenge in compartments (such as the skin, lung, gastrointestinal and urinary-genital tracts, mucosal surfaces, and lymph nodes), which serve as defense barriers for the body. Thus, neurotransmitters and hormones released during stress may increase immunosurveillance and help enhance immune preparedness for potential (or ongoing) immune challenge. Stress-induced immunoenhancement may increase immunoprotection during surgery, vaccination, or infection but may also exacerbate immunopathology during inflammatory (asthma, allergy, dermatitis, cardiovascular disease, gingivitis) or autoimmune (psoriasis, arthritis, multiple sclerosis) diseases that are known to be exacerbated by stress (Amkraut et al, 1971; Ackerman, 2002; Al'Abadie et al, 1994; Garg, 2001; Wright et al, 1998; Wright, 2001).

16.5.7 Protecting Our Health

An important question that health psychologists ask is: What keeps us protected from disease and alive longer? When considering this issue of **resilience** (Rutter, 1985), five factors are often studied in terms of their ability to protect (or sometimes harm) health. They are:

1. Coping
2. Control and Self-Efficacy
3. Social Relationships
4. Dispositions and Emotions
5. Stress Management

16.5.7.1 Coping Strategies

How individuals cope with the stressors they face can have a significant impact on health. Coping is often classified into two categories: problem-focused coping or emotion-focused coping (Carver, Scheier, & Weintraub, 1989). **Problem-focused coping** is thought of as actively addressing the event that is causing stress in an effort to solve the issue at hand. For example, say you have an important exam coming up next week. A problem-focused strategy might be to spend additional time over the weekend studying to make sure you understand all of the material. **Emotion-focused coping**, on the other hand, regulates the emotions that come with stress. In the above examination example, this might mean watching a funny movie to take your mind off the anxiety you are feeling. In the short term, emotion-focused coping might reduce feelings of stress, but problem-focused coping seems to have the greatest impact on mental wellness (Billings & Moos, 1981; Herman-Stabl, Stemmler, & Petersen, 1995). That being said, when events are uncontrollable (e.g., the death of a loved one), emotion-focused coping directed at managing your feelings, at first, might be the better strategy. Therefore, it is always important to consider the match of the stressor to the coping strategy when evaluating its plausible benefits.

16.5.7.2 Control and Self-Efficacy



Figure **16.5.6**: Feeling a sense of control in one's life is important. Something as simple as having control over the care of a houseplant has been shown to improve health and longevity. [Image: JJ Harrison, https://goo.gl/82FsdV, CC BY-SA 2.5, goo.gl/SRAIwa]

Another factor tied to better health outcomes and an improved ability to cope with stress is having the belief that you have **control** over a situation. For example, in one study where participants were forced to listen to unpleasant (stressful) noise, those who were led to believe that they had control over the noise performed much better on proofreading tasks afterwards (Glass & Singer, 1972). In other words, even though participants *did not* have actual control over the noise, the control *belief* aided them in completing the task. In similar studies, perceived control benefited immune system functioning (Sieber et al., 1992) (See Figure 16.5.6)Outside of the laboratory, studies have shown that older residents in assisted living facilities, which are notorious for low control, lived *longer* and showed *better* health outcomes when given control over something as simple as watering a plant or choosing when student volunteers came to visit (Rodin & Langer, 1977; Schulz & Hanusa, 1978). In addition, feeling in control of a threatening situation can actually change stress hormone levels (Dickerson & Kemeny, 2004). Believing that you have control over your own behaviors can also have a positive influence on important outcomes like smoking cessation, contraception use, and weight management (Wallston & Wallston, 1978). When individuals do not believe they have control, they do not try to change. **Self-efficacy** is closely related to control, in that people with high levels of this trait believe they can complete tasks and reach their goals. Just as feeling in control can reduce stress and improve health, higher self-efficacy can reduce stress and negative **health behaviors**, and is associated with better health (O'Leary, 1985).

16.5.7.3 Social Relationships

Research has shown that the impact of social isolation on our risk for disease and death is similar in magnitude to the risk associated with smoking regularly (Holt-Lunstad, Smith, & Layton, 2010; House, Landis, & Umberson, 1988). In fact, the importance of social relationships for our health is so significant that some scientists believe our body has developed a physiological system that encourages us to seek out our relationships, especially in times of stress (Taylor et al., 2000). **Social integration** is the concept used to describe the number of social roles that you have (Cohen & Wills, 1985), as well as the lack of isolation. For example, you might be a daughter, a basketball team member, a Humane Society volunteer, a coworker, and a student. Maintaining these different roles can improve your health via encouragement from those around you to maintain a healthy

lifestyle. Those in your social network might also provide you with **social support** (e.g., when you are under stress). This support might include emotional help (e.g., a hug when you need it), tangible help (e.g., lending you money), or advice. By helping to improve health behaviors and reduce stress, social relationships can have a powerful, protective impact on health, and in some cases, might even help people with serious illnesses stay alive longer (Spiegel, Kraemer, Bloom, & Gottheil, 1989).

16.5.7.4 Dispositions and Emotions: What's Risky and What's Protective?

Negative dispositions and personality traits have been strongly tied to an array of health risks. One of the earliest negative trait-to-health connections was discovered in the 1950s by two cardiologists. They made the interesting discovery that there were common behavioral and psychological patterns among their heart patients that were not present in other patient samples. This pattern included being competitive, impatient, hostile, and time urgent. They labeled it **Type A Behavior**. Importantly, it was found to be associated with *double* the risk of heart disease as compared with **Type B Behavior** (Friedman & Rosenman, 1959). Since the 1950s, researchers have discovered that it is the **hostility** and competitiveness components of Type A that are especially harmful to heart health (Iribarren et al., 2000; Matthews, Glass, Rosenman, & Bortner, 1977; Miller, Smith, Turner, Guijarro, & Hallet, 1996). Hostile individuals are quick to get upset, and this angry arousal can damage the arteries of the heart. In addition, given their negative personality style, hostile people often lack a heath-protective supportive social network.

Positive traits and states, on the other hand, are often health protective. For example, characteristics like positive emotions (e.g., feeling happy or excited) have been tied to a wide range of benefits such as increased longevity, a reduced likelihood of developing some illnesses, and better outcomes once you are diagnosed with certain diseases (e.g., heart disease, HIV) (Pressman & Cohen, 2005). Across the world, even in the most poor and underdeveloped nations, positive emotions are consistently tied to better health (Pressman, Gallagher, & Lopez, 2013). Positive emotions can also serve as the "antidote" to stress, protecting us against some of its damaging effects (Fredrickson, 2001; Pressman & Cohen, 2005; see Figure 16.5.7:). Similarly, looking on the bright side can also improve health. Optimism has been shown to improve coping, reduce stress, and predict better disease outcomes like recovering from a heart attack more rapidly (Kubzansky, Sparrow, Vokonas, & Kawachi, 2001; Nes & Segerstrom, 2006; Scheier & Carver, 1985; Segerstrom, Taylor, Kemeny, & Fahey, 1998).

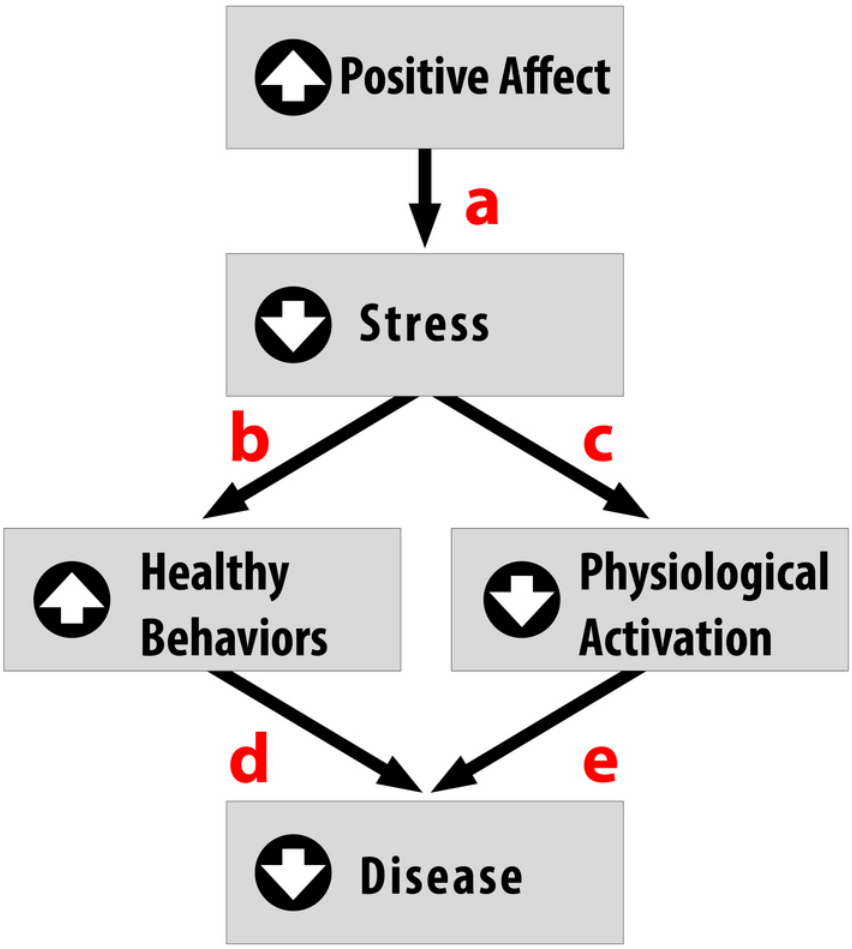


Figure **16.5.7**:This figure illustrates one possible way that positive affect protects individuals against disease. Positive affect can reduce stress perceptions (a), thereby improving health behaviors (b) and lowering physiological stress responses (c) (e.g., decreased cardiovascular reactivity, lower stress hormones, non-suppressed immune activity). As a result, there is likely to be less incidence of disease (d, e). (Adapted from Pressman & Cohen, 2005)

16.5.7.5 Stress Management

About 20 percent of Americans report having stress, with 18–33 year-olds reporting the highest levels (American Psychological Association, 2012). Given that the sources of our stress are often difficult to change (e.g., personal finances, current job), a number of interventions have been designed to help reduce the aversive responses to duress. For example, relaxation activities and forms of meditation are techniques that allow individuals to reduce their stress via breathing exercises, muscle relaxation, and mental imagery. Physiological arousal from stress can also be reduced via **biofeedback**, a technique where the individual is shown bodily information that is not normally available to them (e.g., heart rate), and then taught strategies to alter this signal. This type of intervention has even shown promise in reducing heart and hypertension risk, as well as other serious conditions (e.g., Moravec, 2008; Patel, Marmot, & Terry, 1981). But reducing stress does not have to be complicated! For example, exercise is a great stress reduction activity (Salmon, 2001) that has a myriad of health benefits.

16.5.8 The Importance Of Good Health Practices

As a student, you probably strive to maintain good grades, to have an active social life, and to stay healthy (e.g., by getting enough sleep), but there is a popular joke about what it's like to be in college: you can only pick two of these things (see Figure **16.5.8**: for an example). The busy life of a college student doesn't always allow you to maintain all three areas of your life, especially during test-taking periods. In one study, researchers found that students taking exams were more stressed and, thus, smoked more, drank more caffeine, had less physical activity, and had worse sleep habits (Oaten & Chang, 2005), all of which could have detrimental effects on their health. Positive health practices are *especially* important in times of stress when your immune system is

compromised due to high stress and the elevated frequency of exposure to the illnesses of your fellow students in lecture halls, cafeterias, and dorms.

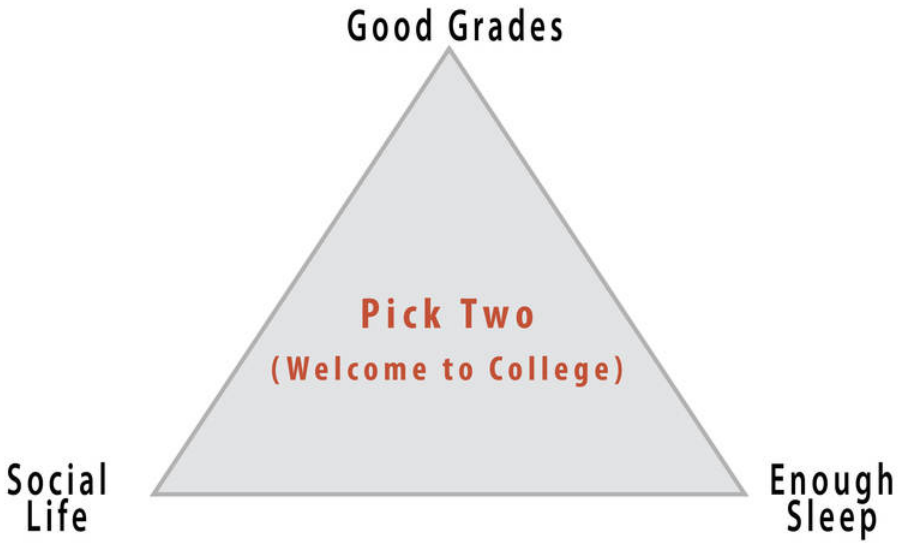


Figure **16.5.8**: A popular joke about how difficult it is to stay balanced and healthy during college.

Psychologists study both **health behaviors** and health habits. The former are behaviors that can improve or harm your health. Some examples include regular exercise, flossing, and wearing sunscreen, versus negative behaviors like drunk driving, pulling all-nighters, or smoking. These behaviors become *habits* when they are firmly established and performed automatically. For example, do you have to think about putting your seatbelt on or do you do it automatically? Habits are often developed early in life thanks to parental encouragement or the influence of our peer group.

While these behaviors sound minor, studies have shown that those who engaged in more of these protective habits (e.g., getting 7–8 hours of sleep regularly, not smoking or drinking excessively, exercising) had fewer illnesses, felt better, and were less likely to die over a 9–12-year follow-up period (Belloc & Breslow 1972; Breslow & Enstrom 1980). For college students, health behaviors can even influence academic performance. For example, poor sleep quality and quantity are related to weaker learning capacity and academic performance (Curcio, Ferrara, & De Gennaro, 2006). Due to the effects that health behaviors can have, much effort is put forward by psychologists to understand *how* to change unhealthy behaviors, and to understand *why* individuals fail to act in healthy ways. Health promotion involves enabling individuals to improve health by focusing on behaviors that pose a risk for future illness, as well as spreading knowledge on existing risk factors. These might be genetic risks you are born with, or something you developed over time like obesity, which puts you at risk for Type 2 diabetes and heart disease, among other illnesses.

16.5.9 Psychology And Medicine

There are many psychological factors that influence medical treatment outcomes. For example, older individuals, (Meara, White, & Cutler, 2004), women (Briscoe, 1987), and those from higher socioeconomic backgrounds (Adamson, Ben-Shlomo, Chaturvedi, & Donovan, 2008) are all *more* likely to seek medical care. On the other hand, some individuals who need care might avoid it due to financial obstacles or preconceived notions about medical practitioners or the illness. Thanks to the growing amount of medical information online, many people now use the Internet for health information and 38% percent report that this influences their decision to see a doctor (Fox & Jones, 2009) (see Figure **16.5.9**). Unfortunately, this is not always a good thing because individuals tend to do a poor job assessing the credibility of health information. For example, college-student participants reading online articles about HIV and syphilis rated a physician's article and a college student's article as *equally* credible if the participants said they were familiar with the health topic (Eastin, 2001). Credibility of health information often means how accurate or trustworthy the information is, and it can be influenced by irrelevant factors, such as the website's design, logos, or the organization's contact information (Freeman & Spyridakis, 2004). Similarly, many people post health questions on online, unmoderated forums where *anyone* can respond, which allows for the possibility of inaccurate information being provided for serious medical conditions by unqualified individuals.



Figure 16.5.9: While the Internet has increased the amount of medical information available to the public and created greater access, there are real concerns about how people are making decisions about their health based on that information. [Image: Mapbox, https://goo.gl/UNhmx5, CC BY 2.0, goo.gl/BRvSA7]

After individuals decide to seek care, there is also variability in the information they give their medical provider. Poor communication (e.g., due to embarrassment or feeling rushed) can influence the accuracy of the diagnosis and the effectiveness of the prescribed treatment. Similarly, there is variation following a visit to the doctor. While most individuals are tasked with a health recommendation (e.g., buying and using a medication appropriately, losing weight, going to another expert), not everyone *adheres* to medical recommendations (Dunbar-Jacob & Mortimer-Stephens, 2010). For example, many individuals take medications inappropriately (e.g., stopping early, not filling prescriptions) or fail to change their behaviors (e.g., quitting smoking). Unfortunately, getting patients to follow medical orders is not as easy as one would think. For example, in one study, over one third of diabetic patients failed to get proper medical care that would prevent or slow down diabetes-related blindness (Schoenfeld, Greene, Wu, & Leske, 2001)! Fortunately, as mobile technology improves, physicians now have the ability to monitor **adherence** and work to improve it (e.g., with pill bottles that monitor if they are opened at the right time). Even text messages are useful for improving treatment adherence and outcomes in depression, smoking cessation, and weight loss (Cole-Lewis, & Kershaw, 2010).

16.5.10 Being A Health Psychologist

Training as a clinical health psychologist provides a variety of possible career options. Clinical health psychologists often work on teams of physicians, social workers, allied health professionals, and religious leaders. These teams may be formed in locations like rehabilitation centers, hospitals, primary care offices, emergency care centers, or in chronic illness clinics. Work in each of these settings will pose unique challenges in patient care, but the primary responsibility will be the same. Clinical health psychologists will evaluate physical, personal, and environmental factors contributing to illness and preventing improved health. In doing so, they will then help create a treatment strategy that takes into account all dimensions of a person's life and health, which maximizes its potential for success. Those who specialize in health psychology can also conduct research to discover new health predictors and risk factors, or develop interventions to prevent and treat illness. Researchers studying health psychology work in numerous locations, such as universities, public health departments, hospitals, and private organizations. In the related field of **behavioral medicine**, careers focus on the application of this type of research. Occupations in this area might include jobs in occupational therapy, rehabilitation, or preventative medicine. Training as a health psychologist provides a wide skill set applicable in a number of different professional settings and career paths.

16.5.11 The Future of Health Psychology

Much of the past medical research literature provides an incomplete picture of human health. “Health care” is often “illness care.” That is, it focuses on the management of symptoms and illnesses as they arise. As a result, in many developed countries, we are faced with several health epidemics that are difficult and costly to treat. These include obesity, diabetes, and cardiovascular disease, to name a few. The National Institutes of Health have called for researchers to use the knowledge we have about risk factors to design effective interventions to reduce the prevalence of *preventable* illness. Additionally, there are a growing number of individuals across developed countries with *multiple* chronic illnesses and/or lasting disabilities, especially with older age. Addressing their needs and maintaining their quality of life will require skilled individuals who understand how to properly treat these populations. Health psychologists will be on the forefront of work in these areas.

With this focus on prevention, it is important that health psychologists move beyond studying risk (e.g., depression, stress, hostility, low socioeconomic status) in isolation, and move toward studying factors that confer resilience and protection from disease. There is, fortunately, a growing interest in studying the positive factors that protect our health (e.g., Diener & Chan, 2011; Pressman & Cohen, 2005; Richman, Kubzansky, Maselko, Kawachi, Choo, & Bauer, 2005) with evidence strongly indicating that people with higher positivity live longer, suffer fewer illnesses, and generally feel better. Seligman (2008) has even proposed a field of “Positive Health” to specifically study those who exhibit “above average” health—something we do not think about enough. By shifting some of the research focus to identifying and understanding these health-promoting factors, we may capitalize on this information to improve public health.

Innovative interventions to improve health are already in use and continue to be studied. With recent advances in technology, we are starting to see great strides made to improve health with the aid of computational tools. For example, there are hundreds of simple applications (apps) that use email and text messages to send reminders to take medication, as well as mobile apps that allow us to monitor our exercise levels and food intake (in the growing mobile-health, or m-health, field). These m-health applications can be used to raise health awareness, support treatment and compliance, and remotely collect data on a variety of outcomes. Also exciting are devices that allow us to monitor physiology in real time; for example, to better understand the stressful situations that raise blood pressure or heart rate. With advances like these, health psychologists will be able to serve the population better, learn more about health and health behavior, and develop excellent health-improving strategies that could be specifically targeted to certain populations or individuals. These leaps in equipment development, partnered with growing health psychology knowledge and exciting advances in neuroscience and genetic research, will lead health researchers and practitioners into an exciting new time where, hopefully, we will understand more and more about how to keep people healthy.

16.5.12 Outside Resources

App: 30 iPhone apps to monitor your health

http://www.hongkiat.com/blog/iphone-health-app/

Quiz: Hostility

http://www.mhhe.com/socscience/hhp/f...sheet_090.html

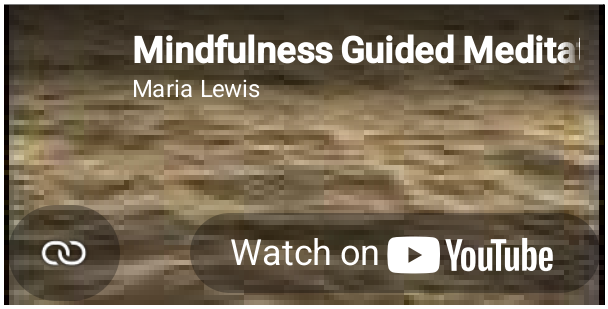
Self-assessment: Perceived Stress Scale

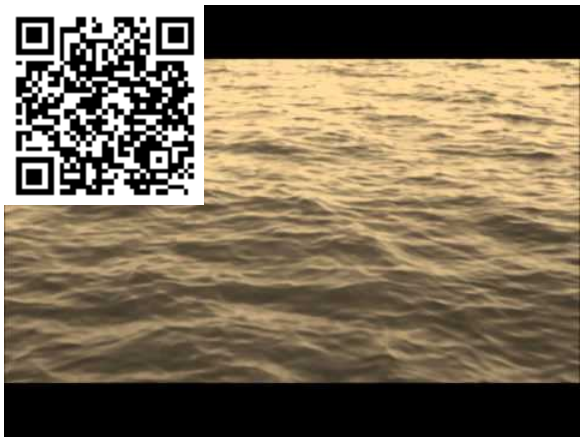
www.ncsu.edu/assessment/resou...ress_scale.pdf

Self-assessment: What’s your real age (based on your health practices and risk factors)?

http://www.realage.com

Video: Try out a guided meditation exercise to reduce your stress





Web: American Psychosomatic Society

http://www.psychosomatic.org/home/index.cfm

Web: APA Division 38, Health Psychology

http://www.health-psych.org

Web: Society of Behavioral Medicine

http://www.sbm.org

16.5.13 References

- Adamson, J., Ben-Shlomo, Y., Chaturvedi, N., & Donovan, J. (2008). Ethnicity, socio-economic position and gender—do they affect reported health—care seeking behaviour? *Social Science & Medicine*, *57*, 895–904.
- American Psychological Association (2012). Stress in American 2012 [Press release]. Retrieved from www.apa.org/news/press/releas...nerations.aspx
- Belloc, N. B., & Breslow, L. (1972). Relationship of physical health status and health practices. *Preventive Medicine*, *1*, 409–421.
- Billings, A. G., & Moos, R. H. (1981). The role of coping responses and social resources in attenuating the stress of life events. *Journal of Behavioral Medicine*, *4*, 139–157.
- Breslow, L., & Enstrom, J. E. (1980). Persistence of health habits and their relationship to mortality. *Preventive Medicine*, *9*, 469–483.
- Briscoe, M. E. (1987). Why do people go to the doctor? Sex differences in the correlates of GP consultation. *Social Science & Medicine*, *25*, 507–513.
- Carver, C. S., Scheier, M. F., & Weintraub, J. K. (1989). Assessing coping strategies: A theoretically based approach. *Journal of Personality and Social Psychology*, *56*, 267–283.
- Cohen, S., & Herbert, T. B. (1996). Health psychology: Psychological factors and physical disease from the perspective of human psychoneuroimmunology. *Annual Review of Psychology*, *47*, 113–142.
- Cohen, S., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological Bulletin*, *98*, 310–357.
- Cohen, S., Alper, C. M., Doyle, W. J., Treanor, J. J., & Turner, R. B. (2006). Positive emotional style predicts resistance to illness after experimental exposure to rhinovirus or influenza A virus. *Psychosomatic Medicine*, *68*, 809–815.

- Cohen, S., Janicki-Deverts, D., & Miller, G. E. (2007). Psychological stress and disease. *Journal of the American Medical Association, 298*, 1685–1687.
- Cohen, S., Tyrrell, D. A., & Smith, A. P. (1991). Psychological stress and susceptibility to the common cold. *New England Journal of Medicine, 325*, 606–612.
- Cole-Lewis, H., & Kershaw, T. (2010). Text messaging as a tool for behavior change in disease prevention and management. *Epidemiologic Reviews, 32*, 56–69.
- Curcio, G., Ferrara, M., & De Gennaro, L. (2006). Sleep loss, learning capacity and academic performance. *Sleep Medicine Reviews, 10*, 323–337.
- DeLongis, A., Folkman, S., & Lazarus, R. S. (1988). The impact of daily stress on health and mood: Psychological and social resources as mediators. *Journal of Personality and Social Psychology, 54*, 486–495.
- Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological Bulletin, 130*, 355–391.
- Dunbar-Jacob, J., & Mortimer-Stephens, M. (2001). Treatment adherence in chronic disease. *Journal of Clinical Epidemiology, 54*(12), S57–S60
- Eastin, M. S. (2001). Credibility assessments of online health information: The effects of source expertise and knowledge of content. *Journal of Computer Mediated Communication, 6*.
- Fox, S. & Jones, S. (2009). The social life of health information. *Pew Internet and American Life Project, California HealthCare Foundation*. Retrieved from www.pewinternet.org/Reports/2...formation.aspx
- Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist, 56*, 218–226.
- Freeman, K. S., & Spyridakis, J. H. (2004). An examination of factors that affect the credibility of online health information. *Technical Communication, 51*, 239–263.
- Friedman, M., & Rosenman, R. (1959). Association of specific overt behaviour pattern with blood and cardiovascular findings. *Journal of the American Medical Association, 169*, 1286–1296.
- Glass, D. C., & Singer, J. E. (1972). Behavioral aftereffects of unpredictable and uncontrollable aversive events: Although subjects were able to adapt to loud noise and other stressors in laboratory experiments, they clearly demonstrated adverse aftereffects. *American Scientist, 60*, 457–465.
- Herman-Stabl, M. A., Stemmler, M., & Petersen, A. C. (1995). Approach and avoidant coping: Implications for adolescent mental health. *Journal of Youth and Adolescence, 24*, 649–665.
- Holt-Lunstad, J., Smith, T. B., & Layton, J. B. (2010). Social relationships and mortality risk: a meta-analytic review. *PLoS Medicine, 7*(7), e1000316.
- House, J. S., Landis, K. R., & Umberson, D. (1988). Social relationships and health. *Science, 241*, 540–545.
- Iribarren, C., Sidney, S., Bild, D. E., Liu, K., Markovitz, J. H., Roseman, J. M., & Matthews, K. (2000). Association of hostility with coronary artery calcification in young adults. *Journal of the American Medical Association, 283*, 2546–2551.
- Kubzansky, L. D., Sparrow, D., Vokonas, P., & Kawachi, I. (2001). Is the glass half empty or half full? A prospective study of optimism and coronary heart disease in the normative aging study. *Psychosomatic Medicine, 63*, 910–916.
- Matthews, K. A., Glass, D. C., Rosenman, R. H., & Bortner, R. W. (1977). Competitive drive, pattern A, and coronary heart disease: A further analysis of some data from the Western Collaborative Group Study. *Journal of Chronic Diseases, 30*, 489–498.
- Meara, E., White, C., & Cutler, D. M. (2004). Trends in medical spending by age, 1963–2000. *Health Affairs, 23*, 176–183.
- Miller, T. Q., Smith, T. W., Turner, C. W., Guijarro, M. L., & Hallet, A. J. (1996). Meta-analytic review of research on hostility and physical health. *Psychological Bulletin, 119*, 322–348.
- Moravec, C. S. (2008). Biofeedback therapy in cardiovascular disease: rationale and research overview. *Cleveland Clinic Journal of Medicine, 75*, S35–S38.
- Nes, L. S., & Segerstrom, S. C. (2006). Dispositional optimism and coping: A meta-analytic review. *Personality and Social Psychology Review, 10*, 235–251.
- Oaten, M., & Cheng, K. (2005). Academic examination stress impairs self–control. *Journal of Social and Clinical Psychology, 24*, 254–279.
- O'Leary, A. (1985). Self-efficacy and health. *Behaviour Research and Therapy, 23*, 437–451.

- Patel, C., Marmot, M. G., & Terry, D. J. (1981). Controlled trial of biofeedback-aided behavioural methods in reducing mild hypertension. *British Medical Journal (Clinical research ed.)*, *282*, 2005–2008.
- Pressman, S. D., & Cohen, S. (2005). Does positive affect influence health? *Psychological Bulletin*, *131*, 925–971.
- Pressman, S. D., Gallagher, M. W., & Lopez, S. J. (2013). Is the emotion-health connection a “first-world problem”? *Psychological Science*, *24*, 544–549.
- Richman, L. S., Kubzansky, L., Maselko, J., Kawachi, I., Choo, P., & Bauer, M. (2005). Positive emotion and health: Going beyond the negative. *Health Psychology*, *24*, 422–429.
- Rodin, J., & Langer, E. J. (1977). Long-term effects of a control-relevant intervention with the institutionalized aged. *Journal of Personality and Social Psychology*, *35*, 897–902.
- Rutter, M. (1985). Resilience in the face of adversity. *British Journal of Psychiatry*, *147*, 598–611.
- Salmon, P. (2001). Effects of physical exercise on anxiety, depression, and sensitivity to stress: A unifying theory. *Clinical Psychology Review*, *21*(1), 33–61.
- Scheier, M. F., & Carver, C. S. (1985). Optimism, coping, and health: assessment and implications of generalized outcome expectancies. *Health Psychology*, *4*, 219–247.
- Schoenfeld, E. R., Greene, J. M., Wu, S. Y., & Leske, M. C. (2001). Patterns of adherence to diabetes vision care guidelines: Baseline findings from the Diabetic Retinopathy Awareness Program. *Ophthalmology*, *108*, 563–571.
- Schulz, R., & Hanusa, B.H. (1978). Long-term effects of control and predictability-enhancing interventions: Findings and ethical issues. *Journal of Personality and Social Psychology*, *36*, 1194–1202.
- Segerstrom, S. C., Taylor, S. E., Kemeny, M. E., & Fahey, J. L. (1998). Optimism is associated with mood, coping, and immune change in response to stress. *Journal of Personality and Social Psychology*, *74*, 1646–1655.
- Seligman, M. E. P. (2008). Positive health. *Applied Psychology*, *57*, 3–18.
- Selye, H. (1946). The general adaptation syndrome and the diseases of adaptation. *Journal of Clinical Endocrinology*, *6*, 117–230.
- Sieber, W. J., Rodin, J., Larson, L., Ortega, S., Cummings, N., Levy, S., ... Herberman, R. (1992). Modulation of human natural killer cell activity by exposure to uncontrollable stress. *Brain, Behavior, and Immunity*, *6*, 141–156.
- Spiegel, D., Kraemer, H., Bloom, J., & Gottheil, E. (1989). Effect of psychosocial treatment on survival of patients with metastatic breast cancer. *The Lancet*, *334*, 888–891.
- Taylor, S. E. (2012) *Health psychology* (8th ed.). New York, NY: McGraw-Hill.
- Taylor, S. E., Klein, L. C., Lewis, B. P., Gruenewald, T. L., Gurung, R. A., & Updegraff, J. A. (2000). Biobehavioral responses to stress in females: Tend-and-befriend, not fight-or-flight. *Psychological Review, 107*, 411–429.
- Twisk, J. W., Snel, J., Kemper, H. C., & van Mechelen, W. (1999). Changes in daily hassles and life events and the relationship with coronary heart disease risk factors: A 2-year longitudinal study in 27–29-year-old males and females. *Journal of Psychosomatic Research*, *46*, 229–240.
- Wallston, B. S., & Wallston, K. A. (1978). Locus of control and health: a review of the literature. *Health Education & Behavior*, *6*, 107–117.
- World Health Organization (2013). *Cardiovascular diseases*. Retrieved from www.who.int/mediacentre/facts.../en/index.html
- World Health Organization. (1946). *Preamble to the Constitution of the World Health Organization*. Retrieved from www.who.int/about/definition/en/print.html

16.5.14 Attributions

The Healthy Life by Emily Hooker and Sarah Pressman, University of Calfornia, Irvine, at Noba Project licensed CC BY-NC-SA 4.0

Enhancing versus Suppressive Effects of Stress on Immune Function: Implications for Immunoprotection versus Immunopathology. by Firdaus S Dhabhar in Allergy, Asthma and Clinical Immunology licensed https://creativecommons.org/licenses/by/2.0

Processes in the primary immune response by: Sciencia58 an the makers of the single images Domdomegg, [1], Fæ, Petr94, Manu5, CC BY-SA 4.0, via Wikimedia Commons

Introduction to the Immune system by Suzanne Wakim and Mandeep Grewal, Butte College, licensed CC BY-NC 4.0

Lymphatic system by Suzanne Wakim and Mandeep Grewal, Butte College, licensed CC BY-NC 4.0

Adaptive immune system by Suzanne Wakim and Mandeep Grewal, Butte College, licensed CC BY-NC 4.0

Concepts of Biology by Openstax, licensed CC BY 4.0

Inflammation and Fever by Openstax, licensed CC BY 4.0

This page titled 16.5: Stress and health is shared under a mixed license and was authored, remixed, and/or curated by ASCCC OERI & Bakhtawar Bhadha (ASCCC Open Educational Resources Initiative (OERI)) .

CHAPTER OVERVIEW

17: Biological Bases of Psychological Disorders

> This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

17.1: Biological Factors in Psychological Disorders- An Introduction
17.2: Biological Basis of Mood Disorders- Major Depressive and Bipolar Disorders
17.3: Biological Basis of Anxiety Disorders
17.4: Biological Basis of Schizophrenia Spectrum Disorder

This page titled 17: Biological Bases of Psychological Disorders is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

17.1: Biological Factors in Psychological Disorders- An Introduction

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Identify 3 biological areas that contribute to psychological disorders.
- Describe the role of neural structures in psychological disorders.
- Identify 2 systems of chemical communication used by the brain.
- Describe how genetics and epigenetics contribute to psychological disorders.

17.1.1 Chapter Overview

This chapter discusses the biological perspective on psychological disorders, exploring how our brain structures, neural systems, and genetics contribute to the etiology of such disorders as schizophrenia, depression, bipolar, anxiety, and obsessive/compulsion.

17.1.2 Introduction

by Amy E. Coren, Ph.D., Pasadena City College

Although the term mental illness is often used when referring to disorders such as depression or anxiety, are these disorders really illnesses in the way that diabetes is an illness – that is a purely biological malfunction? Or are psychological disorders the product of something more intangible than a neurochemical imbalance or malfunctioning amygdala, something such as a conditioned response to stress learned in childhood? The answer is both.

Within the field of psychology, it is generally acknowledged that psychological disorders develop out of a complex interaction of biological, social, and environmental factors. However, exactly how these factors interact with one another to produce a particular psychological disorder is still unclear. This is because diagnosing a psychological disorder is not like diagnosing other illnesses, such as diabetes or pneumonia.

When making a diagnosis of diabetes, a doctor may use a blood test to determine blood glucose levels. Similarly, to diagnose pneumonia, a doctor may order a chest x-ray to look for evidence of the infection. However, when it comes to diagnosing psychological disorders, there is currently no x-ray for depression or blood test for anxiety. Diagnosing psychological disorders is a more subjective process, based on the current symptoms an individual is experiencing.

The Diagnostic and Statistical Manual of Mental Disorders, 5th edition, (American Psychiatric Association, 2013) is the most widely accepted system used by clinicians and researchers for the classification of mental disorders, and outlines the diagnostic criteria for a specific mental disorder based on the symptoms an individual is experiencing. However, individuals suffering from the same disorder often display different symptoms while individuals suffering from different disorders may display many of the same symptoms, making the diagnosis of psychological disorders from outward symptoms alone, difficult.

Note

The American Psychological Association periodically updates the Diagnostic and Statistical Manual of Mental Disorders - also called the **DSM** - so that it will accurately reflect our current understanding of the symptoms, etiology, and treatment of psychological disorders. The revisions to the DSM-5 were released in early 2022, in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR).

The remainder of this chapter focuses on the contributions of biological factors - such as brain structure abnormalities - to the development of psychological disorders.

17.1.3 Biology of Psychological Disorders

Research into the underlying biology of psychological disorders has primarily focused on the following three areas: (1) the structures of the brain, (2) the biochemistry of the brain, and (3) genetics and epigenetics.

17.1.3.0.1 Structures of the Brain

Research has shown that some psychological disorders appear to involve specific structures within the brain. For example, overactivity in an area of the brain known as Brodmann Area 25 (BA25), shown in Figure 17.1.1, is often present in individuals with clinical depression (also known as Major Depressive Disorder) (Mayberg et al., 2005). BA25 acts as a "junction box" that interacts, through the use of neurotransmitters such as dopamine, with other areas of the brain involved in mood, emotion and thinking. When the BA25 region is overactive, a person can experience an increased negative affect (e.g. sadness, anxiety) and decreased positive affect (e.g. happiness, joy) (Jumah & Dossani, 2021). We will explore the role of brain structures in more detail as we discuss specific disorders later in this chapter.

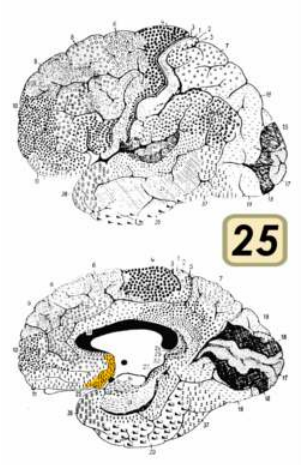


Figure **17.1.1**: **Brodmann Area 25 (BA25) is a narrow region of the brain located in the anterior cingulate cortex and is a part of the limbic system. BA25 is shown here colored in yellow.** (Public Domain; By Brodmann; Colored by was_a_bee, via Wikimedia Commons)

Dig Deeper

Check out the following article from Harvard Health for more on the brain and the biological factors that play a role in depressive disorders:
https://www.health.harvard.edu/mind-and-mood/what-causes-depression

17.1.3.0.1 Biochemistry of the Brain

In addition to the individual structures of the brain, researchers are also interested in what role our chemical communication systems – hormones and neurotransmitters –play in the development of psychological disorders.

The primary communication in the brain occurs between neurons (neurotransmission) using chemicals known as neurotransmitters. When the levels of these neurotransmitters are out of balance, the communication between neurons can be altered or disrupted, resulting in a display of symptoms associated with various psychological disorders.

For example, researchers believe that disruptions in the neurotransmitters, particularly dopamine, play an important role in schizophrenia (NIH, 2007). This link between dopamine and schizophrenia arose out of observations that individuals addicted to cocaine sometimes showed symptoms very similar to those seen in cases of schizophrenia. Cocaine works by reducing the amount of monoamine neurotransmitters (dopamine, norepinephrine, epinephrine, and serotonin) that are taken back into the presynaptic neuron (Richards & Laurin, 2021), as shown in Figure 2. When this happens, more monoamine neurotransmitters, crucially dopamine, are available in the synapse for binding to the receptors on the postsynaptic neuron. We will discuss the dopamine theory of schizophrenia in more detail later in this chapter.

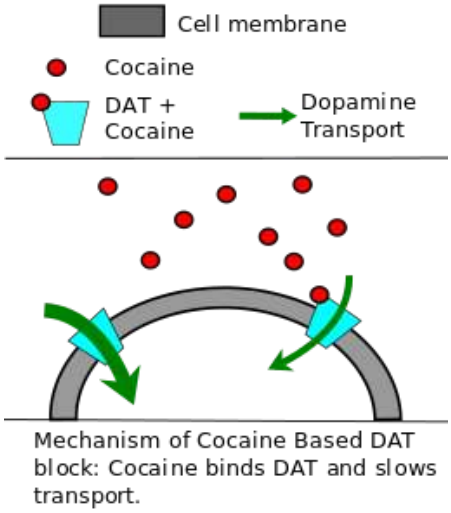


Figure **17.1.2**: **How cocaine increases the amount of dopamine in the synapse by reducing the activity of the dopamine reuptake transporter (DAT)**; (CC BY-SA 3.0; Hard-59 via Wikimedia Commons.)

In addition to neurotransmitters, the brain also communicates via hormones, which are released by the endocrine system, and activate behaviors such as alertness or sleepiness, concentration, and reactions to stress. Elevated or depleted levels of certain hormones may be responsible for some of the symptoms seen in psychological disorders. For example, elevated levels of cortisol, a stress hormone, interfere with learning and memory as well as increase the risk of depression. High levels of cortisol have been shown to alter the function of serotonin receptors in the brain, leading to symptoms of depression (Qin et al., 2018).

17.1.3.0.1 Genetics and Epigenetics

Researchers have long recognized that many psychological disorders tend to run in families, suggesting a potential genetic factor. In family and twin studies, schizophrenia is significantly more likely to be present in an identical twin than a fraternal twin (Coon & Miller, 2007) and in individuals with a first-degree relative (e.g. mother or father) with schizophrenia.

Researchers have also linked several genetic variations, or mutations, to psychological disorders, including variations in two genes which code for the cellular machinery that helps regulate the flow of calcium into neurons (NIH, 2013). One of these calcium channel genes, CACNA1C, is known to affect the brain circuitry involved in emotion, thinking, attention and memory. Variations in this gene have been linked to disorders such as bipolar disorder, schizophrenia, and major depression.



Figure **17.1.3**: **Dr. Bruce Cuthbert, Ph.D., director of NIMH's Division of Adult Translational Research, explains the significance of genetic findings for diagnosis and treatment of mental illnesses.**

Although individual genes, such as CACNA1C, have been linked to psychological disorders, most disorders are believed to be polygenic; that is linked to abnormalities in many genes, rather than just one. It is the complex interaction between multiple genes which may trigger a psychological disorder.

Furthermore, while researchers currently believe that genetic factors are implicated in all psychological disorders, they are not believed to be the sole cause. There are important gene-environment interactions, also known as epigenetic factors, unique to each individual (even identical twins) which may explain why some individuals with a genetic predisposition toward a certain disorder develop that disorder while others do not (e.g., why one identical twin develops schizophrenia, but the other does not).

17.1.3.1 References

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Helen S. Mayberg, Andres M. Lozano, Valerie Voon, Heather E. McNeely, David Seminowicz, Clement Hamani, Jason M. Schwalb, and Sidney H. Kennedy (2005). Deep Brain Stimulation for Treatment-Resistant Depression. Neuron, 45(5), 651-660. https://doi.org/10.1016/j.neuron.2005.02.014

Jumah FR, Dossani RH. (2021) Neuroanatomy, Cingulate Cortex. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK537077/

National Institutes of Health (US); Biological Sciences Curriculum Study. NIH Curriculum Supplement Series [Internet]. Bethesda (MD): National Institutes of Health (US); 2007. Information about Mental Illness and the Brain. Available from: https://www.ncbi.nlm.nih.gov/books/NBK20369/

Richards JR, Laurin EG. Cocaine. [Updated 2021 Jul 19]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK430769/

Qin, Dd., Rizak, J., Feng, Xl. et al. (2016). Prolonged secretion of cortisol as a possible mechanism underlying stress and depressive behavior. Sci Rep 6(30187). https://doi.org/10.1038/srep30187

Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. The Lancet, 381(9875), 1371-1379. https://doi.org/10.1016/S0140-6736(12)62129-1.

17.1.3.2 Attributions

Chapter 17, Biological Basis of Psychological Disorders, 17.1. Biological Basis of Psychological Disorders: An Introduction. Original material written by Amy E. Coren, PhD, Pasadena City College, is licensed under CC BY-NC-SA 4.0.

Figure 17.1.1. Brodman Area 25. By Brodmann; Colured by was_a_bee; File:Brodmann_Cytoarchitectonics.PNG, Public Domain, via Wikimedia Commons.

Figure 17.1.2. Cocaine and DAT. RicHard-59, CC BY-SA 3.0, via Wikimedia Commons.

Figure 17.1.3 Dr. Bruce Cuthbert, Ph.D., director of NIMH's Division of Adult Translational Research, explains the significance of genetic findings for diagnosis and treatment of mental illnesses. (Video retrieved from https://youtu.be/8SDKV29NPaE).

This page titled 17.1: Biological Factors in Psychological Disorders- An Introduction is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

17.2: Biological Basis of Mood Disorders- Major Depressive and Bipolar Disorders

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Describe what is meant by the term “mood disorder”.
- Describe the differences between depressive and bipolar disorders.
- Identify the biology thought to underlie the etiology of Major Depressive Disorder.
- Identify the biology thought to underlie the etiology of Bipolar Disorder (I & II).

17.2.1 Chapter Overview

Mood disorders are a broad category of psychological disorders characterized by severe disturbances in mood and emotion. While these disorders share common features of extreme mood alteration, most research suggests that they have different biological origins (Cuellar et al., 2005).

17.2.2 An Introduction to Mood Disorders

The classification, **mood disorders,** is used by mental health professionals to describe a broad group of psychological disorders characterized by severe disturbances in mood and emotion. Also known as **affective disorders**, disorders in this category are typically classified into two distinct groups: depressive disorders and bipolar disorders (Figure 17.2.1.).

Depressive disorders typically include major depressive disorder, also known as clinical depression, which is characterized by episodes of profound sadness and loss of interest or pleasure in usual activities, among other features, and persistent depressive disorder (formerly **dysthymia**), a milder but long-term, chronic form of depression. Bipolar disorders are characterized by mood states that vacillate between sadness and euphoria and include bipolar disorder types I and II, as well as cyclothymic disorder (APA, 2013).

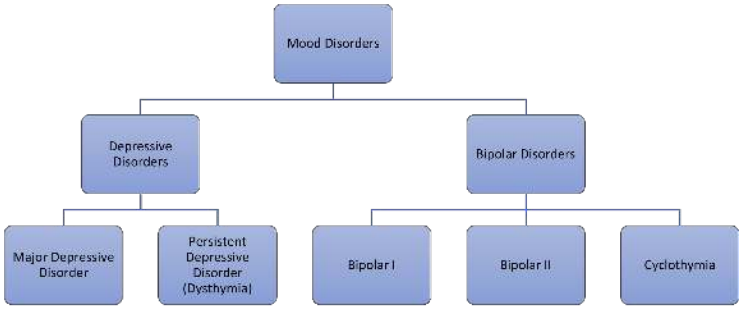


Figure **17.2.1**: **Mood disorders are typically grouped in two categories: (1) Depressive disorders, which include Major Depressive Disorder and Persistent Depressive Disorder (Dysthymia) and (2) Bipolar disorders which includes Bipolar I & II and Cyclothymia.** (CC BY-NC-SA 4.0; by Amy E. Coren, Ph.D.)

17.2.3 Depressive Disorders

Depressive disorders are a group of disorders whose main features include a depressed mood for most of the day and/or feelings of **anhedonia** (the inability to feel pleasure from activities usual found enjoyable) (Post & Warden, 2018).

The DSM-5 describes the main features of depressive disorders as “the presence of sad, empty, or irritable mood, accompanied by somatic and cognitive changes that significantly affect the individual’s capacity to function.” (APA, 2013) Depressed people often

report feeling sad, discouraged, and hopeless (Figure 17.2.2) These individuals lose interest in activities they once enjoyed, and often experience a decrease in drives such as hunger and sex, and frequently doubt their personal worth. When these features are triggered by a traumatic or negative experience (e.g. death of close loved one), it is called **reactive depression**. When there is no apparent reason or cause, it is known as **endogenous depression.**



Figure **17.2.2**: **Individuals suffering from a depressive disorder often feel sad, discouraged, and alone.** (Public Domain; Jack Lucas Smith via Unsplash)

Depressive disorders are often highly **comorbid** (meaning more than one health condition is present in the same person at the same time) with other psychological disorders. Depressive disorders are often seen occurring alongside anxiety disorders, bipolar disorders, and OCD and related disorders (Groen, et al., 2020).

Perhaps the most well-known example of a depressive disorder is **Major Depressive Disorder (MDD)**, also known as **clinical depression**. Major depressive disorder is believed to affect about 2-5% of the global population each year (Han, et al., 2019). According to the DSM-5, the defining symptoms of **major depressive disorder** include “depressed mood most of the day, nearly every day” (feeling sad, empty, hopeless, or appearing tearful to others) and loss of interest and pleasure in usual activities (APA, 2013). When these symptoms last for longer than 2 weeks, and significantly impair an individual’s ability to function, that individual is said to be suffering from MDD.

Researchers and mental health professionals both agree that there is no single cause responsible for any of the depressive disorders. Rather it is thought that depressive disorders arise from the interaction between multiple biological and environmental factors. However, this interaction is thought to be extremely complex and is still not well understood. As such, the biology underlying depressive disorders continues to be a major subject of ongoing research and is explored in more detail below.

17.2.3.1 The Genetics and Epigenetics of Depressive Disorders

Historically, evidence from family, twin, and adoption studies has seemed to support the idea that genetic factors play an important role in the development of depressive disorders, although which genes are involved and the how these genes interact to contribute to the development of depressive disorders is still not understood (Lohoff, 2010).

In 2015, several genome-wide association studies (GWASs), which involve scanning the genomes of large groups of people to find genetic variations associated with a disorder, identified over 100 possible genes which may play a role in increasing an individual’s risk for developing a depressive disorder (Ormel, et al., 2019, Shadrina, et al., 2018).

As a result of these studies, and others, it is now believed that an individual’s genetic predisposition to a depressive disorder is determined by (1) the coordinated action of many genes, (2) the genes’ interactions with each other, and (3) their interactions with environmental factors through epigenetic mechanisms (Shardina, et al., 2018).

Current research is now focusing on the role of epigenetics - how behaviors and environment can cause changes that affect the way our genes work, in the development of depressive disorders such as MDD (Barbu, et al., 2020).

Dig Deeper

To learn more about the interaction of biology and environment in the development of depressive disorders, check out this video lecture provided by Stanford Univeristy professor, Dr. Robert Sapolsky:



17.2.3.2 The Role of Neurotransmitters in Depressive Disorders

Until relatively recently, low levels of neurotransmitters in the brain were believed to be the main culprits behind depressive disorders. However, current research now suggests that, while neurotransmitters do play an important role, the reality is much more complicated than previously thought.

17.2.3.2.1 *Monoamine Theory of Depression*

It has historically been believed that low activity levels of monoamine neurotransmitters (norepinephrine, serotonin, and dopamine) contribute to the symptoms associated with depressive disorders (Nutt, 2008). The discovery that certain drugs which increased levels of serotonin and norepinephrine in the brain, alleviated depressive symptoms, lead to formulation of the monoamine theory of depression by Joseph Schildkraut in the 1960s. According to the monoamine theory, low levels of serotonin, norepinephrine, and/or dopamine in the central nervous system are primarily the basis for the development of depressive disorders (Hirschfeld, 2000). This theory prevailed for more than 50 years before being challenged by an increasing number of research studies and clinical observations.

Today, most current research acknowledges that while monoamine neurotransmitters are important, they are by far not the *only* neurotransmitters involved in depressive disorders. The following video from Yale Medicine explains why the monoamine theory of depression is no longer believed to be correct and what alternative theories researchers are currently exploring. For example, other neurotransmitters, such as glutamate and GABA, are increasingly being investigated for their role in the development of depression.



Furthermore, research is increasingly focused not only on individual neurotransmitters, but also on the *interactions* between the various neurotransmitters themselves as well as how they interact with other biological systems such as the endocrine system (Ding et al., 2014). For example, serotonin production and activity are affected by the hormones secreted by the endocrine system, such as cortisol, in response to threat or stress.

17.2.3.3 The role of neural structures in MDD

Research into the neural basis of MDD has focused on examining the neural architecture of individuals with MDD by using a variety of both structural and functional neuroimaging techniques. Structural techniques, such as CT scans and magnetic resonance imaging (MRI), take static pictures of the brain to determine whether any specific neural structures are different in subjects with MDD than control subjects.

By using these static neuroimaging techniques, researchers have discovered several areas of the brain where the volume of **grey matter** is significantly decreased in those individuals with MDD (Filatova, et al., 2021). These areas include the prefrontal cortex (PFC), hippocampus, amygdala, and cingulate cortex (Ancelin, et al. 2019).

Importantly, these areas comprise most of the major structures in the brain's **corticolimbic system**, a system responsible for regulating multiple behavioral and cognitive functions, including decision making, motivation, emotional processing, and our response to stress and pain (Vachon-Presseau, 2018) (Figure 16.2.3.).

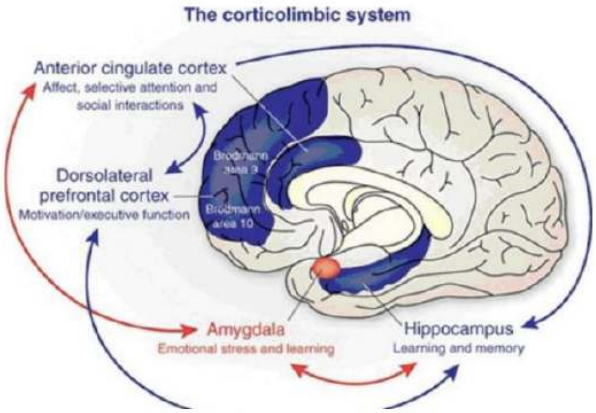


Figure **17.2.3**: **The corticolimbic system in the brain regulates several important behavioral and cognitive functions, including our reactions to stress and pain.** (CC BY-SA 4.0; Benes 2010 via Wikimedia Commons)

In addition to looking at the overall structure of the brain, researchers also use functional neuroimaging techniques, such as positron emission tomography (PET) scans and functional magnetic resonance imaging (fMRI), to examine the brain in action, looking for any areas of atypical activation.

For example, a meta-analysis of neuroimaging studies showed that when viewing negative stimuli (e.g., picture of an angry face, picture of a car accident), participants with MDD have *greater* activation in brain regions involved in stress response, such as the amygdala and anterior cingulate cortex, and *reduced* activation in brain regions involved in positively motivated behaviors, such as the prefrontal cortex, compared with healthy control participants. (Hamilton, et al., 2012).

Other functional imaging studies have examined alterations in the **functional connectivity** - the strength of the connections between regions of the brain. These connections allow multiple brain regions to properly "talk" to one another; that is to perceive, generate, and encode information in concert. By examining these neural connections, researchers can see if there are problems in the way the brain processes specific types of information in individuals with MDD, and where these "faulty" connections may occur (Goldstein-Peikarski, et al. 2018).

Several studies have noted differences in the neural networks of individuals with MDD, compared to those without. For example, abnormally low activity in brain regions related to stopping thoughts and shifting to new ones, referred to in the research as "cognitive control," and hyperactivity in other brain regions that "process emotional thoughts and feelings" has consistently been found in individuals with MDD (Janiri, et al., 2019).

17.2.4 Bipolar Disorders

Some people who suffer from clinical depression may also experience what are known as **manic** or **hypomanic** episodes. When this occurs, the individual is usually diagnosed with **bipolar disorder**.

Hypomanic episodes are fairly short– around 4 days in duration, and are characterized by a positive mood, reduced need for sleep, and high energy. Individuals experiencing a hypomanic episode are often talkative, impulsive, energetic, and very confident. It is important to note that by definition, hypomanic episodes cannot cause impairment, distress, or the need for hospitalization. If any of these three features are present the episode is considered to be manic, rather than hypomanic.

Manic episodes are longer, at least 1 week in duration, and have features similar to hypomania but taken to an extreme. In addition, manic episodes are often characterized by delusions of grandeur, psychosis, and distractibility. Additionally, for an episode to be considered manic, it must cause impairment in functioning, significant distress, or require the individual to be hospitalized.

Bipolar I disorder, previously known as manic-depressive disorder, is diagnosed when there is at least one manic episode. This manic episode may be preceded by or followed by a major depressive episode but that is not required for the Bipolar I diagnosis. In contrast, a diagnosis of **bipolar II disorder** is made when the individual has experienced *both* a hypomanic episode and a depressive episode, but no manic episodes. Alternatively, an individual may suffer from **cyclothymic disorder**, which characterized by multiple, alternating periods of hypomania and depression, lasting at least two years. To qualify for cyclothymic disorder, the person must experience symptoms at least half the time with no more than two consecutive symptom-free months; and the symptoms must cause significant distress or impairment (APA, 2013).

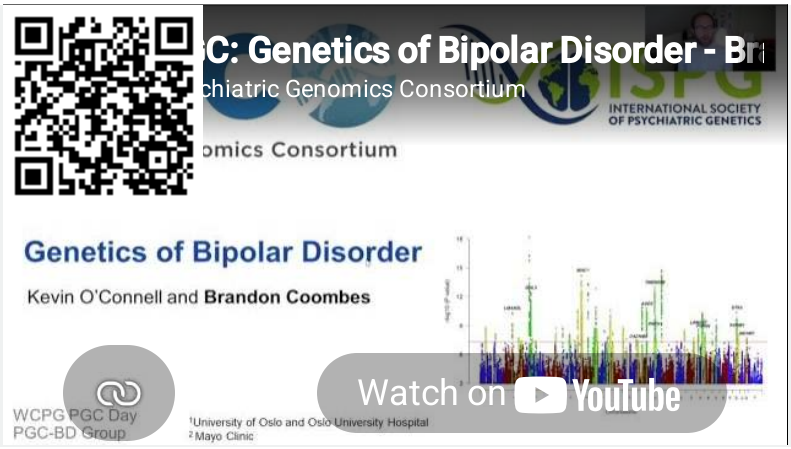
17.2.4.1 The Genetics and Epigenetics of Bipolar Disorders

Numerous twin and family studies have shown that there is a strong genetic component to bipolar disorders, especially Bipolar I (Edvardsen et al., 2008, Escamilla & Zavala, 2008). However, the exact nature of this genetic component remains elusive – numerous studies, include a genome-wide association study (GWAS) of more than 40,000 bipolar disorder cases, have identified hundreds to thousands of genes associated with bipolar disorder (Gandal, et al. 2018).

Given the complicated genetic component, recent research has focused on the identifying epigenetic mechanisms which may play role in the risk and development of bipolar disorders (Duffy, et al. 2019).

Dig Deeper

Genome-wide association studies, or GWAS, have become an increasingly popular tool for exploring the genetic and epigenetic factors involved in psychological disorders, such as Biopolar Disorders. To learn more about GWAS and their use in Bipolar disorder research, check out the following video:



17.2.4.2 The Role of Neurotransmitters and Neural Structures in Bipolar Disorders

As depression was often a characteristic of bipolar disorders, researchers initially believed that norepinephrine, serotonin, and dopamine were all implicated in the development of bipolar disorder. It was thought that manic episodes were the result of drastic increases in serotonin. Unfortunately, research did not support this hypothesis. It is now believed that manic episodes may, in fact, be explained by low levels of serotonin and high levels of norepinephrine (Soreff & McInnes, 2014).

In addition, consistent decreases in gray matter volume in both the prefrontal cortex and areas of the limbic system, such as the hippocampus, have been reported, as have enlargement of the ventricles (Jiang, et al. 2020).

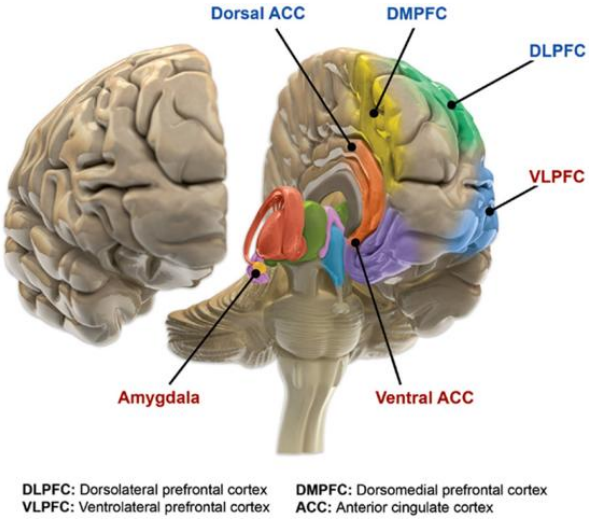


Figure 17.2.4: **Regions of the brain implicated in Bipolar disorders.** (CC BY 4.0; Maletic V, Raison C via Wikimedia Commons)

Studies of brain activity have led to the view that bipolar disorder may be a suite of related neurological issues with interconnected functional abnormalities that often appear early in life and worsen over time. In support of this hypothesis, fMRI techniques have noted atypical activation of the frontal cortex and basil ganglia, as well as disruption in the connectivity between these structures (Maletic & Rasion, 2014). The issues in connectivity seem to cluster around those networks in the brain associated with emotional processing.

17.2.5 References

Ancelin, M. L., Carrière, I., Artero, S., Maller, J., Meslin, C., Ritchie, K., Ryan, J., & Chaudieu, I. (2019). Lifetime major depression and grey-matter volume. *Journal of psychiatry & neuroscience, JPN*, *44*(1), 45–53. https://doi.org/10.1503/jpn.180026

Cuellar, A. K., Johnson, S. L., & Winters, R. (2005). Distinctions between bipolar and unipolar depression. *Clinical psychology review*, *25*(3), 307–339. https://doi.org/10.1016/j.cpr.2004.12.002

Duffy, A., Goodday, S.M., Keown-Stoneman, C. *et al.* (2019). Epigenetic markers in inflammation-related genes associated with mood disorder: a cross-sectional and longitudinal study in high-risk offspring of bipolar parents. *Int J Bipolar Disord* **7(**17). https://doi.org/10.1186/s40345-019-0152-1

Escamilla, M. A., & Zavala, J. M. (2008). Genetics of bipolar disorder. *Dialogues in clinical neuroscience*, *10*(2), 141–152. https://doi.org/10.31887/DCNS.2008.10.2/maescamilla

Filatova, E. V., Shadrina, M. I., & Slominsky, P. A. (2021). Major Depression: One Brain, One Disease, One Set of Intertwined Processes. *Cells*, *10*(6), 1283. https://doi.org/10.3390/cells10061283

Goldstein-Piekarski, A.N., Staveland, B.R., Ball, T.M. *et al.* Intrinsic functional connectivity predicts remission on antidepressants: a randomized controlled trial to identify clinically applicable imaging biomarkers. *Transl. Psychiatry* 8, 57 (2018). https://doi.org/10.1038/s41398-018-0100-3

Groen, RN, Ryan, O., Wigman, JTW., Harriëtte R., Brenda, WJH, Penninx, EJ, Giltay, MW, & Hartman, CA (2020). Comorbidity between depression and anxiety: assessing the role of bridge mental states in dynamic psychological networks. *BMC Med*, 18(308). https://doi.org/10.1186/s12916-020-01738-z

Hamilton JP, Etkin A, Furman DJ, Lemus MG, Johnson RF, Gotlib IH. (2012). Functional neuroimaging of major depressive disorder: a meta-analysis and new integration of base line activation and neural response data. *Am J Psychiatry*, *169*(7), 693-703. doi: 10.1176/appi.ajp.2012.11071105. PMID: 22535198.

Han, MH, Russo, SJ, & Nestler, EJ (2019). Molecular, Cellular, and Circuit Basis of Depression Susceptibility and Resilience. In *Neurobiology of Depression* (pp. 123-136). Academic Press. https://doi.org/10.1016/B978-0-12-813333-0.00012-3.

Hirschfeld, R. (2000). History and evolution of the monoamine hypothesis of depression. *The Journal of Clinical Psychiatry*, 61(6), 4-6.

Janiri D, Moser DA, Doucet GE, et al. Shared Neural Phenotypes for Mood and Anxiety Disorders: A Meta-analysis of 226 Task-Related Functional Imaging Studies. JAMA Psychiatry. 2020;77(2):172–179. doi:10.1001/jamapsychiatry.2019.3351

Jiang, W., Andreassen, O. A., Agartz, I., Lagerberg, T. V., Westlye, L. T., Calhoun, V. D., & Turner, J. A. (2020). Distinct structural brain circuits indicate mood and apathy profiles in bipolar disorder. *NeuroImage. Clinical*, *26*, 101989. https://doi.org/10.1016/j.nicl.2019.101989

Lohoff F. W. (2010). Overview of the genetics of major depressive disorder. *Current psychiatry reports*, *12*(6), 539–546. https://doi.org/10.1007/s11920-010-0150-6

Maletic V & Raison C. (2014) Integrated neurobiology of bipolar disorder. *Front Psychiatry*. 5(98). doi: 10.3389/fpsyt.2014.00098. PMID: 25202283; PMCID: PMC4142322.

Nutt DJ. (2008). Relationship of neurotransmitters to the symptoms of major depressive disorder. *J Clin Psychiatry*, 69(Suppl E1), 4-7.

Ormel J, Hartman CA, Snieder H. (2019). The genetics of depression: successful genome-wide association studies introduce new challenges. *Transl Psychiatry*. *9*(1), 114. doi: 10.1038/s41398-019-0450-5. PMID: 30877272; PMCID: PMC6420566.

Post, RJ & Waren, MR (2018). Melancholy, Anhedonia, Apathy: The search for separable behaviors and neural circuits in depression. *Current Opinions in Neurobiology*, *49*, 192-200.

Shadrina, M., Bondarenko, E. A., & Slominsky, P. A. (2018). Genetics Factors in Major Depression Disease. *Frontiers in psychiatry*, *9*, 334. https://doi.org/10.3389/fpsyt.2018.00334

Vachon-Presseau, E. (2018) Effects of stress on the corticolimbic system: implications for chronic pain. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *87*(Part B), 216-223.

17.2.6 Attributions

Chapter 17, Biological Basis of Psychological Disorders, 17.2. Biological Bases of Mood Disorders original material written by Amy E. Coren, PhD, Pasadena City College, is licensed under CC BY-NC-SA 4.0. Some text of sections 17.2.3 and 17.2.4 adapted and modified by Amy E. Coren, Ph.D., Pasadena City College, from 6.1 Depressive Disorders Etiology & 6.2 Bipolar Disorders

Etiology in *Essentials of Abnormal Psychology* by Alexis Bridley & Lee W. Daffin Jr., Washington State University; licensed under CC-BY-NC-SA 4.0 International License. Retrieved from: https://opentext.wsu.edu/abnormalpsy...g-information/

Fig. 17.2.1 Original creation by author, Amy E. Coren, Ph.D.

Fig. 17.2.2 Photo by Jack Lucas Smith on Unsplash

Fig. 17.2.3. Benes 2010, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons CC BY 4.0

Fig. 17.2.4 Maletic V, Raison C, CC BY 4.0 <https://creativecommons.org/licenses/by/4.0>, via Wikimedia Commons CC BY 4.0

This page titled 17.2: Biological Basis of Mood Disorders- Major Depressive and Bipolar Disorders is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

17.3: Biological Basis of Anxiety Disorders

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Describe the principal features of anxiety disorders.
- Describe the difference between fear and anxiety.
- Identify the major biological systems involved in anxiety and fear.
- Identify the biology thought to underlie the etiology of anxiety disorders.

17.3.1 Overview

Anxiety disorders are the most prevalent of all psychological disorders. Large scale surveys estimate that around 33% of the population is likely to suffer from an anxiety disorder at some point in their lives (Bandelow & Michaelis, 2015). The category of anxiety disorders encompasses a broad range of disorders marked by severe levels of fear and anxiety. Similar to the other disorders discussed in this chapter, the etiology of anxiety disorders is thought to be a complex combination of biological and environmental factors.

17.3.2 Introduction to Anxiety Disorders

The main feature common to all individuals suffering from an anxiety disorder is the experience of excessive **anxiety**. Similar to the emotion, fear, anxiety acts as a signal of danger, threat, or motivational conflict, and prepares the body for action. While *fear* is a physical and emotional response to a real or external imminent threat, *anxiety* is a generalized response to an unknown threat or internal conflict, often accompanied by bodily symptoms such as increased heart rate, muscle tension, a sense of unease, and apprehension about the future (APA, 2013). The category of anxiety disorders includes generalized anxiety disorder, panic disorder, specific phobia, and social anxiety disorder (social phobia) (APA, 2013).



Figure 17.3.1: **Anxiety is generalized, unfocused fear that may be triggered by a variety of sources, both internal and external.** (Public Domain; Mohammad_Hassan via Pixabay)

Like many psychological disorders, anxiety disorders are believed to arise from a complex blend of factors, both biological and environmental, that when combined with stress, lead to the development of a particular anxiety disorder.

Although all the anxiety disorders share the feeling of anxiety as a symptom, there are many different ways in which this anxiety may manifest in an individual, leading to the diversity seen among the disorders in this category.

17.3.3 Genetic and epigenetic factors in anxiety disorders

Anxiety disorders are believed to be highly complex and polygenic (Meier & Decker, 2019). While genetics have been known to contribute to the presentation of anxiety symptoms, the interaction between genetics and stressful environmental influences

accounts for more anxiety disorders than genetics alone (Bienvenu, et al., 2011). Nevertheless, several genes have been identified that may contribute to the increased risk of the development of anxiety disorders.

One of the first promising genes to be identified was a serotonin transporter or SERT gene. The serotonin transporter is of major importance in regulating levels of serotonin in the synapse. Serotonin synapses play a central role in the neural circuitry controlling mood and temperament (Houwing, et al., 2017). Disturbances in the serotonin system are known to contribute to many psychological disorders (Andrews et al., 2015). Mutation of the SERT gene have been found to be related to a reduction in serotonin activity and an increase in anxiety-related personality traits (Munafo, et al, 2008).

In 2020, a large research study of over 200,000 participants identified a gene known as SATB1 (Levey, et al. 2020). This gene is believed to influence the activity ("expression") of multiple other genes involved in neuronal development, including a gene known as CRH1. The CRH1 gene codes for the protein, Corticotropin Releasing Hormone (*CRH*), which is a hormone that plays an essential role in our body's hypothalamic-pituitary-adrenal (HPA) axis, the pathway that modulates our stress and fear/anxiety responses.

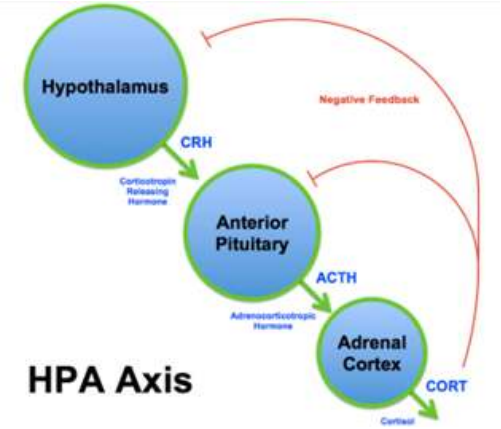


Figure **17.3.2**: **The body's HPA axis regulates stress responses. Genes which help maintain this system, may play a role in the risk of developing an anxiety disorder.** (CC BY-SA 3.0; BrianMSweis via Wikimedia Commons)

17.3.4 Neural structures and neurotransmitters involved anxiety disorders

Researchers have identified several brain structures and pathways that are may be responsible for the excessive anxiety responses often seen in anxiety disorders. In particular, the atypical activation in the prefrontal cortex, hippocampus, and amygdala has been implicated in anxiety disorders (Shin & Liberzon, 2010, Maron & Nutt, 2017).

A region of the brain called the locus coeruleus has been of particular interest to researchers studying panic disorder. Located in the brainstem, the locus coeruleus is the brain's major source of norepinephrine, a neurotransmitter that triggers the body's fight-or-flight response (Figure 16.3.4.1). Research with nonhuman primates has shown that stimulating the locus coeruleus either electrically or through drugs produces panic-like symptoms (Charney et al., 1990). Such findings have led to the theory that individuals with panic disorder may have a hyperactive locus coeruleus, leaving them an increased likelihood of experiencing more intense and frequent physiological arousal than the general public (Gorman, et al., 2000). This theory is supported by studies in which individuals experienced increased panic symptoms following injection of norepinephrine (Bourin, et al., 1995).

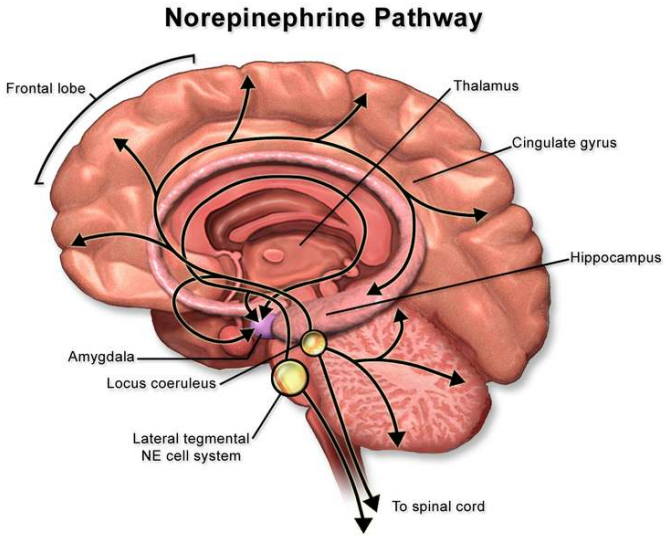


Figure **17.3.3**: **The locus coeruleus and the norepinephrine pathway.** (CC BY-SA 4.0; BruceBlaus via Wikimedia Commons)

Unfortunately, norepinephrine and the locus coeruleus fail to fully explain the development of panic disorder, and a more complex neuropathway is likely implicated in the development of panic disorder. More specifically, the corticostriatal-thalamocortical (CSTC) circuit, also known as the fear-specific circuit, is theorized as a major contributor to panic symptoms (Gutman, et al., 2004). When an individual is presented with a frightening object or situation, the amygdala is activated, sending a fear response to the anterior cingulate cortex and the orbitofrontal cortex. Additional projections from the amygdala to the hypothalamus activate endocrinologic responses to fear- releasing adrenaline and cortisol to help prepare the body to fight or flight (Gutman, et al. 2004).

17.3.5 References

Andrews P. W., Bharwani A., Lee K. R., Fox M., Thomson J. A. (2015). Is serotonin an upper or a downer? The evolution of the serotonergic system and its role in depression and the antidepressant response. *Neurosci. Biobehav. Rev. 51*, 164–188. DOI: 10.1016/j.neubiorev.2015.01.018.

Bandelow, B., & Michaelis, S. (2015). Epidemiology of anxiety disorders in the 21st century. *Dialogues in clinical neuroscience*, *17*(3), 327–335. https://doi.org/10.31887/DCNS.2015.17.3/bbandelow

Craig KJ., Brown KJ., Baum A. Environmental factors in the etiology of anxiety. In: Bloom FE, Kupfer DJ, eds. Psychopharmacology: the Fourth Generation of Progress. New York, NY: Raven Press; 1995:1325–1339.

Houwing DJ, Buwalda B., van der Zee EA, de Boer SF, Olivier JDA (2017). The Serotonin Transporter and Early Life Stress: Translational Perspectives. *Frontiers in Cellular Neuroscience*, *11*(Article 117). DOI=10.3389/fncel.2017.00117

Levey DF, Gelernter J, Polimanti R, Zhou H, Cheng Z, Aslan M, Quaden R, Concato J, Radhakrishnan K, Bryois J, Sullivan PF; Million Veteran Program, Stein MB. (2020). Reproducible Genetic Risk Loci for Anxiety: Results From ∼200,000 Participants in the Million Veteran Program. *Am J Psychiatry*, 177(3), 223-232. doi: 10.1176/appi.ajp.2019.19030256

Steimer T. (2002). The biology of fear- and anxiety-related behaviors. *Dialogues in clinical neuroscience*, *4*(3), 231–249. https://doi.org/10.31887/DCNS.2002.4.3/tsteimer

Shin, L. M., & Liberzon, I. (2010). The neurocircuitry of fear, stress, and anxiety disorders. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, *35*(1), 169–191. https://doi.org/10.1038/npp.2009.83

17.3.6 Attributions

Chapter 16, Biological Basis of Psychological Disorders, 17.3. Biological Bases of Anxiety Disorders, original material written by Amy E. Coren, PhD, Pasadena City College, is licensed under CC BY-NC-SA 4.0. Text of Section 17.3.4. Neural structures and neurotransmitters involved anxiety disorders modified and adapted by Amy E. Coren, Ph.D., Pasadena City College, from 4.6 Anxiety Disorders Etiology in *Essentials of Abnormal Psychology*, by Alexis Bridley & Lee W. Daffin Jr. at Washington State

University;licensed under CC-BY-NC-SA 4.0 International License. Retrieved from: https://opentext.wsu.edu/abnormalpsy...g-information/

Figure 17.3.1 Public Domaine, Mohammad_Hassan, via Pixabay.

Figure 17.3.2 BrianMSweis, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons.

Figure 17.3.3 BruceBlaus, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons. File retrieved from: https://upload.wikimedia.org/wikiped...ine_Part_1.png

---

This page titled 17.3: Biological Basis of Anxiety Disorders is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

17.4: Biological Basis of Schizophrenia Spectrum Disorder

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

- Describe the principal features of schizophrenia spectrum disorders.
- Identify the difference between positive and negative symptoms in schizophrenia.
- Identify the biology thought to underlie the etiology of schizophrenia.
- Describe the dopamine hypothesis of schizophrenia.

17.4.1 Section Overview

Schizophrenia spectrum and other psychotic disorders are a category of psychological disorders encompassing an incredibly diverse range of characteristics, including distortions in perception, impairments in cognition, and abnormal motor activity. In addition, the symptoms vary greatly among individuals, as there are rarely two cases similar in presentation, course, or responsiveness to treatment (APA, 2013).

This category also includes schizophreniform disorder (a briefer version of schizophrenia), schizoaffective disorder (a mixture of psychosis and depression/mania symptoms), delusional disorder (the experience of only delusions), and brief psychotic disorder (psychotic symptoms that last only a few days or weeks). However, as many of the symptoms of these disorders overlap with schizophrenia, this disorder is the primary focus of this section.

17.4.2 Introduction to Schizophrenia

Originally termed *dementia praecox* (Latin for premature dementia), schizophrenia involves a diverse and complex set of symptoms, such as delusions, hallucinations, disorganized speech and behavior, abnormal motor behavior (including catatonia), anhedonia/amotivation and blunted affect/reduced speech (APA, 2013). Given this diversity of symptoms, and the fact that the presentation of symptoms can vary widely from case to case, it has been difficult for researchers and clinicians to precisely define schizophrenia as a disorder (Tandor, 2013). Growing evidence suggests that schizophrenia represents a broad and heterogenous syndrome, rather than a specific disorder, and is better classified along a psychosis spectrum (APA, 2013, Cuthbert & Morris, 2021).

When considering the symptoms of schizophrenia, it is common practice to divide the symptoms into **positive symptoms** (symptoms that demonstrate an excess of function) and **negative symptoms** (symptoms that represent a loss or reduction in function) (see Figure 17.4.2.1).

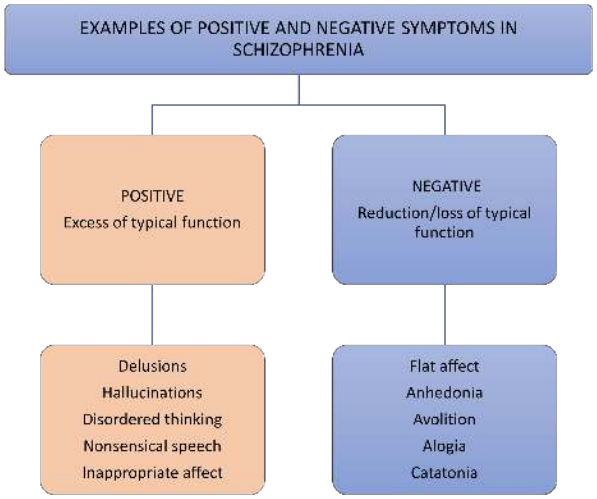


Figure **17.4.1**: **Schizophrenia symptoms are often classified as positive (an excess of typical function) or negative (a loss/reduction in typical function).** (CC BY-SA 4.0; Amy E. Coren, Ph.D.)

The most common positive symptoms are **delusions** and **hallucinations**. Delusions are "fixed beliefs that are not amenable to change in light of conflicting evidence" (APA, 2013, pp. 87). An individual suffering from delusions remains convinced in their belief, even when presented with clear, contradicting evidence. Researchers believe that delusions primarily relate to the social, emotional, educational, and cultural background of the individual (Arango & Carpenter, 2010). For example, an individual with schizophrenia who comes from a highly religious family is more likely to experience religious delusions (delusions of grandeur) than another type of delusion.

Hallucinations are perceptual experiences with no external stimulus (such as hearing voices when no one else is present). Hallucinations can occur in any sensory modality: auditory, visual, olfactory (smell), gustatory (taste), or somatic (touch), however the most common type are auditory hallucinations (APA, 2013). The content of these auditory hallucinations is frequently negative, making critical comments ("you're worthless") or telling the person what to do.

Negative symptoms typically involve an impairment in functions – cognitive, physical, and/or emotional. For example, anhedonia reflects a lack of apparent drive to engage in social or recreational activities. Importantly, this does not seem to reflect a lack of enjoyment in pleasurable activities or events (as is often seen in Depressive Disorders) (Llerena, et al., 2012) but rather a reduced drive or ability to take the steps necessary to obtain the potentially positive outcomes (Barch & Dowd, 2010). Flat affect and reduced speech (alogia) reflect a lack of showing emotions through facial expressions, gestures, and speech intonation, as well as a reduced amount of speech and increased pause frequency and duration.

17.4.2.1 Genetics and Epigenetics in Schizophrenia

Numerous family and twin studies have provided evidence that genetics play a significant role in an individual's risk of developing schizophrenia (Henriksen, et al., 2017). However, as is the case with many psychological disorders, the genetics underlying schizophrenia have proven to be highly complex, heterogeneous, and polygenic. Furthermore, the wide variations in symptoms among cases (i.e., one individual may demonstrate only negative symptoms, while another may present only positive symptoms) makes it even more challenging to identify specific genes associated with risk for schizophrenia.

Genome-wide studies have identified more than 70 genes suspected of playing a role in producing schizophrenia (Flint & Munafò, 2014), and it is believed that the development of schizophrenia involves the cumulative effects of multiple genes, with most of these having only a small effect by themselves.

While our understanding of the genetic basis underlying schizophrenia is still developing, recent research has noted several genes associated with various aspects of brain development, including myelination and synaptic transmission (Smigielski, et al. 2020), that are also associated with an increased susceptibility to schizophrenia (Gürel, et al. 2020). It is now believed that this genetic susceptibility may be triggered by early environmental factors, through epigenetic mechanisms, and alter the course of neural development, leading to the later development of schizophrenia. (Gürel, et al. 2020.)

17.4.2.2 The Role of Neurotransmitters in Schizophrenia

Since the 1960s, it has been believed that many of the symptoms of schizophrenia were the result of an overactivity of dopamine. As with many of the psychological disorders, it is now understood that the biochemistry underlying schizophrenia is complex and involves multiple biological systems.

17.4.2.2.1 *The Dopamine Hypothesis*

Although the first **antipsychotic** drugs appeared in the 1950s, it was not well understood how these drugs were alleviating some of the symptoms of schizophrenia, nor why they also produced side effects, similar to those seen in patients with Parkinson's disease (muscular rigidity, tremors, decrease in voluntary movement). It wasn't until the late 1960s when further research made the notable discovery that the striatums of individuals with Parkinson's disease were completely depleted of dopamine (Goetz, 2011). This discovery led researchers to further speculate that antipsychotic drugs might be producing their rigidity and tremor side effects by reducing dopamine activity, which then implied that many of the psychotic symptoms alleviated by these drugs were being caused by the overactivity of dopamine in the brain. These speculations developed into what eventually became known as the dopamine theory of schizophrenia (see Brisch, et al., 2014).

As researchers came to better understand the biological complexities of schizophrenia, the original version of the dopamine hypothesis no longer fit with our understanding of the disorder. This has led to the "revised dopamine hypothesis" (Brisch, et al., 2014). This revised hypothesis proposes that the increased dopamine transmission in the mesocorticolimbic areas (see Figure 17.4.2.2) and decreased dopamine transmission in the prefrontal cortex seen in individuals with schizophrenia account for many of their symptoms. (Pogarell, et al. 2012).

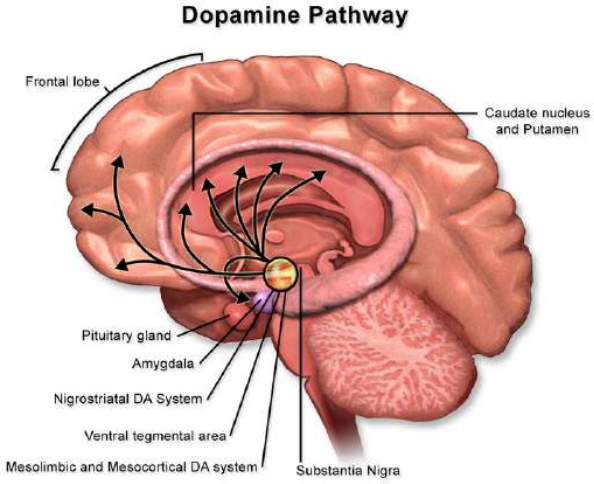


Figure **17.4.2**: **The mesocorticolimibic dopamine pathway. The revised dopamine hypothesis proposes that many of the symptoms of schizophrenia arise from dopamine dysfunction in this pathway.** (CC BY-SA 4.0; BruceBlaus via Wikimedia Commons)

In addition to the mesolimbic brain areas, dopamine dysregulation has also been observed in the amygdala and prefrontal cortex, which are important for emotional processing, and differences in dopamine contents in the prefrontal cortex, cingulate cortex, and hippocampus between schizophrenia patients and healthy control subjects (Patel, et al., 2010). In particular, the dopamine system in the hippocampus is overactive in schizophrenia patients (Grace, 2012).

17.4.2.3 Neural Structures in Schizophrenia

There is a long and reliable history of research into the neural basis of schizophrenia. Neuroimaging studies have found a significant reduction in overall and specific brain region volumes, as well as tissue density of individuals with schizophrenia compared to healthy controls (Brugger & Howes, 2017). Additionally, there has been evidence of ventricle enlargement as well as volume reductions in the medial temporal lobe, which contains such structures as the amygdala (involved in emotion regulation), the hippocampus (involved in memory), as well as the neocortical surface of the temporal lobes (processing of auditory information) (Kurtz, 2015). Additional studies also indicate a reduction in the orbitofrontal regions of the brain, a part of the frontal lobe that is responsible for response inhibition (Kurtz, 2015).

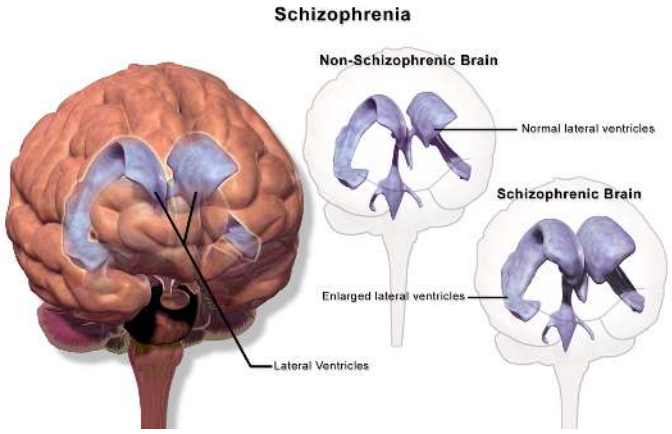


Figure 17.4.3: **The enlarged ventricles are often seen in the brains of individuals diagnosed with Schizophrenia.** (CC BY-SA 4.0; BruceBlaus via Wikimedia Commons)

17.4.3 References

Barch DM, Dowd EC. (2010). Goal representations and motivational drive in schizophrenia: the role of prefrontal-striatal interactions. *Schizophr Bull.*, *36*(5), 919-34. doi: 10.1093/schbul/sbq068.

Brisch, R., Saniotis, A., Wolf, R., Bielau, H., Bernstein, H. G., Steiner, J., Bogerts, B., Braun, K., Jankowski, Z., Kumaratilake, J., Henneberg, M., & Gos, T. (2014). The role of dopamine in schizophrenia from a neurobiological and evolutionary perspective: old fashioned, but still in vogue. *Frontiers in psychiatry*, *5*, 47. https://doi.org/10.3389/fpsyt.2014.00047.

Brugger SP, Howes OD. (2017). Heterogeneity and Homogeneity of Regional Brain Structure in Schizophrenia: A Meta-analysis. *JAMA Psychiatry*. *74*(11):1104-1111. doi:10.1001/jamapsychiatry.2017.2663. PMID: 28973084; PMCID: PMC5669456.

Cuthbert B.N. & Morris S.E. (2021). Evolving Concepts of the Schizophrenia Spectrum: A Research Domain Criteria Perspective. *Frontiers in Psychiatry*. *12*. DOI=10.3389/fpsyt.2021.641319

Flint J, Munafò MR. (2014) Genetics: finding genes for schizophrenia. *Curr Biol.*, 24(16), R755-7. doi: 10.1016/j.cub.2014.07.018.

Grace A. A. (2012). Dopamine system dysregulation by the hippocampus: implications for the pathophysiology and treatment of schizophrenia. *Neuropharmacology*, *62*(3), 1342–1348. https://doi.org/10.1016/j.neuropharm.2011.05.011.

Gürel C, Kuşçu GC, Yavaşoğlu, A, Avcı, CB. (2020). The clues in solving the mystery of major psychosis: The epigenetic basis of schizophrenia and bipolar disorder. *Neuroscience & Biobehavioral Reviews*. *113*. 51-61. https://doi.org/10.1016/j.neubiorev.2020.03.005.

Henriksen MG, Nordgaard J, Jansson LB. (2017). Genetics of Schizophrenia: Overview of Methods, Findings and Limitations. *Frontiers in Human Neuroscience*. *11*. DOI=10.3389/fnhum.2017.00322

Kurtz MM, Trask CL, Rosengard R, Hyman S, Kremen L, Mehta S, Olfson R, Rispaud S, Zaidman S, Choi J. (2017). Verbal Learning and Memory Enhancement Strategies in Schizophrenia: A Randomized, Controlled Investigation. *J Int Neuropsychol Soc.* *23*(4). 352-357. doi: 10.1017/S1355617717000042.

Llerena, K., Strauss, G. P., & Cohen, A. S. (2012). Looking at the other side of the coin: A meta-analysis of self-reported emotional arousal in people with schizophrenia. *Schizophrenia Research, 142*(1-3), 65–70. https://doi.org/10.1016/j.schres.2012.09.005.

Patel NH, Vyas NS, Puri BK, Nijran KS, Al-Nahhas A. (2010) Positron emission tomography in schizophrenia: a new perspective. *J Nucl Med., 51*(4). 511-20. doi: 10.2967/jnumed.109.066076.

Pogarell O, Koch W, Karch S, Dehning S, Müller N, Tatsch K, Poepperl G, Möller HJ. (2012). Dopaminergic neurotransmission in patients with schizophrenia in relation to positive and negative symptoms. *Pharmacopsychiatry. 45*(Suppl 1). 36-41. doi: 10.1055/s-0032-1306313.

Smigielski L, Jagannath V, Rössler W, Walitza S, Grünblatt E. (2020). Epigenetic mechanisms in schizophrenia and other psychotic disorders: a systematic review of empirical human findings. *Mol Psychiatry. 25*(8). 1718-1748. doi: 10.1038/s41380-019-0601-3. Epub 2020 Jan 6. PMID: 31907379.

Tandon R. (2013). Schizophrenia and other psychotic disorders in DSM-5. *Clin Schizophr Relat Psychoses*. *7*(1), 16-9. doi: 10.3371/CSRP.TA.032513. PMID: 23538289.

17.4.4 Attributions

Chapter 17, Biological Basis of Psychological Disorders, 17.4. Biological Bases of Schizophrenia Spectrum and other psychotic disorders. Original material written by Amy E. Coren, PhD, Pasadena City College, is licensed under CC BY-NC-SA 4.0

. Text of section 17.4.5 adapted from Alexis Bridley & Lee W. Daffin Jr., Essentials of Abnormal Psychology, 6.1 Schizophrenia Spectrum and Other Psychotic Disorders by Washington State University licensed under CC-BY-NC-SA 4.0 International License. Retrieved from: https://opentext.wsu.edu/abnormalpsy...g-information/

Figure 17.4.1 CC BY-SA 4.0, Amy E. Coren, Ph.D. via original creation

Figure 17.4.2. CC BY-SA 4.0, BruceBlaus via Wikimedia Commons

Figure 17.4.3. CC BY-SA 4.0; BruceBlaus via Wikimedia Commons

This page titled 17.4: Biological Basis of Schizophrenia Spectrum Disorder is shared under a mixed license and was authored, remixed, and/or curated by Open Education Resource Initative (ASCCC Open Educational Resources Initiative (OERI)) .

CHAPTER OVERVIEW

18: Supplemental Content

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

18.1: Chapter 3- Population Evolution (and the Modern Synthesis)
18.2: Chapter 3- Phenotypic Variation and Population Genetics
18.3: Chapter 3- Adaptive Evolution
18.4: Chapter 3- Modern Primates and Their Evolution
18.5: Chapter 3- Material Culture
18.6: Chapter 10- Computer Models and Learning in Artificial Neural Networks
18.7: Chapter 10- Acetylcholine, Enriched Experience, and Memory
18.8: Chapter 10- Memory Types and Stages
18.9: Chapter 10- The Psychology of Memory Encoding, Storage, and Retrieval
18.10: Chapter 10- Psychological Factors Affecting Learning, Memory, and Forgetting
18.11: Chapter 14- Traditional Models of Human Intelligence
18.12: Chapter 14- Problem Solving, Categories and Concepts
18.13: Chapter 14- The Brain and Evolution of Cognition
18.14: Chapter 14- Mirror Neurons, Theory of Mind, Social Cognition, and Neuroscience of Its Disorders
18.15: Chapter 15- A Gene Essential for Speech
18.16: Chapter 15- The Origins of Language

18: Supplemental Content is shared under a mixed license and was authored, remixed, and/or curated by Kenenth A. Koenigshofer.

18.1: Chapter 3- Population Evolution (and the Modern Synthesis)

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Define population genetics and describe how population genetics is used in the study of the evolution of populations.
2. Explain what is meant by "the modern synthesis."
3. Define phenotypic change and genotypic change.
4. Discuss the importance of allele frequency in modern conceptions of evolution.
5. Define genetic drift.
6. Explain Hardy-Weinberg equilibrium and the significance of deviations from it.

18.1.1 Overview

Natural selection is the most dominant of evolutionary forces. Natural selection acts to promote traits and behaviors that increase an organism's chances of survival and reproduction, while eliminating those traits and behaviors that are to the organism's detriment. But natural selection can only, as its name implies, select—it cannot create. The introduction of novel traits and behaviors falls on the shoulders of another evolutionary force—mutation. Mutation and other sources of variation among individuals, as well as the evolutionary forces that act upon them, alter populations and species. This combination of processes has led to the world of life we see today.

1. Introduction. All life on Earth is related. Evolutionary theory states that humans, beetles, plants, and bacteria all share a common ancestor, but that millions of years of evolution have shaped each of these organisms into the forms seen today. Scientists consider evolution a key concept to understanding life. Natural selection is one of the most dominant evolutionary forces.
2. Population Evolution. Initially, the newly discovered particulate nature of genes made it difficult for biologists to understand how gradual evolution could occur. But over the next few decades, genetics and evolution were integrated in what became known as the modern synthesis—the coherent understanding of the relationship between natural selection and genetics that took shape by the 1940s and is generally accepted today.
3. Population Genetics. Individuals of a population often display different phenotypes, or express different alleles of a particular gene, referred to as polymorphisms. Populations with two or more variations of particular characteristics are called polymorphic. The distribution of phenotypes among individuals, known as the population variation, is influenced by a number of factors, including the population's genetic structure and the environment.
4. Adaptive Evolution. Fitness is often quantifiable and is measured by scientists in the field. However, it is not the absolute fitness of an individual that counts, but rather how it compares to the other organisms in the population. This concept, called relative fitness, allows researchers to determine which individuals are contributing additional offspring to the next generation, and thus, how the population might evolve.

18.1.2 Introduction



Figure **18.1.1**: Living things may be single-celled or complex, multicellular organisms. They may be plants, animals, fungi, bacteria, or archaea. This diversity results from evolution. (credit "wolf": modification of work by Gary Kramer; credit "coral": modification of work by William Harrigan, NOAA; credit "river": modification of work by Vojtěch Dostál; credit "fish" modification of work by Christian Mehlführer; credit "mushroom": modification of work by Cory Zanker; credit "tree": modification of work by Joseph Kranak; credit "bee": modification of work by Cory Zanker)

All life on Earth is related. Evolutionary theory states that humans, beetles, plants, and bacteria all share a common ancestor, but that millions of years of evolution have shaped each of these organisms into the forms seen today. Scientists consider evolution a key concept to understanding life. Natural selection is one of the most dominant evolutionary forces. Natural selection acts to promote traits and behaviors that increase an organism's chances of survival and reproduction, while eliminating those traits and behaviors that are to the organism's detriment. But natural selection can only, as its name implies, select—it cannot create. The introduction of novel traits and behaviors falls on the shoulders of another evolutionary force—mutation. Mutation and other sources of variation among individuals, as well as the evolutionary forces that act upon them, alter populations and species. This combination of processes has led to the world of life we see today--the physical and psychological traits of each species.

18.1.3 Population Evolution

The mechanisms of inheritance, or genetics, were not understood at the time Charles Darwin and Alfred Russel Wallace were developing their idea of natural selection. This lack of understanding was a stumbling block to understanding many aspects of evolution. In fact, the predominant (and incorrect) genetic theory of the time, blending inheritance, made it difficult to understand how natural selection might operate. Darwin and Wallace were unaware of the genetics work by Austrian monk Gregor Mendel, which was published in 1866, not long after publication of Darwin's book, *On the Origin of Species* (1859). Mendel's work was rediscovered in the early twentieth century at which time geneticists were rapidly coming to an understanding of the basics of inheritance. Initially, the newly discovered particulate nature of genes made it difficult for biologists to understand how gradual evolution could occur. But over the next few decades, genetics and evolution were integrated in what became known as **the modern synthesis**—the coherent understanding of the relationship between natural selection and genetics that took shape by the 1940s and is generally accepted today. In sum, the modern synthesis describes how evolutionary processes, such as natural selection, can affect a population's genetic makeup, and, in turn, how this can result in the gradual evolution of populations and species. The theory also connects this change of a population over time, called microevolution, with the processes that gave rise to new species and higher taxonomic groups with widely divergent characters, called macroevolution.

Everyday Connection: Evolution and Flu Vaccines

Every fall, the media starts reporting on flu vaccinations and potential outbreaks. Scientists, health experts, and institutions determine recommendations for different parts of the population, predict optimal production and inoculation schedules, create vaccines, and set up clinics to provide inoculations. You may think of the annual flu shot as a lot of media hype, an important health protection, or just a briefly uncomfortable prick in your arm. But do you think of it in terms of evolution?

The media hype of annual flu shots is scientifically grounded in our understanding of evolution. Each year, scientists across the globe strive to predict the flu strains that they anticipate being most widespread and harmful in the coming year. This knowledge is based in how flu strains have evolved over time and over the past few flu seasons. Scientists then work to create the most effective vaccine to combat those selected strains. Hundreds of millions of doses are produced in a short period in order to provide vaccinations to key populations at the optimal time.

Because viruses, like the flu, evolve very quickly (especially in evolutionary time), this poses quite a challenge. Viruses mutate and replicate at a fast rate, so the vaccine developed to protect against last year's flu strain may not provide the protection needed against the coming year's strain. Evolution of these viruses means continued adaptions to ensure survival, including adaptations to survive previous vaccines.

18.1.4 Population Genetics

Recall that a gene for a particular character may have several alleles, or variants, that code for different traits associated with that character. For example, in the ABO blood type system in humans, three alleles determine the particular blood-type protein on the surface of red blood cells. Each individual in a population of diploid organisms (organisms that reproduce sexually and have paired chromosomes, unlike haploid organisms that reproduce asexually and have a single set of chromosomes) can only carry two alleles for a particular gene, but more than two may be present in the individuals that make up the population. Mendel followed alleles as they were inherited from parent to offspring. In the early twentieth century, biologists in a field of study known as population genetics began to study how selective forces change a population through changes in allele and genotypic frequencies.

The allele frequency (or gene frequency) is the rate at which a specific allele appears within a population. Until now, we have discussed evolution as a change in the characteristics of a population of organisms, but behind that phenotypic change is genetic change. In population genetics, the term evolution is defined as a change in the frequency of an allele in a population. Using the ABO blood type system as an example, the frequency of one of the alleles, A, is the number of copies of that allele divided by all the copies of the ABO gene in the population. For example, a study by Sahar et al. (2007) found a frequency of A to be 26.1 percent. The B and 0 alleles made up 13.4 percent and 60.5 percent of the alleles respectively, and all of the frequencies added up to 100 percent. A change in this frequency over time would constitute evolution in the population.

The allele frequency within a given population can change depending on environmental factors; therefore, certain alleles become more widespread than others during the process of natural selection. Natural selection can alter the population's genetic makeup; for example, if a given allele confers a phenotype that allows an individual to better survive or have more offspring. Because many of those offspring will also carry the beneficial allele, and often the corresponding phenotype, they will have more offspring of their own that also carry the allele, thus, perpetuating the cycle. Over time, the allele will spread throughout the population. Some alleles will quickly become fixed in this way, meaning that every individual of the population will carry the allele, while detrimental mutations may be swiftly eliminated if derived from a dominant allele from the gene pool. The gene pool is the sum of all the alleles in a population.

Sometimes, allele frequencies within a population change randomly with no advantage to the population over existing allele frequencies. This phenomenon is called **genetic drift**. Natural selection and genetic drift usually occur simultaneously in populations and are not isolated events. It is hard to determine which process dominates because it is often nearly impossible to determine the cause of change in allele frequencies at each occurrence. An event that initiates an allele frequency change in an isolated part of the population, which is not typical of the original population, is called the **founder effect**. Natural selection, random genetic drift, and founder effects can lead to significant changes in the genome of a population.

18.1.5 Hardy-Weinberg Principle of Equilibrium

In the early twentieth century, English mathematician Godfrey Hardy and German physician Wilhelm Weinberg stated the principle of equilibrium to describe the genetic makeup of a population. The theory, which later became known as the Hardy-Weinberg principle of equilibrium, states that a population's allele and genotype frequencies are inherently stable—neither the allele nor the genotypic frequencies would change, unless some kind of evolutionary force is acting upon the population. The Hardy-Weinberg principle assumes conditions with no mutations, migration, emigration, or selective pressure for or against any genotype, plus an infinite population; while no population can satisfy those conditions, the principle offers a useful model against which to compare real population changes.

In theory, if a population is at equilibrium—that is, there are no evolutionary forces acting upon it—generation after generation would have the same gene pool and genetic structure, and the equilibrium equations of Hardy-Weinberg would all hold true all of the time. Of course, even Hardy and Weinberg recognized that no natural population is immune to evolution. Populations in nature are constantly changing in genetic makeup due to drift, mutation, possibly migration, natural selection, and in social species, kin selection. As a result, the only way to determine the exact distribution of phenotypes in a population is to go out and count them. But the Hardy-Weinberg principle gives scientists a mathematical baseline of a non-evolving population to which they can compare evolving populations and thereby infer what evolutionary forces might be at play. If the frequencies of alleles or genotypes deviate from the value expected from the Hardy-Weinberg equilibrium equation, then the population is evolving.

18.1.6 Additional Concepts in Evolutionary Science

18.1.6.1 Species

Although all life on earth shares various genetic similarities, only certain organisms combine genetic information by sexual reproduction and have offspring that can then successfully reproduce. Scientists call such organisms members of the same biological species.

A **species** is a group of individual organisms that interbreed and produce fertile, viable offspring. According to this definition, one species is distinguished from another when, in nature, it is not possible for matings between individuals from each species to produce fertile offspring. The biological definition of species, which works for sexually reproducing organisms, is a group of actually or potentially interbreeding individuals.

Members of the same species share both external and internal characteristics, which develop from their DNA. The closer relationship two organisms share, the more DNA they have in common, just like people and their families. Organisms of the same species have the highest level of DNA alignment and therefore share characteristics and behaviors that lead to successful reproduction.

Populations of species share a **gene pool**: a collection of all the variants of genes in the species. Remember that any evolutionary changes in a population of organisms must be genetic because genes are the only way to share and pass on heritable traits. Only heritable traits can evolve. Therefore, reproduction plays a paramount role for genetic change to take root in a population or species. Differential rates of reproduction among the genetic variants in a population drive evolutionary change.

18.1.6.2 Speciation

Given the extraordinary diversity of life on the planet, there must be mechanisms for speciation: the formation of two species from one original species. Darwin envisioned this process as a branching event.

Biologists think of speciation events as the splitting of one ancestral species into two descendant species.

For speciation to occur, two new populations must be formed from one original population and they must evolve in such a way that it becomes impossible for individuals from the two new populations to interbreed. Biologists have proposed mechanisms by which this could occur that fall into two broad categories. Allopatric speciation (allo- = "other"; -patric = "homeland") involves geographic separation of populations from a parent species and subsequent evolution. Sympatric speciation (sym- = "same"; -patric = "homeland") involves speciation occurring within a parent species remaining in one location. Isolation of populations leading to allopatric speciation can occur in a variety of ways: a river forming a new branch, erosion forming a new valley, a group of organisms traveling to a new location without the ability to return, or seeds floating over the ocean to an island. Additionally,

scientists have found that the further the distance between two groups that once were the same species, the more likely it is that speciation will occur.

In some cases, a population of one species disperses throughout an area, and each finds a distinct niche or isolated habitat. Over time, the varied demands of their new lifestyles lead to multiple speciation events originating from a single species. This is called adaptive radiation because many adaptations evolve from a single point of origin; thus, causing the species to radiate into several new ones. Island archipelagos like the Hawaiian Islands provide an ideal context for adaptive radiation events because water surrounds each island which leads to geographical isolation for many organisms.

18.1.6.3 Rates of Speciation

Speciation occurs over a span of evolutionary time, so when a new species arises, there is a transition period during which the closely related species continue to interact.

In terms of how quickly speciation occurs, two patterns are currently observed: gradual speciation model and punctuated equilibrium model.

In the gradual speciation model, species diverge gradually over time in small steps. In the punctuated equilibrium model, a new species undergoes changes quickly from the parent species, and then remains largely unchanged for long periods of time afterward (Figure 18.3.218.3.2). This early change model is called punctuated equilibrium, because it begins with a punctuated or periodic change and then remains in balance afterward. While punctuated equilibrium suggests a faster tempo, it does not necessarily exclude gradualism.

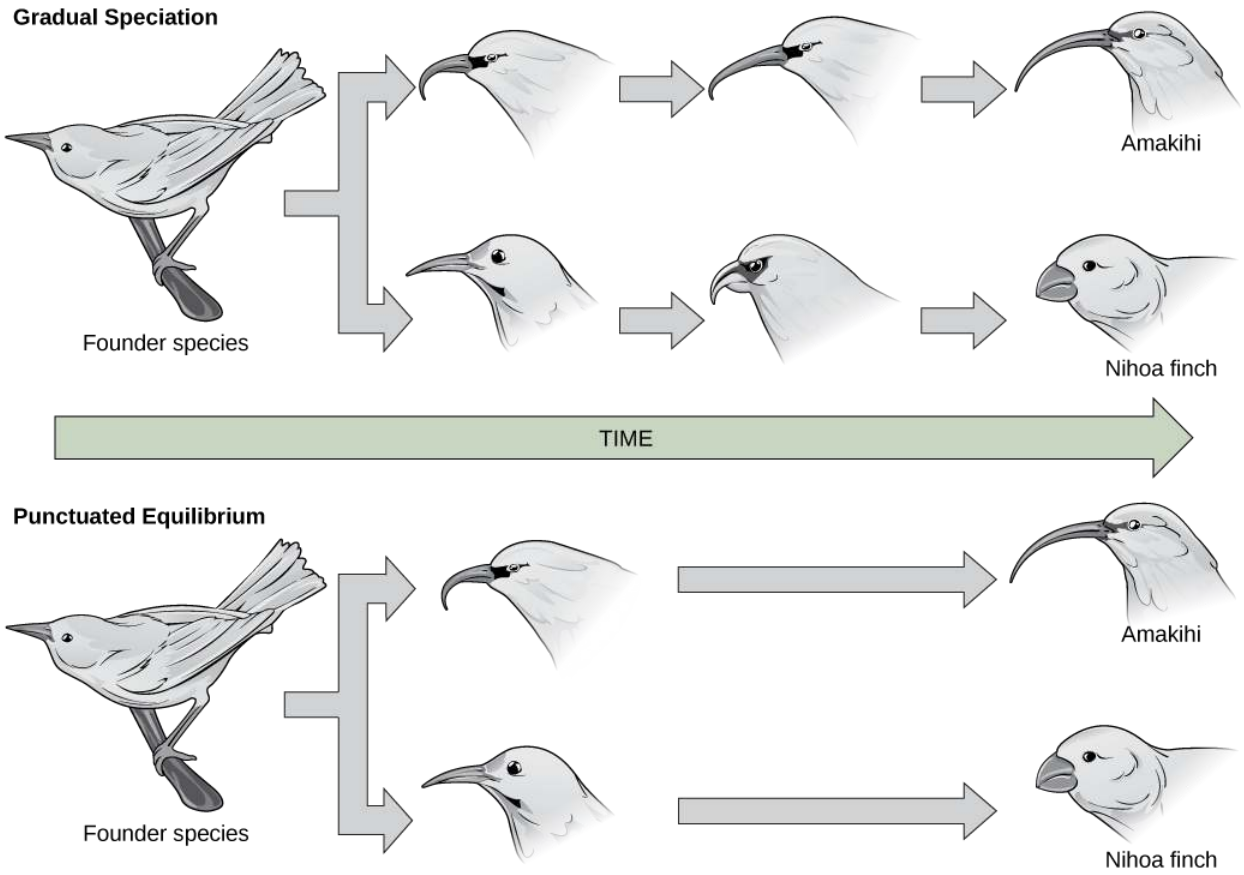


Figure 18.1.2: In (a) gradual speciation, species diverge at a slow, steady pace as traits change incrementally. In (b) punctuated equilibrium, species diverge quickly and then remain unchanged for long periods of time.

The primary influencing factor on changes in speciation rate is environmental conditions. Under some conditions, selection occurs quickly or radically. Consider a species of snails that had been living with the same basic form for many thousands of years. Layers of their fossils would appear similar for a long time. When a change in the environment takes place—such as a drop in the water

level—a small number of organisms are separated from the rest in a brief period of time, essentially forming one large and one tiny population. The tiny population faces new environmental conditions. Because its gene pool quickly became so small, any variation that surfaces and that aids in surviving the new conditions becomes the predominant form.

Which of the following statements is false? (Answer at end of this module).

1. Punctuated equilibrium is most likely to occur in a small population that experiences a rapid change in its environment.
2. Punctuated equilibrium is most likely to occur in a large population that lives in a stable climate.
3. Gradual speciation is most likely to occur in species that live in a stable climate.
4. Gradual speciation and punctuated equilibrium both result in the divergence of species.

18.1.7 Summary

The modern synthesis of evolutionary theory grew out of the combination of Darwin's and Wallace's formulations of evolution with Mendel's analysis of heredity, along with the more modern study of population genetics. The modern synthesis describes the evolution of populations and species, from small-scale changes among individuals to large-scale changes over paleontological time periods. To understand how organisms evolve, scientists can track populations' allele frequencies over time. If they differ from generation to generation, scientists can conclude that the population is not in Hardy-Weinberg equilibrium, and is thus evolving.

Speciation is not a precise division: overlap between closely related species can occur in areas called hybrid zones. Organisms reproduce with other similar organisms. The fitness of these hybrid offspring can affect the evolutionary path of the two species. Scientists propose two models for the rate of speciation: one model illustrates how a species can change slowly over time; the other model demonstrates how change can occur quickly from a parent generation to a new species. Both models continue to follow the patterns of natural selection.

(Answer to review question above: 2 is false. "Punctuated equilibrium is most likely to occur in a large population that lives in a stable climate").

18.1.8 Art Connections

> [link] In plants, violet flower color (V) is dominant over white (v). If p=.8 and q = 0.2 in a population of 500 plants, how many individuals would you expect to be homozygous dominant (VV), heterozygous (Vv), and homozygous recessive (vv)? How many plants would you expect to have violet flowers, and how many would have white flowers?

18.1.9 References

Sahar S. Hanania, Dhia S. Hassawi, and Nidal M. Irshaid, "Allele Frequency and Molecular Genotypes of ABO Blood Group System in a Jordanian Population," *Journal of Medical Sciences* 7 (2007): 51-58, doi:10.3923/jms.2007.51.58.

18.1.10 Glossary

allele frequency

founder effect

gene pool

genetic structure

macroevolution

microevolution

modern synthesis

population genetics

18.1.11 Attributions

Adapted by Kenneth A. Koenigshofer, Ph.D., from The Evolution of Populations by OpenStax, licensed CC BY 4.0.

This page titled 18.1: Chapter 3- Population Evolution (and the Modern Synthesis) is shared under a mixed license and was authored, remixed, and/or curated by Keneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

- **19.1: Population Evolution** by OpenStax is licensed CC BY 4.0.

18.2: Chapter 3- Phenotypic Variation and Population Genetics

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe the different types of variation in a population.
2. Explain why only heritable variation can be acted upon by natural selection.
3. Describe genetic drift and the bottleneck effect.
4. Explain how each evolutionary force can influence the allele frequencies of a population.

18.2.1 Overview

Individuals of a population often display different phenotypes, or express different alleles of a particular gene, referred to as polymorphisms. Populations with two or more variations of particular characteristics are called polymorphic. The distribution of phenotypes among individuals, known as the population variation, is influenced by a number of factors, including the population's genetic structure and the environment. Understanding the sources of a phenotypic variation in a population is important for determining how a population will evolve in response to different evolutionary pressures.



Figure **18.2.1**: The distribution of phenotypes in this litter of kittens illustrates population variation. (credit: Pieter Lanser)

18.2.2 Genetic Variance

Natural selection and some of the other evolutionary forces can only act on heritable traits, namely an organism's genetic code. Because alleles are passed from parent to offspring, those that confer beneficial traits or behaviors may be selected for, while deleterious alleles may be selected against. Acquired traits, for the most part, are not heritable. For example, if an athlete works out in the gym every day, building up muscle strength, the athlete's offspring will not necessarily grow up to be a body builder. If there is a genetic basis for the ability to run fast, on the other hand, this may be passed to a child.

Heritability is the fraction of phenotype variation that can be attributed to genetic differences, or genetic variance, among individuals in a population. The greater the hereditability of a population's phenotypic variation, the more susceptible it is to the evolutionary forces that act on heritable variation.

The diversity of alleles and genotypes within a population is called **genetic variance**. When scientists are involved in the breeding of a species, such as with animals in zoos and nature preserves, they try to increase a population's genetic variance to preserve as much of the phenotypic diversity as they can. This also helps reduce the risks associated with inbreeding, the mating of closely

related individuals, which can have the undesirable effect of bringing together deleterious recessive mutations that can cause abnormalities and susceptibility to disease.

Changes in allele frequencies that are identified in a population can shed light on how it is evolving. In addition to natural selection, there are other evolutionary forces that could be in play: genetic drift, gene flow, mutation, nonrandom mating, and environmental variances.

18.2.3 Genetic Drift

The theory of natural selection stems from the observation that some individuals in a population are more likely to survive longer and have more offspring than others; thus, they will pass on more of their genes to the next generation. A big, powerful male gorilla, for example, is much more likely than a smaller, weaker one to become the population's silverback, the pack's leader who mates far more than the other males of the group. The pack leader will father more offspring, who share half of his genes, and are likely to also grow bigger and stronger like their father. Over time, the genes for bigger size will increase in frequency in the population, and the population will, as a result, grow larger on average. That is, this would occur if this particular selection pressure, or driving selective force, were the only one acting on the population. In other examples, better camouflage or a stronger resistance to drought might provide a selection pressure.

Another way a population's allele and genotype frequencies can change is **genetic drift**, which is simply the effect of chance. By chance, some individuals will have more offspring than others—not due to an advantage conferred by some genetically-encoded trait, but just because one male happened to be in the right place at the right time (when the receptive female walked by) or because others happened to be in the wrong place at the wrong time (when a predator was hunting).

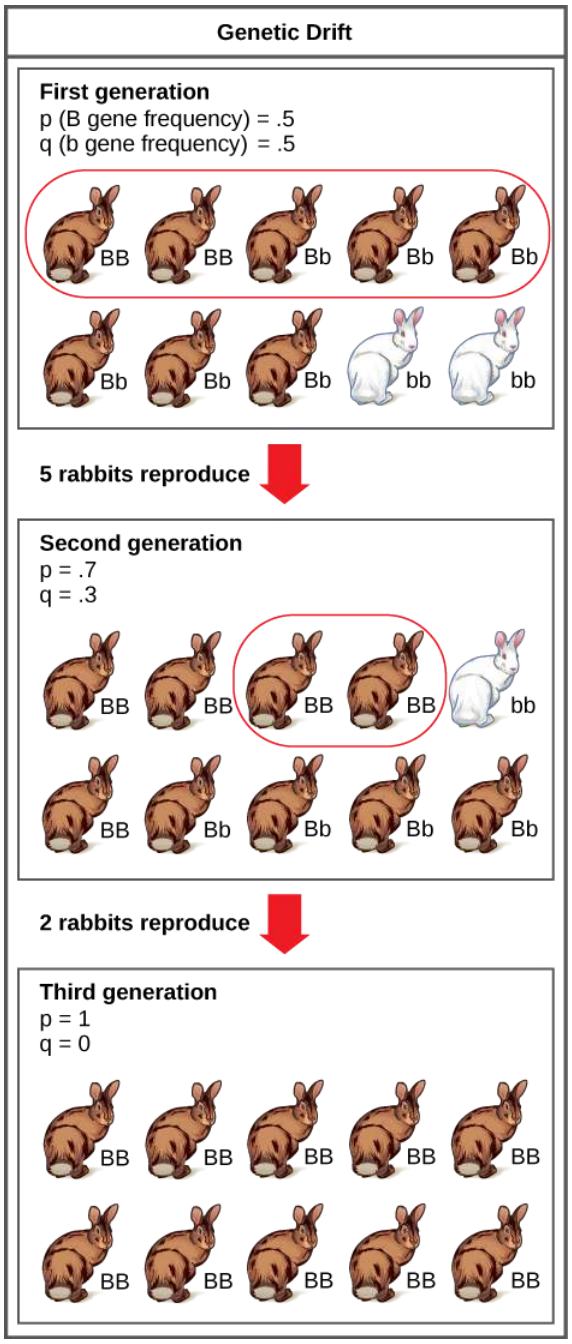


Figure **18.2.2**: Genetic drift in a population can lead to the elimination of an allele from a population by chance. In this example, rabbits with the brown coat color allele (B) are dominant over rabbits with the white coat color allele (b). In the first generation, the two alleles occur with equal frequency in the population, resulting in p and q values of 0.5. Only half of the individuals reproduce, resulting in a second generation with p and q values of .7 and .3, respectively. Only two individuals in the second generation reproduce, and by chance these individuals are homozygous dominant for brown coat color. As a result, in the third generation the recessive b allele is lost.

Do you think genetic drift would happen more quickly on an island or on the mainland?

Small populations are more susceptible to the forces of genetic drift. Large populations, on the other hand, are buffered against the effects of chance. If one individual of a population of 10 individuals happens to die at a young age before it leaves any offspring to the next generation, all of its genes—1/10 of the population's gene pool—will be suddenly lost. In a population of 100, that's only 1 percent of the overall gene pool; therefore, it is much less impactful on the population's genetic structure.

Genetic drift can also be magnified by natural events, such as a natural disaster that kills—at random—a large portion of the population. Known as the bottleneck effect, it results in a large portion of the genome suddenly being wiped out. In one fell swoop, the genetic structure of the survivors becomes the genetic structure of the entire population, which may be very different from the pre-disaster population.

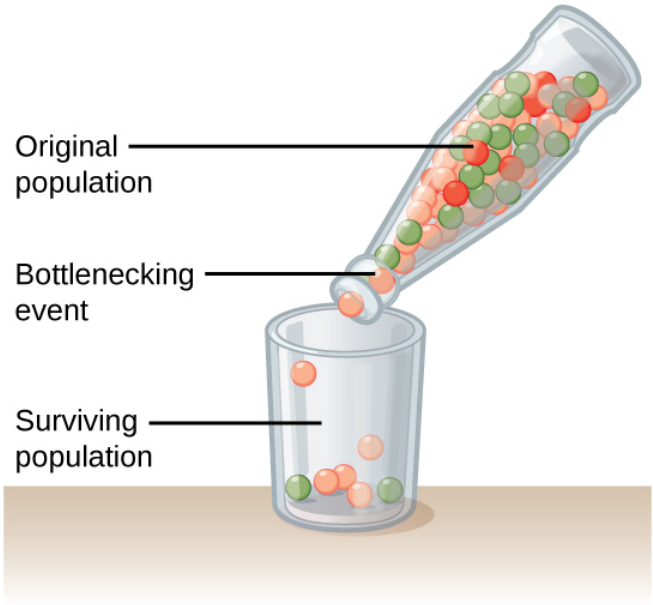


Figure **18.2.3**: A chance event or catastrophe can reduce the genetic variability within a population.

Another scenario in which populations might experience a strong influence of genetic drift is if some portion of the population leaves to start a new population in a new location or if a population gets divided by a physical barrier of some kind. In this situation, those individuals are unlikely to be representative of the entire population, which results in the founder effect. The founder effect occurs when the genetic structure changes to match that of the new population's founding fathers and mothers. The founder effect is believed to have been a key factor in the genetic history of the Afrikaner population of Dutch settlers in South Africa, as evidenced by mutations that are common in Afrikaners but rare in most other populations. This is likely due to the fact that a higher-than-normal proportion of the founding colonists carried these mutations. As a result, the population expresses unusually high incidences of Huntington's disease (HD) and Fanconi anemia (FA), a genetic disorder known to cause blood marrow and congenital abnormalities—even cancer.

Scientific Method Connection: Testing the Bottleneck Effect

- **Question**: How do natural disasters affect the genetic structure of a population?
- **Background**: When much of a population is suddenly wiped out by an earthquake or hurricane, the individuals that survive the event are usually a random sampling of the original group. As a result, the genetic makeup of the population can change dramatically. This phenomenon is known as the bottleneck effect.
- **Hypothesis**: Repeated natural disasters will yield different population genetic structures; therefore, each time this experiment is run, the results will vary.
- **Test the hypothesis**: Count out the original population using different colored beads. For example, red, blue, and yellow beads might represent red, blue, and yellow individuals. After recording the number of each individual in the original population, place them all in a bottle with a narrow neck that will only allow a few beads out at a time. Then, pour 1/3 of the bottle's contents into a bowl. This represents the surviving individuals after a natural disaster kills a majority of the population. Count the number of the different colored beads in the bowl, and record it. Then, place all of the beads back in the bottle and repeat the experiment four more times.
- **Analyze the data**: Compare the five populations that resulted from the experiment. Do the populations all contain the same number of different colored beads, or do they vary? Remember, these populations all came from the same exact parent population.

- **Form a conclusion**: Most likely, the five resulting populations will differ quite dramatically. This is because natural disasters are not selective—they kill and spare individuals at random. Now think about how this might affect a real population. What happens when a hurricane hits the Mississippi Gulf Coast? How do the seabirds that live on the beach fare?

18.2.4 Gene Flow

Another important evolutionary force is **gene flow**: the flow of alleles in and out of a population due to the migration of individuals or gametes. While some populations are fairly stable, others experience more flux. Many plants, for example, send their pollen far and wide, by wind or by bird, to pollinate other populations of the same species some distance away. Even a population that may initially appear to be stable, such as a pride of lions, can experience its fair share of immigration and emigration as developing males leave their mothers to seek out a new pride with genetically unrelated females. This variable flow of individuals in and out of the group not only changes the gene structure of the population, but it can also introduce new genetic variation to populations in different geological locations and habitats.

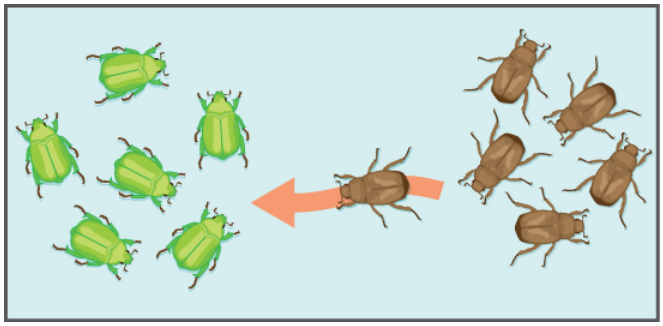


Figure 18.2.4: Gene flow can occur when an individual travels from one geographic location to another.

18.2.5 Mutation

Mutations are changes to an organism's DNA and are an important driver of diversity in populations. Species evolve because of the accumulation of mutations that occur over time. The appearance of new mutations is the most common way to introduce novel genotypic and phenotypic variance. Some mutations are unfavorable or harmful and are quickly eliminated from the population by natural selection. Others are beneficial and will spread through the population. Whether or not a mutation is beneficial or harmful is determined by whether it helps an organism survive to sexual maturity and reproduce. Some mutations do not do anything and can linger, unaffected by natural selection, in the genome. Some can have a dramatic effect on a gene and the resulting phenotype.

18.2.6 Nonrandom Mating

If individuals nonrandomly mate with their peers, the result can be a changing population. There are many reasons nonrandom mating occurs. One reason is simple mate choice (sexual selection); for example, female peahens may prefer peacocks with bigger, brighter tails. Sexual attractiveness, hypothesized by some evolutionary psychologists to be a phenotypic indicator of health and reproductive potential, increases mating opportunities. Traits that lead to more matings for an individual become selected for by natural selection. One common form of mate choice, called assortative mating, is an individual's preference to mate with partners who are phenotypically similar to themselves. On the other hand, you may have heard of the saying, opposites attract.

Humans frequently engage in behaviors to enhance their attractiveness to potential mates. For example, women in many modern cultures shape their eyebrows making them thinner and more arched above the eyes. Why this works to increase attractiveness is a matter for speculation, but one possibility is that traits that exaggerate differences between the sexes can enhance attractiveness. In this case, thinner arched eyebrows contrast more sharply with the human male "brow ridge," a ridge of bone above the eye sockets in human adult males that gives their brows a heavier and lower profile on the human male face. When women thin out and arch their eyebrows they are exaggerating the difference between female facial bone structure and male facial bone structure. This enhances a sexual cue on the human face which identifies it as female. Evolution may have selected for male brain circuitry that stimulates sexual feeling in males more readily in the presence of a face with this identifying feature of femaleness (Koenigshofer, 2011). In addition, there are many examples of humans using various adornments on the body and face and other strategies to

enhance their attractiveness (think of makeup and jewelry; men working out to get more muscular or buying expensive cars to signal high status). Human mate choice is the result of complex interacting factors, but it is clearly nonrandom.

Another cause of nonrandom mating is physical location. This is especially true in large populations spread over large geographic distances where not all individuals will have equal access to one another. Some might be miles apart through woods or over rough terrain, while others might live immediately nearby.

18.2.7 Environmental Variance

Genes are not the only players involved in determining population variation. Phenotypes are also influenced by other factors, such as the environment. A beachgoer is likely to have darker skin than a city dweller, for example, due to regular exposure to the sun, an environmental factor. Some major characteristics, such as gender, are determined by the environment for some species. For example, some turtles and other reptiles have temperature-dependent sex determination (TSD). TSD means that individuals develop into males if their eggs are incubated within a certain temperature range, or females at a different temperature range.



Figure **18.2.5**: The sex of the American alligator (Alligator mississippiensis) is determined by the temperature at which the eggs are incubated. Eggs incubated at 30°C produce females, and eggs incubated at 33°C produce males. (credit: Steve Hillebrand, USFWS)

Geographic separation between populations can lead to differences in the phenotypic variation between those populations. Such geographical variation is seen between most populations and can be significant. One type of geographic variation, called a cline, can be seen as populations of a given species vary gradually across an ecological gradient. Species of warm-blooded animals, for example, tend to have larger bodies in the cooler climates closer to the earth's poles, allowing them to better conserve heat. This is considered a latitudinal cline. Alternatively, flowering plants tend to bloom at different times depending on where they are along the slope of a mountain, known as an altitudinal cline.

If there is gene flow between the populations, the individuals will likely show gradual differences in phenotype along the cline. Restricted gene flow, on the other hand, can lead to abrupt differences, even speciation.

18.2.8 Summary

Both genetic and environmental factors can cause phenotypic variation in a population. Different alleles can confer different phenotypes, and different environments can also cause individuals to look or act differently. Only those differences encoded in an individual's genes, however, can be passed to its offspring genetically and, thus, be a target of natural selection. Natural selection works by selecting for alleles that confer beneficial traits or behaviors, while selecting against those for deleterious qualities. Genetic drift stems from the chance occurrence that some individuals in the germ line have more offspring than others. When individuals leave or join the population, allele frequencies can change as a result of gene flow. Mutations to an individual's DNA may introduce new variation into a population. Allele frequencies can also be altered when individuals do not randomly mate with others in the group (sexual selection, some mates are more desirable than others and therefore mate more frequently and consequently tend to leave more offspring).

18.2.9 Art Connections

[link] Do you think genetic drift would happen more quickly on an island or on the mainland?

[link] Genetic drift is likely to occur more rapidly on an island where smaller populations are expected to occur.

18.2.10 Footnotes

1. 1 A. J. Tipping et al., “Molecular and Genealogical Evidence for a Founder Effect in Fanconi Anemia Families of the Afrikaner Population of South Africa,” *PNAS* 98, no. 10 (2001): 5734-5739, doi: 10.1073/pnas.091402398.

18.2.11 Attributions

Adapted by Kenneth A. Koenigshofer, PhD. (with modifications to section titled "Nonrandom Mating") from Population Genetics by OpenStax, licensed CC BY 4.0

This page titled 18.2: Chapter 3- Phenotypic Variation and Population Genetics is shared under a mixed license and was authored, remixed, and/or curated by Keneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

- **19.2: Population Genetics** by OpenStax is licensed CC BY 4.0.

18.3: Chapter 3- Adaptive Evolution

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Explain the different ways natural selection can shape populations
2. Describe how these different forces can lead to different outcomes in terms of the population variation

18.3.1 Overview

In this module we examine the ways in which natural selection, and its action under different conditions, leads to adaptive evolution under a variety of conditions. Different types of selection include stabilizing, directional, diversifying, and frequency-dependent selection. Sexual selection, resulting from competition for mating opportunities among individuals in sexually dimorphic species, can lead to evolution of traits which increase chances for mating but actually reduce chances of survival (e.g. male peacock's tail feathers). Evolutionary fitness refers to the genetic contribution that individuals make to the next generation (i.e. reproductive success). Natural selection does not create a perfect organism. It can generate populations of organism that are better adapted to survive and reproduce in their environments. Natural selection can only select on existing variation in the population; it does not create anything from scratch. Thus, it is limited by a population's existing genetic variance and whatever new alleles arise through mutation and gene flow.

18.3.2 Natural Selection and Adaptive Evolution

Natural selection drives adaptive evolution by selecting for and increasing the occurrence of beneficial traits (traits which increase survival and reproduction) in a population and by selecting against non-beneficial traits.

18.3.2.1 Key Points

- Natural selection increases or decreases biological traits within a population, thereby selecting for individuals with greater evolutionary fitness.
- An individual with a high evolutionary fitness will provide more beneficial contributions to the gene pool of the next generation.
- Relative fitness, which compares an organism's fitness to the others in the population, allows researchers to establish how a population may evolve by determining which individuals are contributing additional offspring to the next generation.
- Stabilizing selection, directional selection, diversifying selection, frequency-dependent selection, and sexual selection all contribute to the way natural selection can affect variation within a population.

18.3.2.2 Key Terms

- **natural selection**: a process in which individual organisms or phenotypes that possess favorable traits are more likely to survive and reproduce
- **fecundity**: number, rate, or capacity of offspring production
- **Darwinian fitness**: the average contribution to the gene pool of the next generation that is made by an average individual of the specified genotype or phenotype

18.3.2.3 Introduction: Natural Selection and Adaptive Evolution

Natural selection only acts on the population's heritable traits: selecting for beneficial alleles and thus increasing their frequency in the population, while selecting against deleterious alleles and thereby decreasing their frequency—a process known as **adaptive evolution**. Natural selection does not act on individual alleles, however, but on entire organisms. An individual may carry a very beneficial genotype with a resulting phenotype that, for example, increases the ability to reproduce (fecundity), but if that same individual also carries an allele that results in a fatal childhood disease, that fecundity phenotype will not be passed on to the next generation because the individual will not live to reach reproductive age. Natural selection acts at the level of the individual; it

selects for individuals with greater contributions to the gene pool of the next generation, known as an organism's evolutionary (Darwinian) fitness.

Fitness is often quantifiable and is measured by scientists in the field. However, it is not the absolute fitness of an individual that counts, but rather how it compares to the other organisms in the population. This concept, called relative fitness, allows researchers to determine which individuals are contributing additional offspring to the next generation, and thus, how the population might evolve.

There are several ways selection can affect population variation: stabilizing selection, directional selection, diversifying selection, frequency-dependent selection, and sexual selection. As natural selection influences the allele frequencies in a population, individuals can either become more or less genetically similar and the phenotypes displayed can become more similar or more disparate.

18.3.3 Stabilizing Selection

If natural selection favors an average phenotype, selecting against extreme variation, the population will undergo stabilizing selection. In a population of mice that live in the woods, for example, natural selection is likely to favor individuals that best blend in with the forest floor and are less likely to be spotted by predators. Assuming the ground is a fairly consistent shade of brown, those mice whose fur is most closely matched to that color will be most likely to survive and reproduce, passing on their genes for their brown coat. Mice that carry alleles that make them a bit lighter or a bit darker will stand out against the ground and be more likely to fall victim to predation. As a result of this selection, the population's genetic variance will decrease.

18.3.4 Directional Selection

When the environment changes, populations will often undergo directional selection, which selects for phenotypes at one end of the spectrum of existing variation. A classic example of this type of selection is the evolution of the peppered moth in eighteenth- and nineteenth-century England. Prior to the Industrial Revolution, the moths were predominately light in color, which allowed them to blend in with the light-colored trees and lichens in their environment. But as soot began spewing from factories, the trees became darkened, and the light-colored moths became easier for predatory birds to spot. Over time, the frequency of the melanic form of the moth increased because they had a higher survival rate in habitats affected by air pollution because their darker coloration blended with the sooty trees. Similarly, the hypothetical mouse population may evolve to take on a different coloration if something were to cause the forest floor where they live to change color. The result of this type of selection is a shift in the population's genetic variance toward the new, fit phenotype.

18.3.5 Diversifying Selection

Sometimes two or more distinct phenotypes can each have their advantages and be selected for by natural selection, while the intermediate phenotypes are, on average, less fit. Known as diversifying selection, this is seen in many populations of animals that have multiple male forms. Large, dominant alpha males obtain mates by brute force, while small males can sneak in for furtive copulations with the females in an alpha male's territory. In this case, both the alpha males and the "sneaking" males will be selected for, but medium-sized males, which can't overtake the alpha males and are too big to sneak copulations, are selected against. Diversifying selection can also occur when environmental changes favor individuals on either end of the phenotypic spectrum. Imagine a population of mice living at the beach where there is light-colored sand interspersed with patches of tall grass. In this scenario, light-colored mice that blend in with the sand would be favored, as well as dark-colored mice that can hide in the grass. Medium-colored mice, on the other hand, would not blend in with either the grass or the sand, and would thus be more likely to be eaten by predators. The result of this type of selection is increased genetic variance as the population becomes more diverse.

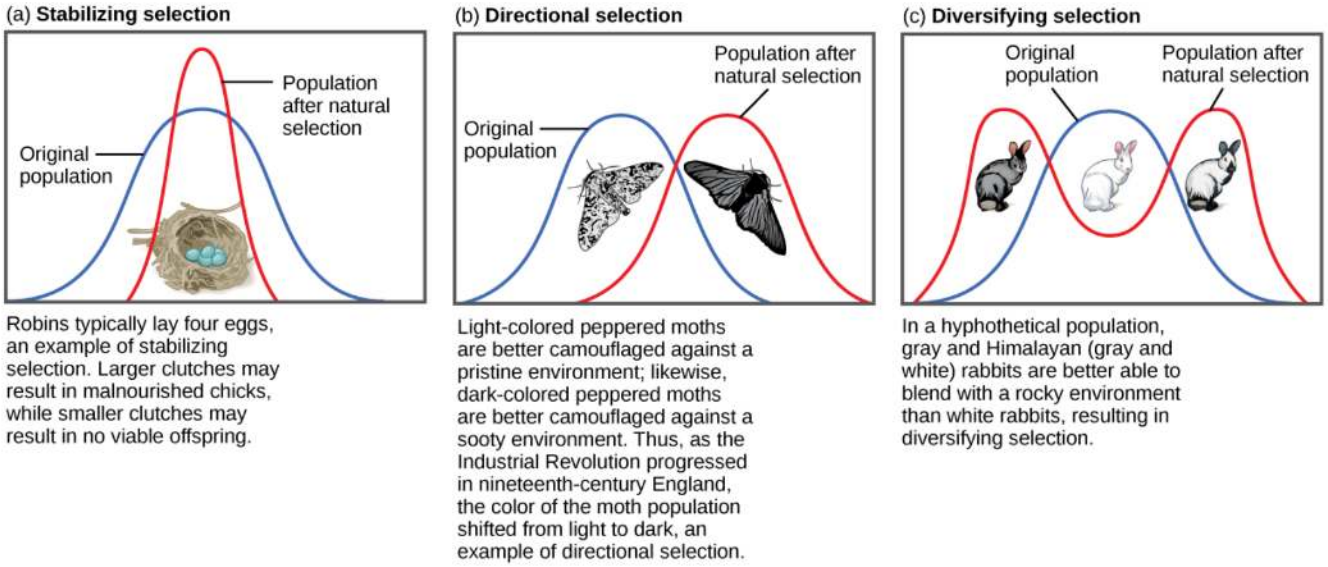


Figure **18.3.1**: Different types of natural selection can impact the distribution of phenotypes within a population. In (a) stabilizing selection, an average phenotype is favored. In (b) directional selection, a change in the environment shifts the spectrum of phenotypes observed. In (c) diversifying selection, two or more extreme phenotypes are selected for, while the average phenotype is selected against. Part (a) shows a robin clutch size as an example of stabilizing selection. Robins typically lay four eggs. Larger clutches may result in malnourished chicks, while smaller clutches may result in no viable offspring. A wide bell curve indicates that, in the original population, there was a lot of variability in clutch size. Overlaying this wide bell curve is a narrow one that represents the clutch size after natural selection, which is much less variable. Part (b) shows moth color as an example of directional selection. Light-colored pepper moths are better camouflaged against a pristine environment, while dark-colored peppered moths are better camouflaged against a sooty environment. Thus, as the Industrial Revolution progressed in nineteenth-century England, the color of the moth population shifted from light to dark, an example of directional selection. A bell curve representing the original population and one representing the population after natural selection only slightly overlap. Part (c) shows rabbit coat color as an example of diversifying selection. In this hypothetical example, gray and Himalayan (gray and white) rabbits are better able to blend into their rocky environment than white ones. The original population is represented by a bell curve in which white is the most common coat color, while gray and Himalayan colors, on the right and left flank of the curve, are less common. After natural selection, the bell curve splits into two peaks, indicating gray and Himalayan coat color have become more common than the intermediate white coat color.

? Exercise 18.3.1

In recent years, factories have become cleaner, and less soot is released into the environment. What impact do you think this has had on the distribution of moth color in the population?

Answer

Moths have shifted to a lighter color.

18.3.6 Frequency-dependent Selection

Another type of selection, called frequency-dependent selection, favors phenotypes that are either common (positive frequency-dependent selection) or rare (negative frequency-dependent selection). An interesting example of this type of selection is seen in a unique group of lizards of the Pacific Northwest. Male common side-blotched lizards come in three throat-color patterns: orange, blue, and yellow. Each of these forms has a different reproductive strategy: orange males are the strongest and can fight other males for access to their females; blue males are medium-sized and form strong pair bonds with their mates; and yellow males are the smallest, and look a bit like females, which allows them to sneak copulations. Like a game of rock-paper-scissors, orange beats blue, blue beats yellow, and yellow beats orange in the competition for females. That is, the big, strong orange males can fight off the blue males to mate with the blue's pair-bonded females, the blue males are successful at guarding their mates against yellow sneaker males, and the yellow males can sneak copulations from the potential mates of the large, polygynous orange males.



Figure **18.3.2**: A yellow-throated side-blotched lizard is smaller than either the blue-throated or orange-throated males and appears a bit like the females of the species, allowing it to sneak copulations. (credit: “tinyfroglet”/Flickr)

In this scenario, orange males will be favored by natural selection when the population is dominated by blue males, blue males will thrive when the population is mostly yellow males, and yellow males will be selected for when orange males are the most populous. As a result, populations of side-blotched lizards cycle in the distribution of these phenotypes—in one generation, orange might be predominant, and then yellow males will begin to rise in frequency. Once yellow males make up a majority of the population, blue males will be selected for. Finally, when blue males become common, orange males will once again be favored.

Negative frequency-dependent selection serves to increase the population’s genetic variance by selecting for rare phenotypes, whereas positive frequency-dependent selection usually decreases genetic variance by selecting for common phenotypes.

18.3.7 Sexual Selection

Males and females of certain species are often quite different from one another in ways beyond the reproductive organs. Males are often larger, for example, and display many elaborate colors and adornments, like the peacock’s tail, while females tend to be smaller and duller in decoration. Such differences are known as sexual dimorphisms, which arise from the fact that in many populations, particularly animal populations, there is more variance in the reproductive success of the males than there is of the females. That is, some males—often the bigger, stronger, or more decorated males—get the vast majority of the total matings, while others receive none. This can occur because the males are better at fighting off other males, or because females will choose to mate with the bigger or more decorated males. In either case, this variation in reproductive success generates a strong selection pressure among males to get those matings, resulting in the evolution of bigger body size and elaborate ornaments to get the females’ attention. Females, on the other hand, tend to get a handful of selected matings; therefore, they are more likely to select more desirable males.

Sexual dimorphism varies widely among species, of course, and some species are even sex-role reversed. In such cases, females tend to have a greater variance in their reproductive success than males and are correspondingly selected for the bigger body size and elaborate traits usually characteristic of males.



(a) (b) (c)

Figure **18.3.3**: Sexual dimorphism is observed in (a) peacocks and peahens, (b) Argiope appensa spiders (the female spider is the large one), and in (c) wood ducks; males are much more colorful than females. At left, a peacock with a bright blue body and flared tail feathers standing next to a brown, drab peahen. The middle photo shows a large female spider sitting on a web next to its male counterpart. The photo on the right shows a brightly colored male wood duck swimming next to a drab brown female.(credit "spiders": modification of work by "Sanba38"/Wikimedia Commons; credit "duck": modification of work by Kevin Cole)

The selection pressures on males and females to obtain matings is known as sexual selection; it can result in the development of secondary sexual characteristics that do not benefit the individual's likelihood of survival but help to maximize its reproductive success. Sexual selection can be so strong that it selects for traits that are actually detrimental to the individual's survival. Think, once again, about the peacock's tail. While it is beautiful and the male with the largest, most colorful tail is more likely to win the female, it is not the most practical appendage. In addition to being more visible to predators, it makes the males slower in their attempted escapes. There is some evidence that this risk, in fact, is why females like the big tails in the first place. The speculation is that large tails carry risk, and only the best males survive that risk: the bigger the tail, the more fit the male. This idea is known as the handicap principle.

The good genes hypothesis states that males develop these impressive ornaments to show off their efficient metabolism or their ability to fight disease. Females then choose males with the most impressive traits because it signals their genetic superiority, which they will then pass on to their offspring. Though it might be argued that females should not be picky because it will likely reduce their number of offspring, if better males father more fit offspring, it may be beneficial. Fewer, healthier offspring may increase the chances of survival more than many, weaker offspring.

In both the handicap principle and the good genes hypothesis, the trait is said to be an honest signal of the males' quality, thus giving females a way to find the fittest mates— males that will pass the best genes to their offspring.

18.3.8 No Perfect Organism

Natural selection is a driving force in evolution and can generate populations that are better adapted to survive and successfully reproduce in their environments. But natural selection cannot produce the perfect organism. Natural selection can only select on existing variation in the population; it does not create anything from scratch. Thus, it is limited by a population's existing genetic variance and whatever new alleles arise through mutation and gene flow.

Natural selection is also limited because it works at the level of individuals, not alleles, and some alleles are linked due to their physical proximity in the genome, making them more likely to be passed on together (linkage disequilibrium). Any given individual may carry some beneficial alleles and some unfavorable alleles. It is the net effect of these alleles, or the organism's fitness, upon which natural selection can act. As a result, good alleles can be lost if they are carried by individuals that also have several overwhelmingly bad alleles; likewise, bad alleles can be kept if they are carried by individuals that have enough good alleles to result in an overall fitness benefit.

Furthermore, natural selection can be constrained by the relationships between different polymorphisms. One morph may confer a higher fitness than another, but may not increase in frequency due to the fact that going from the less beneficial to the more beneficial trait would require going through a less beneficial phenotype. Think back to the mice that live at the beach. Some are light-colored and blend in with the sand, while others are dark and blend in with the patches of grass. The dark-colored mice may

be, overall, more fit than the light-colored mice, and at first glance, one might expect the light-colored mice be selected for a darker coloration. But remember that the intermediate phenotype, a medium-colored coat, is very bad for the mice—they cannot blend in with either the sand or the grass and are more likely to be eaten by predators. As a result, the light-colored mice would not be selected for a dark coloration because those individuals that began moving in that direction (began being selected for a darker coat) would be less fit than those that stayed light.

Finally, it is important to understand that not all evolution is adaptive. While natural selection selects the fittest individuals and often results in a more fit population overall, other forces of evolution, including genetic drift and gene flow, often do the opposite: introducing deleterious alleles to the population's gene pool. Evolution has no purpose—it is not changing a population into a preconceived ideal. It is simply the sum of the various forces described in this chapter and how they influence the genetic and phenotypic variance of a population.

18.3.9 Summary

Because natural selection acts to increase the frequency of beneficial alleles and traits while decreasing the frequency of deleterious qualities, it results in adaptive evolution. Natural selection acts at the level of the individual, selecting for those that have a higher overall fitness compared to the rest of the population. If the fit phenotypes are those that are similar, natural selection will result in stabilizing selection, and an overall decrease in the population's variation. Directional selection works to shift a population's variance toward a new, fit phenotype, as environmental conditions change. In contrast, diversifying selection results in increased genetic variance by selecting for two or more distinct phenotypes.

Other types of selection include frequency-dependent selection, in which individuals with either common (positive frequency-dependent selection) or rare (negative frequency-dependent selection) are selected for. Finally, sexual selection results from the fact that one sex has more variance in the reproductive success than the other. As a result, males and females experience different selective pressures, which can often lead to the evolution of phenotypic differences, or sexual dimorphisms, between the two. Sexual selection can result in the evolution of traits that attract the opposite sex, but at the same time might reduce chances of survival--one example is the peacock's large tail feather display which may attract females but also make escape from predators more difficult.

18.3.10 Attributions

Adapted by Kenneth A. Koenigshofer, PhD., from Adaptive Evolution by OpenStax, licensed CC BY 4.0

Adaptive Evolution by Lumen Learning, licensed CC BY-SA 4.0

This page titled 18.3: Chapter 3- Adaptive Evolution is shared under a mixed license and was authored, remixed, and/or curated by Kenenth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

18.4: Chapter 3- Modern Primates and Their Evolution

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe the distinguishing features of mammals and of primates
2. Discuss the range of physical and behavioral traits in several different species of primate
3. Describe general principles of brain evolution as proposed by Striedter (2006) and discuss one short coming of his principles
4. List major steps in primate evolution
5. Describe the taxonomic classification of our species, Homo sapiens

18.4.1 Overview

Humans are primates. We belong to the taxonomic order Primates. This order encompasses humans as well as non-human primates. Non-human primates are our closest biological and evolutionary relatives. So, we study them in order to learn more about ourselves. All primates are mammals.



Figure **18.4.1**: Golden Lion Tamarin, a primate found in Brazil. Woodland Park Zoo, Seattle, WA 9/30/2011

18.4.2 General Mammalian Characteristics

The earliest evidence of mammals is from the Mesozoic era (see table of geological time below), however, there is limited fossil evidence and the fossils that have been found are mouse-like forms with quadrupedal locomotion. Primates evolved from an ancestral mammal during the Cenozoic era and share many characteristics with other mammals. The Cenozoic is the era of the adaptive radiation of mammals with thirty different mammalian orders evolving.

Some of the general characteristics of mammals includes:

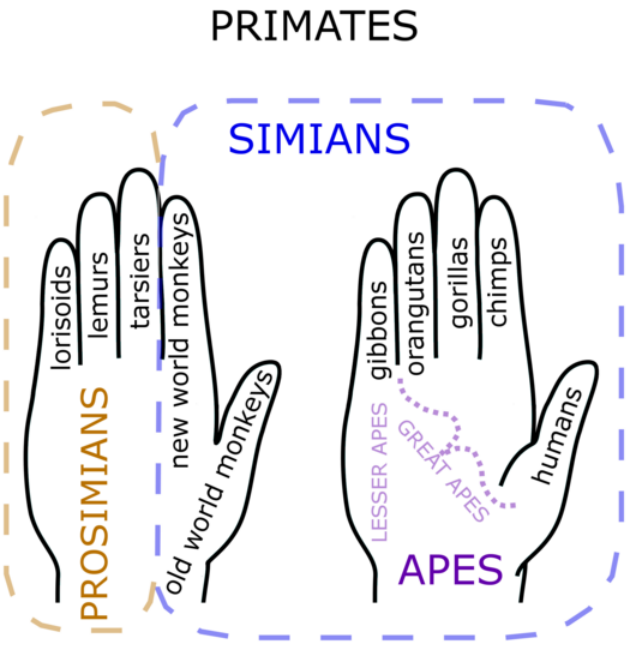
- mammary glands: females produce milk to feed young during their immediate post-natal growth period
- mammals have hair (sometimes called fur) that covers all or parts of their body
- the lower jaw is a single bone
- the middle ear contains three bones: stapes (stirrup), incus (anvil), and malleus (hammer)
- four-chambered heart
- main artery leaving the heart curves to the left to form the aortic arch
- mammals have a diaphragm
- mammals regulate their body temperature to maintain homeostais (a constant body temperature)

- teeth are replaced only once during lifetime

18.4.3 Primate Evolution in Summary

The first fifty million years of primate evolution was a series of adaptive radiations (diversification of a group of organisms into forms filling different ecological niches) leading to the diversification of the earliest lemurs, monkeys, and apes. The primate story begins in the forest canopy (overlapping branches and leaves high above ground) and understory of conifer-dominated forests, with our small, furtive ancestors subsisting at night, beneath the notice of day-active dinosaurs. From the archaic primates to the earliest groups of true primates (euprimates), the origin of our own order (the Primates) is characterized by the struggle for new food sources and microhabitats in the arboreal setting. Climate change forced major extinctions as the northern continents became increasingly dry, cold, and seasonal and as tropical rainforests gave way to deciduous forests, woodlands, and eventually grasslands. Lemurs, lorises, and tarsiers—once diverse groups containing many species—became rare, except for lemurs in Madagascar where there were no anthropoid competitors and perhaps few predators. Meanwhile, anthropoids (monkeys and apes) emerged in the Old World (Africa, Europe, and Asia), then dispersed across parts of the northern hemisphere, Africa, and ultimately South America. Meanwhile, the movement of continents, shifting sea levels, and changing patterns of rainfall and vegetation contributed to the developing landscape of primate biogeography, morphology, and behavior. Today's primates provide a few reminders of the past diversity and remarkable adaptations of their extinct relatives.

18.4.4 A Simplified Classification of Living Primates



Drawing of two hands showing prosimians on left palm, apes on the right palm, and name of a primate or group on each finger.

Figure 18.4.2: This image is a hand mnemonic used to help students learn a categorization of primates. The right hand (held palm upward) correlates to the apes, including the great apes and lesser apes. Humans are distinct from that group as the thumb is distinct from the other four digits of the hand. All apes evolved to be tailless, whereas species grouped onto the left hand are characterized as having tails (with certain exceptions, such as the Barbary macaque). Old world monkeys have a family name that means "tailed ape" and are indicated on the left thumb that points toward the right hand of apes. Old world monkeys are grouped with all of the apes in the parvorder called Catarrhini. New world monkeys on the left index finger form their own parvorder called Platyrrhini (meaning "flat nosed"). These are the only monkeys with prehensile tails. The group of simians (higher primates) are all apes and

monkeys, so includes all of the above (Catarrhini and Platyrrhini). The three remaining digits on the left hand form the group of prosimians (lower primates).

The hand phalanges are not to be mistaken for a phylogeny as the branching geometry is not accurate. And the ten hand digits do not correspond to any single particular level of taxonomy. The specific correspondence of digits is:

- right thumb = genus *Homo* => 1 species: humans,
- right index = genus *Pan* => 2 species: common chimpanzees (4 subspecies) and bonobos,
- right middle = subfamily Gorillinae => genus *Gorilla* => 2 species: western gorillas (2 subspecies) and eastern gorillas (2 subspecies),
- right ring = subfamily Ponginae => genus *Pongo* => 2 species: Bornean orangutans (3 subspecies) and Sumatran orangutans,
- right pinky = family Hylobatidae => 4 genera of gibbons => nearly twenty species in total, including siamangs, lar gibbons and hoolock gibbons,
- left thumb = superfamily Cercopithecoidea (old world monkeys) => well over one hundred species including baboons and macaques,
- left index = parvorder Platyrrhini => superfamily Ceboidea (new world monkeys) => well over one hundred species including marmosets, tamarins, titis, howlers and squirrel monkeys,
- left middle = infraorder of tarsiers,
- left ring = superfamily of lemurs,
- left pinky = superfamily of lorisoids.

Categorization of humans has not been without controversy. One position is that homo sapiens are above apes and that it is improper to categorize the species as one of the great apes. The opposite extreme is the view that humans are "the third chimpanzee", based upon the very high percentage of genetic commonality between the species. This image follows a convention found between these two positions.

(Image and caption for Figure 18.4.2 from Wikimedia Commons; File:Primate Hand Mnemonic.png; https://commons.wikimedia.org/wiki/F...d_Mnemonic.png; by Tdadamemd; licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license).

18.4.5 General Primate Characteristics

While primates share traits with other mammals, such as mammary glands and regulation of body temperature, there are a number of traits that all primates share. Because humans and non-human primates share a common evolutionary history (we have a common primate ancestor), human and non-human primates have a number of similar biological and behavioral traits.

18.4.5.1 Body

Primates have a **flexible** and **generalized** limb structure that's able to move readily in many directions. Compare the flexibility of the spider monkeys on the left with the horse on the right.

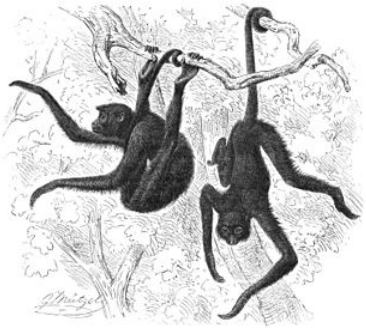


Figure **18.4.3**: Spider monkeys.



Figure **18.4.4**: Horse.

18.4.5.2 Hands

Primates have **prehensile** hands (and most of them have prehensile feet also). This means that they have the ability to grasp and manipulate objects. Primates have 5 digits on their hands and feet. Note that a few primates, like spider monkeys, have what are called **vestigial** thumbs. This means that they have either a very small or non-existent external thumb (but in that case, they will still have a small internal thumb bone).



Figure **18.4.5**: Primate hands and feet.

18.4.5.3 Teeth

Primates have different teeth that perform different tasks when processing food via biting or chewing. Thus, they have the ability for a more generalized diet (compared to a specialist diet), meaning that they have more dietary flexibility. Why would this be a good thing to have?

Compare the teeth of the cayman (a relative of crocodiles and alligators) with the human teeth on the right. While the cayman's teeth do vary in size, they don't vary in structure. However, the human has incisors, canines, premolars, and molars -- all of which perform different food processing tasks.



Figure **18.4.6**: Cayman.



Figure **18.4.7**: Human jawbone.

18.4.5.4 Senses

Compared to most other mammals, primates have an increased reliance on vision and a decreased reliance on their senses of smell and hearing. Associated with this are their smaller, flattened noses, loss of whiskers, and relatively small, hairless ears. Also associated with this are their forward-facing eyes with accompanying binocular or stereoscopic vision. This type of vision means that both eyes have nearly the same field of vision with a lot of overlap between them. This arrangement provides good depth perception (but a loss of peripheral vision)--very useful for moving through tree branches.

Note first how the eyes of the monkey on the left are more front-facing than the eyes of the cow on the right. Also note how flat the monkey's nose is and how small its ears are, when compared to those of the cow.



Figure 18.4.8: Monkey.



Figure 18.4.9: Cow. Note placement of the eyes compared to the monkey.

18.4.5.5 Brain Evolution

In relation to other mammals, primates have a more expanded and elaborate brain, including expansion of the cerebral cortex. Compare the complexity of the human brain on the left to the cat brain on the right.

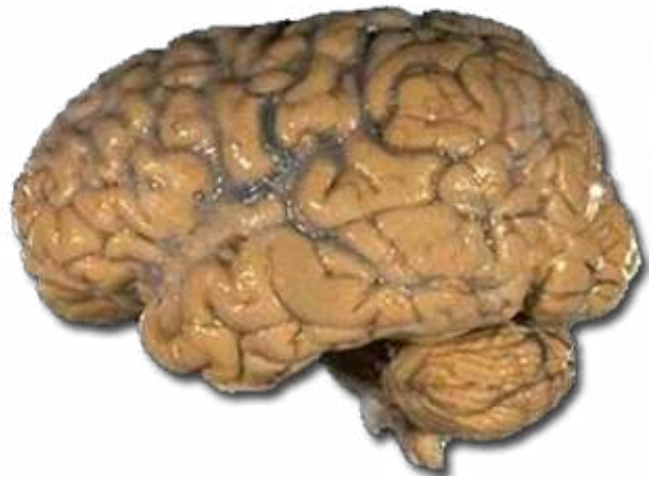


Figure 18.4.10: Human Brain.



Figure 18.4.11: Cat Brain.

Clearly, significant anatomical changes have taken place during the evolution of the brain in primates, in other mammals, and in animals in general. Are there generalizations that can be made about the evolution of animal brains?

Striedter (2006) has identified a number of general principles of brain evolution applicable across a wide range of species (i.e. not just primates or mammals).

1) Embryonic brains across species are more similar than adult brains, because brains tend to diversify more as they grow toward adult form;

2) relative brain size to body size in the vertebrates (animals with backbones) has tended to increase more often than decrease over evolutionary time;

3) it appears that increases in relative brain size were generally accompanied by increases in social or food foraging complexity;

4) most increases in relative brain size were accompanied by increases in absolute body size;

5) increases in absolute brain size require changes in the brain's internal connections which imply greater modularity (specialized processing modules increasing the "division of labor") of brain anatomy and functioning;

6) evolution generally enlarges brains by extending the period of brain development while conserving (keeping the same) the "birth order" of different brain regions, so that big-brained animals tend to have disproportionately larger late-"born" (late-developing) regions ("late equals large"), such as cerebral cortex, leading to disproportionately more cerebral cortex (increased corticalization) in big-brained mammals (non-mammals don't have cerebral cortex); however there are exceptions to this rule, for example, at any

given absolute brain size, there is more cerebral cortex in simians than in prosimians, and in parrots there is an unusually large telencephalon which is not accounted for by the rule;

7) changes in size proportions of brain areas, although "automatic" within the scaling (allometric) rules above, can still be adaptive and undergo natural selection;

8) as brain regions increase in absolute or proportional size, they tend to become laminated--organized into sheets of neurons--allowing point for point corresponding connections between sensory and motor maps with minimal axonal and dendritic wiring, saving space and metabolic energy;

9) as brain size increases, more regional subdivisions occur from ancestral parts subdividing into new parts, as in the dorsal thalamus, or, as in the case of neocortex, a new part was added to an ancestral set of conserved parts;

10) Deacon's rule is that "large equals well connected," meaning that as the relative size of a brain structure increases it tends to receive more connections and to project more outputs to other structures.

Striedter (2006) adds a number of additional generalizations about the mammal and primate brains, including human:

11) six-layered mammalian neocortex probably evolved from a 3-layered reptilian precursor (something like that in turtles) called dorsal cortex by adding several layers;

12) aside from neocortex, the mammalian brain is similar to the reptilian brain (which also has hippocampus, for example) but even with a "fundamental scheme" of brain regions and circuitry, many minor changes in wiring can drastically change how information flows through a brain and thus how it functions--thus, the mammal brain is not just an upscale version of the reptilian brain;

13) increasing corticalization in mammals cannot be explained in terms of the above scaling (allometric) rules and involved highly specialized changes in brain anatomy presumably due to natural selection which expanded precursor sensory and motor cortical regions;

14) bird forebrains evolved along a very different path with expansion of their dorsal ventricular ridge (DVR), the major sensorimotor region of the avian telencephalon, highly similar in function to mammalian neocortex, making "many birds at least as intelligent as most mammals."

Striedter adds a number of points about the human brain in an attempt to identify features that make it special compared to the brains of other mammals.

15) In the six million years since bipedal apes (hominins) diverged from other apes, absolute brain size increased radically (about fourfold), not gradually, but in bursts--from when genus *Homo* first evolved, absolute brain size doubled from 400 to 800 cubic centimeters, then remained relatively steady in *Homo erectus* during the next 1.5 million years, but then exploded again in the transition to *Homo sapien* until about 100,000 years ago, at which time absolute brain size reached its current value of about 1,200 to 1,800 cubic centimeters. The first jump in Homo brain size was likely related to change in diet involving transition to meat and later the cooking of meat. The second leap was perhaps stimulated by competition among humans for mates and other resources;

16) the principle of "late equals large" predicts large neocortex in humans (the human neocortex to medulla ratio is twice that of chimpanzees);

17) the principle of "large equals well connected" is consistent with known expanded numbers of projections from human neocortex to motor neurons in medulla and spinal cord permitting greater precision of control over muscles serving hands, lips, tongue, face, jaw, respiratory muscles, and vocal folds, required for the development of human language about 50,000 to 100,000 years ago;

18) once human language appeared, dramatic changes in human behavior became possible without further increases in brain size;

19) increase in brain size has some disadvantages: including increased metabolic costs because the brain utilizes so much metabolic energy (20% of human metabolic energy even though it is only 2% of human body weight; being so metabolically expensive increases in brain size must be paid for by improved diet or reduction of other metabolic energy demands); decreased connectivity perhaps making the two hemispheres more independent of one another, and perhaps explaining why the two cerebral hemispheres became functionally specialized; and size limits of neonatal brain due to constraints imposed by size of the human mother's pelvis and birth canal. According to Striedter, these costs may explain why human brain size plateaued about 100,000 years ago;

20) within neocortex, the lateral prefrontal cortex has become relatively enlarged in the human brain, likely increasing its role in behavior (see Chapter 14 for discussion of the the lateral prefrontal cortex and higher cognitive functions);

21) some key evolutionary changes in brain structure were not caused by increases in absolute or relative brain size, such as evolution of the neocortex of mammals, but require additional explanation; comparing distantly related species on absolute brain size alone misses important factors, for example, brains of some large whales weight 5 times as much as the human brain but whale brains have poorly laminated and thin neocortex; in cases of distantly related species, comparisons of relative brain size are more useful, for example, human and some toothed whales have relative brain sizes significantly larger than average mammals of similar body size;

22) two general hypotheses about brain evolution are that individual brain systems evolve independently by natural selection (the mosaic hypothesis) or alternatively that components of such systems evolve together because of functional constraints (the concerted or constraint hypothesis), with a third view being that all brain evolution is simultaneously both mosaic and concerted.

One problem with Striedter's approach is that it doesn't explicate the forces of natural selection that may account for more specific features of brain evolution in specific species, including our own. He admits that evolution of the neocortex cannot be explained by evolution of bigger brains and that neocortex evolved independently of absolute brain size. However, aside from restating the theory that complex social life in ancestral humans, along with competition among humans for resources, stimulated the evolution of increases in the size and complexity of the neocortex, he offers little about what role natural selection played in the evolution of neocortex, or in brain evolution in general. As Adkins-Regan (2006, p.12-13) states in her critique of Striedter's work, "There is relatively little discussion of tests of hypotheses about the selective pressures responsible for the origin and maintenance of traits . . . The author would seem to be experiencing symptoms of discomfort with the concept of adaptation. . ..Given that brain mechanisms are products of natural selection, a central strategy in neuroscience should be to use the methods of evolutionary biology, which have been so successful in helping us to understand the mechanisms and design of organisms generally." In Chapter 14 of this text, on Intelligence, Cognition, and Language, consistent with this advice, you will find an extensive discussion of the role of natural selection in evolution of brain systems involved in intelligence and thinking.

18.4.5.6 Primate Life History

Life history refers to the pattern that an organism takes from conception to death. When compared to other mammals, primates have:

- longer gestation (pregnancy) periods
- reduced number of offspring (usually one, but some species commonly have twins)
- delayed maturation, with a long infancy and juvenile learning period, reflecting substantial brain development after birth
- longer lifespan



Figure 18.4.12: Chimpanzee mother and baby.

18.4.6 Primate Ecology

Primate ecology refers to the relationship between primates and their environment. Their environment includes not only the physical environment (e.g., trees, water, weather) but also the other animals in the environment, including other non-human primates and even humans. Why should we care about primates and their ecology? Remember that evolution is always environmentally dependent. Also, keep in mind that primates are not only affected by their environments, they affect their environments as well (by eating plants and insects, dispersing seeds, etc.).



Figure 18.4.13: Monkey feeding.

There are two primary environmental factors to consider: Food and Predators.

1) **Food**

- Primate researchers examine the quality, quantity, and distribution of food in a primate's environment.
- Why? Because food = babies. This means that those individuals who acquire more high-quality food more efficiently are likely to have more offspring and therefore will be more evolutionarily successful.
- There is greater pressure for females than for males, because of the higher biological costs of reproduction for primate females (**reproductive asymmetry**)--remember that primates spend a very long time as dependent nursing infants, so primate mothers have a heavy burden when caring for them (the only exceptions are some New World monkeys: the father cares for the infants, but the mother still nurses them).

2) **Predators**

- the types and distribution of potential predators in a primate's ecosystem is important.
- Why? Because no one wants to get eaten, right? Those who can most efficiently and effectively avoid predation are likely to have more offspring and therefore will be more evolutionarily successful.

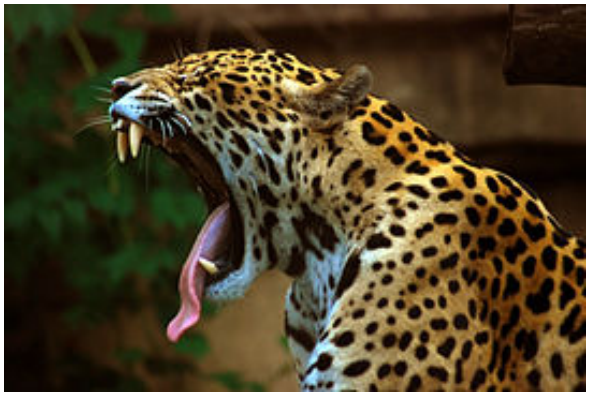


Figure **18.4.14**: Jaguar, a primate predator.

Other important environmental factors are:

- weather
- the distribution of water
- the distribution of sleeping sites, and
- the primate's relationships with other individuals within their social group, other groups of their own species, and other species (including humans).

18.4.6.1 Food



Figure **18.4.15**: Baboon examining a leaf.

Most primates are **herbivores** (they eat plant foods) and are dietary generalists. Some primates are **omnivores** and eat lots of things (plant and animal). However, some primates are more specialized.

- **Folivores**: eat mainly leaves.
- **Frugivores**: eat mainly fruit.

- **Insectivores**: eat mainly insects.
- **Gummivores**: eat mainly tree sap.



Figure 18.4.16: Lemur eating fruit.

One of the challenges that primates face in their day-to-day life is a type of **evolutionary arms race** they have with their food. Some food items, like fruit, "want" to be eaten. In other words, it's evolutionarily advantageous for a fruit to be eaten by a primate, who then carries the fruit's seeds far away from the parent plant in its stomach, to deposit them elsewhere. The fruit benefits from this **seed dispersal**, thereby limiting competition between the parent plant and its offspring. This is one of the hypothesized reasons that fruits have high sugar or fat contents, to make them attractive to seed dispersers, such as primates and birds.

However, some food items, like leaves, don't "want" to be eaten. It's not evolutionarily advantageous for a plant to be stripped of its leaves. A leafless tree would not be able to perform photosynthesis and would promptly die. This is why plants have developed certain chemicals in their leaves that make the leaves unpalatable or indigestible to primates. In these cases, most leaf-eating monkeys prefer to eat young leaves, which are visually identifiable (either lighter in color or a different color) and higher in nutrients than mature leaves.

Some primates have evolved color vision in order to more successfully forage for food and other primates have developed specialized digestive systems to deal with low-quality (high-fiber) foods.



Figure 18.4.17: Black and white colobus monkey resting.

18.4.6.2 Activity budgets

How a primate spends its time is called its **activity budget**. These are comprised of the major categories of:

- foraging/feeding
- mating
- social behavior
- locomotion/traveling, and

- resting/sleeping

18.4.6.2.1 Why should we care about activity budgets?

Examining a primate's activity budget gives a good idea of how that particular primate (or primate group) "makes a living". For example, time spent foraging, gives clues about how much food in its area and how much time it has to spend searching for food. Eating meat provides a concentrated source of calories; meat eating after a successful hunt in meat-eating primates such as chimpanzees and humans leaves more time and calories for activities other than food gathering.

Remember also that if a primate has to spend more time eating or moving around in order to find something to eat, then it has less time for either resting or socializing. Primates only have a finite amount of time in their day, so they have to maximize the use of it the best they can within their environmental parameters.

Activity budgets denote **when** a primate is "making a living":

- **Diurnal**: active during the day, generally inactive at night. Most primates are diurnal.
- **Nocturnal**: active during the night, generally inactive during the day.
- **Crepuscular**: active during dawn and dusk. Ring-tailed lemurs (pictured below left) have this activity pattern.
- **Cathemeral**: active during irregular periods during day and night. Black lemurs (pictured below right) are one of the few primate species who have this activity pattern.



Figure 18.4.18: Ringtailed lemur.



Figure 18.4.19: Black lemurs.

Note that these activity patterns are not hard and fast rules. As with diet, there is a lot of variation in how flexible a particular species (or group within a species) is when it comes to their activity patterns. For example, although traditionally thought of as strictly nocturnal, the activity patterns of owl monkeys have been found to be highly sensitive to environmental factors such as temperature and amount of moonlight.

18.4.6.2.2 Food and Feeding Competition

One of the main factors that affects a primate "making a living" is **feeding competition** (competing with others for food).

There are two groups that a primate competes with;

- other primates in its social group (**within-group competition**)
- primates (of the same species) in other groups (**between-group competition**)



Figure 18.4.20: Macaques fighting.

There are also two types of competition. The type of competition prevalent in a situation depends on the quality and quantity of food available in the environment.

- **Scramble competition**: happens when there is a lot of low quality food. This is more common for leaf-eating primates, because trees tend to have large quantities of edible leaves (although some types, such as young leaves, may be preferable to others). It's an **indirect competition** in which whoever finds food faster or eats faster gets more food than do other individuals.
- **Contest competition**: happens when there is a small amount of high quality food. This is more common for fruit-eating primates, because a fruiting tree has a limited amount of high-quality fruit. It's a **direct competition**, where certain individuals (stronger, higher social status) get more food than do other individuals via squabbling or fighting.

In total, one can have:

- within-group scramble competition
- within-group contest competition
- between-group scramble competition, and
- between-group contest competition.

Feeding competition can be severe and have serious effects on the health, well-being, and reproductive capacities of primates. In other words, it has a direct effect on their evolutionary fitness. As expected, primates demonstrate a lot of **behavioral strategies** to deal with the feeding competition they face. For example, a group with **high scramble competition** may **spread out** more when feeding so as to access more resources OR in a group with **high contest competition**, some individuals may form **coalitions** in order to "gang up on" other primates to take their food.

18.4.6.2.3 Spatial use: home ranges and territories

The area that a primate uses is its **home range**. The part of the home range that's used most often is the **core area**. Some species, such as snub-nosed monkeys, have very large home ranges (32 sq. km) and other species, such as the pygmy marmoset, have very small ones (0.003 sq. km). A primate's home range has to contain all of the resources it needs in order to survive (food, water, sleeping sites, etc.). The home ranges of different primate groups often overlap, sometimes only a little, but sometimes a lot. If the home range is physically defended, it's a **territory**. Chimpanzees are famous for their territorial conflicts, which are so organized and violent that scientists have begun calling it "warfare".

18.4.7 Primate Behavior

The behavioral traits we share with other primates include:

- a greater dependence on flexible, learned behavior, reflecting neuroplasticity of the brain and corticalization (increased development of cerebral cortex)
- a tendency to live in social groups



Figure 18.4.21: Baboons crossing the road. Baboons live in complex social groups.

18.4.7.1 Primate Social Behavior

One of the fundamental traits that primates share is that we are **social** creatures. All primates share some form of social structure with others of their own species, whether that be in a large permanent social group or as an individual who has repeated short-term interactions with others. The social relationships that primates have with other members of their own species have a huge impact on the individuals involved and are incredibly important to their health, well-being, and reproductive success. This makes the study of primate social behavior very important.



Figure 18.4.22: Chimpanzees, our closest living relative, in captivity.

Primate behavior studies come in three formats:

- **Captive**: The animals are in captivity. Variables are easy to control in these situations (for example, the number of individuals in a group or food availability) and the primates are easier to observe at close-range. However, because of the artificiality of the situation, the primates may not be exhibiting normal behavior.
- **Semi-captive (semi-free-ranging)**: The animals are captive, but in a very large area like an island or a fenced-in compound. Variables are still relatively easy to control and the primates are easier to observe (than in the wild). The primates tend to exhibit more natural behavior patterns in this type of setting than when compared to a completely captive situation.
- **Free-ranging**: The animals are living in the wild, in their natural environment. This is the most logistically difficult type of primate study. The animals are more difficult to observe (often requiring a **habituation** period to get them accustomed to the presence of observers). However, the primates are most likely to perform their normal array of behaviors in this type of study.

18.4.7.2 Social structure: the "whys" and the "hows"

The basic premise of why primates have certain social behaviors is due to **reproductive asymmetry**. Females are under a lot more pressure than males to forage effectively because of the biological pressures they're under (due to the burden of reproduction). It is precisely this inequality that causes the sexes to have different evolutionary priorities:

- For **females**, their priority is **food**. Females can only reproduce at a certain rate, due to the length of time it takes to be pregnant, raise an infant to independence, and have their bodies recover to do it all over again. Access to lots of good quality food through little effort is the key to accomplishing this task.
- For **males**, their priority is finding **mates**. Males can reproduce at a much faster rate than can females. The only real constraint on their reproductive rate is how frequently they can get fertile females to mate with them.

So, **primate social structure** works like this:

- male distribution follows how the females distribute themselves, and
- the females distribute themselves depending on how the food is laid out in their environment.

Social groups have two **immigration/emigration** patterns:



Figure **18.4.23**: Baboon social group.

Female philopatry:

- Females do not emigrate (depart) from their birth group at sexual maturity.
- Males usually do emigrate from their birth group at sexual maturity.
- Females form the core of the group, are biologically related, and have tight social bonds (often exhibited through social cooperation and grooming: seen in these baboons, pictured right). Males have few positive social relationships.
- This is the most common form of social system in primates.

Male philopatry:

- Males do not emigrate from their birth group at sexual maturity.
- Females usually do emigrate from their birth group at sexual maturity.
- Males form the core of the group, are related and have tight social bonds. Females have few positive social relationships.
- This type of social system tends to occur when resources are widely dispersed, so females are widespread and difficult for males to consistently access.

18.4.7.2.1 Social strategies

In order to achieve their priorities, male and females must utilize certain social strategies.

Dominance

- As with chickens, primates have a "pecking order" or **dominance hierarchy** in their social groups.
- One's position in the dominance hierarchy often gives them access to preferred resources, including not only food and mates, but other resources such as sleeping sites and water.
- Sometimes dominance hierarchies are determined through fighting, but more often they are sorted out through a series of aggressive/submissive non-contact interactions, such as approach/avoidance, facial expressions, and body postures. Primates try to avoid direct aggressive contact in an effort to avoid risk of bodily harm.



Figure 18.4.24: Macaques grooming.

Cooperation

- Dominance hierarchies aren't completely linear. One's rank in the hierarchy often depends on who they can get to cooperate with them during conflicts. For example, Monkey 2 may be submissive to Monkey 1 when alone, but when her buddy Monkey 3 is around, the two of them cooperate and chase Monkey 1 away from food together. Therefore, Monkey 2's position in the dominance hierarchy is situationally dependent.
- These cooperative relationships are usually between relatives. In male philopatric groups, they're usually brothers. In female philopatric groups, it's often mother-offspring or siblings.
- The relationships are cultivated through **affiliative behaviors**, such as play, grooming, and other forms of body contact (e.g., hugging). So, primates practice *You scratch my back, I'll scratch yours* both literally and figuratively.

18.4.7.2.2 Why be in a social group?



Figure 18.4.25: Harpy Eagle. Predator of smaller primates. (Copyright; author via source).

So, why be in a social group if it is just one big mess of food competition, mate competition, and political posturing? Well, because there are benefits, of course!

Living in a social group:

- provides access to **mates**. While there is, of course, competition for mates within the group, a social group ensures that there are at least mates available to fight over.
- provides more eyes looking for **food** and more brains remembering where food is in the ecosystem
- provides **anti-predator defense**
- There are more eyes on the lookout for predators.
- An additional anti-predator benefit is known as "The Selfish Herd" (safety in numbers). It basically comes down to "you don't have to run faster than the predator, just faster than the other guy".

- Some social groups will "mob" predators and drive them away. This only works on smaller predators, especially those who rely on surprise attacks (such as the Harpy Eagle, pictured above).

18.4.7.3 Types of primate social groups

Solitary:



Figure **18.4.26**: Galago. These small nocturnal primates are also called "bush babies" and are native to sub-Saharan Africa. Social groups consist of closely related females and their young. Males leave the group at puberty.

- Males are alone most of the time, except when seeking out mates. Females live with their dependent offspring.
- One male will have territorial overlap with several females.
- Some prosimians have this social system, like the galago, pictured above.

Monogamy:



Figure **18.4.27**: Titi monkey.

- A social group is comprised of a male, a female, and their dependent offspring.
- This is also called "pair bonded".
- This is common in New World monkeys (for example, titi monkeys, pictured above) and the "lesser apes" (siamangs and gibbons).

Single-male multi-female:



Figure 18.4.28: Diana monkey.

- A social group is comprised of one adult male, several (to many) adult females, and their dependent offspring.
- Males that don't belong to a social group may either live alone or in multimale (bachelor) groups that have no females.
- It's a difficult life for the resident male in these social groups, because not only did he have to fight his way into the group (either driving the previous resident male out or stealing females from another male), he has to continually fight to keep other males out of his group, while keeping the females in the group.
- When males take over groups, some may commit infanticide (killing the previous resident male's infants).
- This is a common social group in Old World monkeys (like the Diana monkey, pictured above)

Multi-male multi-female



Figure 18.4.29: Hanuman langur.

- A social group is comprised of more than one adult male, more than one adult female, and their dependent offspring
- In these groups, in order to have preferential access to females, males will form dominance hierarchies or develop biological characteristics that the females find attractive. For example, during the mating season, squirrel monkey males begin storing large amounts of water and fat on their bodies.
- Some species (such as Hanuman langurs - pictured above) have both single-male and multimale-multifemale social groups in the same population.

Fission-fusion



Figure 18.4.30: Bonobo. Chimps and bonobos are our closest living relatives. Humans, chimps and bonobos descended from a single ancestor species that lived six or seven million years ago.

- In this situation, individuals belong to a large group called a community. Each community has a home range and consistent community membership. Within the community, individuals form temporary foraging groups called parties. Parties have an unpredictable membership. So, on any given day (or even part of a day), one cannot predict who will be foraging with whom.
- Females always travel with their dependent offspring.
- Sometimes males form partnerships and forage together.
- This type of social group is thought to have been an adaptation to environments with patchy, unpredictable fruit availability.
- It's found in chimpanzees, bonobos (pictured above), and spider monkeys.

Single-female multi-male



Figure 18.4.31: Tamarins.

- A social group is comprised of one adult female, more than one adult male, and their dependent offspring.
- This is rare in primates.
- Found in marmosets and tamarins (pictured right), this type of social system is thought to be an adaptation to the twinning (giving birth to twins) that it common to these small primates. The males carry the dependent infants.

18.4.8 Primate Evolution

Now that you have an understanding of living primates' morphology and behavior, it is time to learn about the origins of primates. the study of primate evolution is multidisciplinary in nature and incorporates data and methods from paleontology, geology, anthropology, and archaeology to study the fossil record of primates.

18.4.9 Fossils



Figure **18.4.32**: Permineralized wood.

Fossils are at the center of the study of ancestral primates. Animal fossils provide insight into morphology and behavior of ancient organisms while plant fossils help to determine what the landscape may have looked like during the time that a particular organism, including primates, lived there (remember, evolution is environmentally dependent!).

When a fossil is found, it can be dated using a variety of methods. Some methods date the fossil itself; other methods use the surrounding indicators There are relative dating techniques that provide an order of occurrence, but no absolute dates, and there are absolute, or chronometric, dating techniques that provide dates (usually a range of dates) for an object.

18.4.9.1 Geologic time

From geology we have an understanding of the time involved in the development of earth. In this section we are specifically focused on the late Mesozoic/early Cenozoic eras in relation to primate evolution (see chart below).

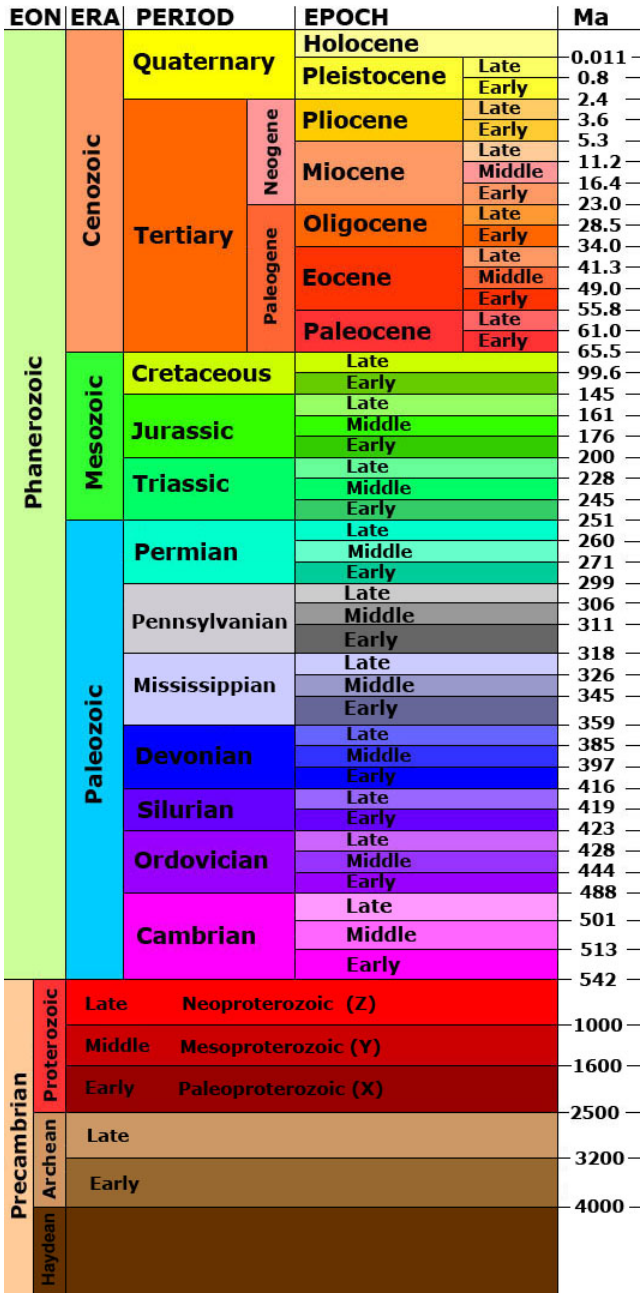


Figure 18.4.33: Geologic time scale (Ma = millions of years ago).

18.4.10 Tracking Primate Evolution

18.4.10.1 The Mesozoic: the origin of mammals

The Mesozoic era is known as "the age of the dinosaurs" due to their ecological dominance at the time; however, it is during this era that the first mammals evolved, including a primate-like mammalian ancestor. Research using DNA analysis and fossils suggests that by 75 mya (million years ago) all of the mammalian orders had diverged.

18.4.10.2 The Cenozoic era

18.4.10.2.1 The Paleocene epoch

Proto-primates, primate-like mammals, evolved in the Paleocene epoch, about 65 mya. The proto-primates from this epoch are controversial; some argue that they are related to primates but are not actually primates (hence, "proto-primates").

18.4.10.2.2 The Eocene epoch

The Eocene epoch (56-33 mya), with its warmer and wetter climate and diversification of rainforests and flowering plants, sees the emergence of the first true primates.

Eocene primates were (Jurmain et al. 2013: 189):

- about the same size as modern squirrels, with prehensile hands and feet and more forward-facing eyes that gave them stereoscopic vision.
- widely distributed
- mostly extinct by the end of the Eocene
- most are not ancestral to later primates

18.4.10.2.3 The Oligocene epoch: the monkey/ape divergence (34 to 23 mya)

The global climate shifted again, cooling and drying, during the Oligocene epoch (33-23 mya). There is a reduction in the amount of rainforests, which retreated toward the equator, but an expansion of grasslands. So, there was an increase in terrestrial (ground) niches and a decrease in arboreal (tree) niches. While most of the fossils from this time are Old World monkeys, some are ancestors of New World monkeys; at this point the American continents had separated from Africa and Eurasia.

18.4.10.2.4 The Miocene epoch: the Old World monkey/ape divergence (23 to 5 mya)

There were two climate shifts during this period, first there was a warm period with heavy forestation, followed by drier and cooler climates with decreasing forest and increased grasslands.

At this point, the Old World monkeys and apes split. The fossils are from east Africa.

The ape ancestors had the following traits:

- no tail
- arboreal quadrupeds (limb structure was still monkey-like)

The Old World monkey ancestors at this point in time had the following traits:

- A tail
- Had macaque-like faces
- Were medium sized 7-11 lbs

18.4.11 Where did primates come from?

There are three primary hypotheses about the origins of primates. The earliest hypothesis, the arboreal hypothesis, claims that the first primates evolved a suite of traits for living in trees, e.g., grasping hands and feet and stereoscopic vision. This hypothesis held sway from the early 1900s until the 1970s when the visual predation hypothesis was proposed. This hypothesis suggests that primates evolved because they provided an advantage for hunting small insects, e.g., orbital convergence that allowed for 3d-vision. Shortly thereafter, the angiosperm hypothesis was proposed, which claims that primate evolution is related to the adaptive radiation (diversification into different ecological niches) of angiosperms or flowering plants. Each of these hypotheses have been challenged over time, and it may be that each may play a part in explaining primate evolution.

18.4.12 The Taxonomy of Humans

At this point, it is useful to review taxonomy, the biological classification of organisms, using the human species to illustrate the taxonomic classifications.

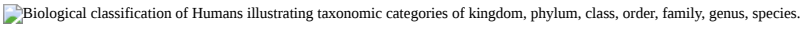


Figure **18.4.34**: Taxonomic classification of humans illustrating taxonomic categories used for all living things. (Image from Wikimedia Commons; File:Biological classification of Humans Britfix.png; https://commons.wikimedia.org/wiki/F...ns_Britfix.png; by L Pengo PD-User, Modified by Britfix; made available under the Creative Commons CC0 1.0 Universal Public Domain Dedication).

18.4.13 Summary

Humans are primates and primates are mammals. Both mammals and primates have distinguishing features, anatomical, physiological, and behavioral. One distinguishing feature of mammals with direct relevance for behavior is that only mammals have six-layered neocortex (cerebral cortex), while non-mammals lack six-layered cortex. Larger brains tend to be found in larger animals. Cerebral cortex developed as a result of natural selection, perhaps related to the increased information processing demands of living in complex social groups, although this is only one theory of the evolutionary origins of cerebral cortex. Brains are metabolically expensive organs--they require a lot of metabolic energy (in humans, the brain is only 2% of body weight, but it consumes 20% of the body's metabolic energy). Larger brains in human ancestors requiring more metabolic energy may have come about after ancestral humans began to eat meat, a concentrated source of calories and nutrients, and later to cook the meat making it easier to digest, leaving more metabolic energy from the meat to support energy requirements of bigger brains. In subsequent modules, we consider more details of hominin and human evolution.

18.4.14 References

Adkins-Regan, E. (2006). Commentary on G.F. Striedter's "Précis of principles of brain evolution." *Behavioral and Brain Sciences*, *29* (1), 12-13.

Bloch JI, Boyer DM. (2002). Grasping primate origins. Science 298(5598): 1606-1610.

Castro J. (2013). How do fossils form? [Internet]. LiveScience [cited 2015 Jul 06]. Available from: http://www.livescience.com/37781-how-do-fossils-form-rocks.html

Jurmain R, Kilgore L, Trevathan W. (2013). Essentials of physical anthropology, 4th edition. Belmont (CA): Wadsworth, Cengage Learning.

Klaus HD. (2011). Non-human primate and human evolution. In: Hutchins M, editor. Grizmek's animal life encyclopedia: evolution. Detroit (MI): Gale; 2011. p. 299-309.

Larsen CS. (2014). Our origins: discovering physical anthropology, 3rd edition. New York (NY): W.W Norton & Company, Inc. 478 p.

Lerner KL, Lerner BW, editors. (2014). Paleoecology. In: The Gale encyclopedia of science, 5th edition, Vol. 6. Farmington Hills (MI): Gale; p. 3217-3218.

O'Neill D. 1998-2014. Early primate evolution. Behavioral Sciences Department, Palomar College [Internet] [cited 2015 Jul 06]. Available from: anthro.palomar.edu/earlyprimates/Default.htm

Panell M. (20214). Fossils and fossilization. In: Lerner KL, Lerner BW, editors. The Gale encyclopedia of science, 5th edition, Vol. 3. Farmington Hills (MI): Gale; p. 1837-1842.

Peppe DJ, Deino AL. (2013). Dating rocks and fossils using geologic methods [Internet]. The Nature Education Knowledge Project [cited 2105 Jul 06]. Available from: http://www.nature.com/scitable/knowledge/library/dating-rocks-and-fossils-using-geologic-methods-107924044

Saneda TM. (2010). Dating techniques. In: Birx HJ, editor. 21st century anthropology: a reference handbook, Vol. 1. Thousand Oaks (CA): SAGE Reference; p. 352-360.

Striedter, G. F. (2006). Précis of principles of brain evolution. *Behavioral and Brain Sciences*, *29* (1), 1-12.

Thies ML. Mammalian characteristics [Internet] [cited 2015 Aug 2]. Available from: http://www.shsu.edu/~bio_mlt/mammals.html

18.4.15 Attributions

"Primate Evolution in Summary" adapted and modified by Kenneth A. Koenigshofer, Ph.D., Chaffey College, from *Explorations: An Open Invitation to Biological Anthropology;* https://explorations.americananthro.....php/chapters/; Chapter 8: Primate Evolution by Jonathan M. G. Perry, Ph.D. and Stephanie L. Canington, B.A.; Editors: Beth Shook, Katie Nelson, Kelsie Aguilera and Lara Braff, American Anthropological Association Arlington, VA 2019; CC BY-NC 4.0 International, except where otherwise noted.

Adapted and modified by Kenneth A. Koenigshofer, Ph.D., Chaffey College, from Biological Anthropology (Saneda and Field), chapters:

Modern Primates by Tori Saneda & Michelle Field via LibreTexts

Primate Ecology by Tori Saneda & Michelle Field via LibreTexts

Primate Evolution by Tori Saneda & Michelle Field via LibreTexts

"Brain Evolution" adapted by Kenneth A. Koenigshofer, Ph.D., Chaffey College, from Striedter (2006).

This page titled 18.4: Chapter 3- Modern Primates and Their Evolution is shared under a mixed license and was authored, remixed, and/or curated by Kenenth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

- **2.3: Modern Primates** by Tori Saneda & Michelle Field is licensed CC BY-SA 4.0. Original source: https://wikieducator.org/Biological_Anthropology.
- **2.5: Primate Ecology** by Tori Saneda & Michelle Field is licensed CC BY-SA 4.0. Original source: https://wikieducator.org/Biological_Anthropology.
- **2.10: Primate Evolution** by Tori Saneda & Michelle Field is licensed CC BY-SA 4.0. Original source: https://wikieducator.org/Biological_Anthropology.

18.5: Chapter 3- Material Culture

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Briefly explain why material culture is important to biological psychologists as well as anthropologists
2. Briefly list the types of tool industries, the date ranges associated with each, and the species of hominin or human associated with each if known

18.5.1 Overview

The earliest evidence of material culture is in the form of stone tools found on sites dated to 2.4 millions years ago (mya). This does not mean that early hominins did not use tools. New finds from Ethiopia indicate that *A. afarensis* used stone tools to extract marrow from bones 3.4 million years ago. What is not known is whether *A. afarensis* was making tools or simply using found rocks as tools. From material culture, biological psychologists and anthropologists can make inferences about cognitive capacities of hominins and early humans.

18.5.2 Material Culture

Harmand et al. (2015) reported that they had found 3.3 million-year-old (myo) stone tools at West Turkana, Kenya, which they proposed calling Lomekwian as the tools predate Oldowan tools (see below) by 700,000 years. What is particularly interesting is that the oldest *Homo* fossils found in West Turkana are 2.34 myo, although *Au. afarensis* is known from 3.39 mya. Questions remain as to which hominin left behind the assemblage (an assemblage is a group of artifacts found together at a specific site) of 149 artifacts, including flake fragments, worked cobbles, and cores, and how it compares with Oldowan tools. Material culture can give clues about cognitive capacities in hominins based on the assumption that more sophisticated tools would require more advanced cognitive abilities.

18.5.3 Oldowan Tool Industry

The oldest stone tool assemblage is the Oldowan tool industry (at least it is the oldest until the field comes to a consensus about Lomekwian tools). First identified at Olduvai Gorge, Tanzania by Louis and Mary Leakey, Oldowan tools are stone pebble tools manufactured using a hard percussion technique. This technique involves striking two stones together to knock off a flake or create an edge on a piece of stone. While this seems like a simple technique, to make one of these tools, the individual needs to be able to understand how the stone will break when struck. The presence of Oldowan tools is an indication of changing cognitive abilities.



Figure 18.5.1: Olduwan flake tool.

Oldowan tools were used for cutting, chopping and scraping.



Figure 18.5.2: Olduwan chopper.

Homo hablis was not reliant primarily on hunting but instead scavenging. Researchers used Oldowan tools to scavenge a carcass and found that in about 10 minutes, it was possible to extract enough meat and bone to meet about 60% of the estimated daily caloric intake (approximately 1500 calories). Furthermore, researchers found that it was easy to scare off other scavengers and even some predators in order to gain access to the carcass. As a subsistence strategy, scavenging does have some advantages:

- it is less dangerous as the scavenger does not have to risk themselves for the kill
- it is quicker as the scavenger can simply follow the roar of the lion or look for vultures circling overhead
- there is less energy expenditure for the reasons listed above and the short amount of time it would take to butcher the remaining carcass using stone tools

The first definitive evidence of hunting is from Schöningen, Germany in the form of wooden spears dated to 400,000 years. The artifacts were identified as spears because they have similar morphology to modern javelins, e.g., the balance point is 1/3 the way from the spear point. The spears are about 7 feet long with sharpened points and were found with the remains of butchered horses.

All researchers agree that *Homo* was making and using tools. Recently reported finds of stone tools 3.3 myo at West Turkana, Kenya may end up demonstrating that an early australopith was the first tool maker.

18.5.4 Acheulean Tool Industry

One thing we see with tool technologies is that as time passes the tools become more and more sophisticated. About 1.9 million years ago, *Homo erectus* invented a new sophisticated technology for making stone tools, which started with the hard percussion technique, but then employed a soft hammer technique to get more refined and sharper edges. This new tool industry is called the Acheulean.

The Acheulean tool industry, first found at St. Acheul, France, is characterized by bifacial tools. This means that the stone is worked on both sides. This tool industry is a marked step in the cognitive abilities of hominins because the tool has to be conceptualized prior to manufacturing. Dozens of flakes have to be removed precisely in order to maintain the symmetry of the tool and keep the edges straight. The signature tool of the Acheulean tool industry is the tear-drop shaped handaxe. Often referred to as the Swiss Army knife of the Pleistocene, the handaxe was an all-purpose tool used for a multitude of activities including digging, sawing, and cutting.

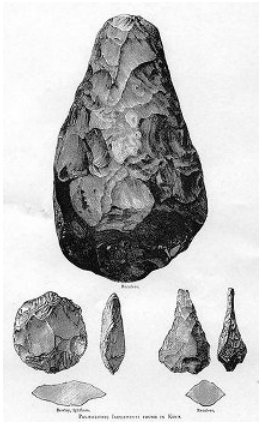


Figure 18.5.3: Acheulean Handaxes.

18.5.5 Mousterian Tool Industry

Neanderthals took the next step in the evolution of stone tools by making tools for specialized tasks. These flake tools found in a cave in France developed out of a manufacturing technique characterized by preparing the core of raw material from which flakes can be struck and then worked. Sharper tools with a finer edge are produced using this technique. Neanderthals shaped these flakes into tools like scrapers, blades, and projectile points, specifically spear points. In fact, at Neanderthal caves sites in the Middle East, there are a higher percentage of spear points found than at neighboring Homo sapiens sites. Mousterian tools are a technological advance, taking a high degree of conceptualization and knowledge of the properties of the stone.



Figure 18.5.4: Mousterian Tool Industry.

18.5.6 Upper Paleolithic Tool Industries

The Upper Paleolithic of Europe begins 45,000 years ago and ushers in further advances in tool technology. Not only are there a wider variety of tools made, but new materials are used, including bone and antler. Blade tools emerge. A blade tool is a tool that is at least twice as long as it is wide. The benefit of blade tool technology is that blades can be easily knocked off a prepared core and then made into a wide range of tools, e.g., projectile points, drills, needles, scrapers, burins. By 31,000 years ago, these tools are widespread throughout Europe, allowing archaeologists to trace the movement of modern *Homo sapiens*.



Figure 18.5.5: Aurignacian Backed Knives (Wellcome M0011849).

This tool industry disappears from the archeological record by 29,000 years ago. It is replaced by another tool industry, which is found at European sites until around 21,000 years ago. This tool industry is characterized by small blades and serrated knives, as well as projectile points that can be hafted onto a shaft. The small size of some of the projectile points leads some archeologists to surmise that the bow and arrow was invented during this period, although the first definitive evidence of arrows comes from Stellmoor, Germany (10,500 years ago). It does appear that the spearthrower was invented during this time frame. This is an important advance as it allowed a hunter to throw farther and with more force, making hunting the large animals of the period a little safer.



Figure 18.5.6: Gravettian tools (Fleche Font Robert 231.4 (2)).



Figure 18.5.7: Solutrean Point.

Solutrean points are some of the best made points of the Upper Paleolithic. The technology flourished from around 21,000 to 16,000 years ago, but then disappears for thousands of years until a similar manufacturing process appears in North America during the Clovis period. To explain this, some archeologists propose that there was a migration of peoples from the Iberian Peninsula to North America in the late Pleistocene who carried the technology with them; however, there is little other evidence to support this contention. It is probable that the manufacturing techniques were rediscovered by North America's early inhabitants.

One of the reasons that Solutrean points were finely made was because the stone was heat treated before it was worked. Heat treatment means that the stone was placed in a fire for a period of time, making it possible to make pressure flaking more precise. Heat treating was also a hallmark of the Magdelenian tool industry, 16,000-11,000 years ago. Bone and antler tools flourish during the Magdelenian. Harpoons appear in the archeological record, with true barbed harpoons showing up around 13,000 years ago.



Figure 18.5.8: Magdalenian Barbed Harpoons.

18.5.7 Why Material Culture is Important to Biological Psychology

For biological psychology some familiarity with the development of tool making is important for what it suggests about the cognitive abilities of various hominin species and early humans. Some of the tools pictured above and the uses that they were put to show sophisticated cognitive abilities to imagine stone tools, how to shape them, and the purposes that they could serve once completed. Evolutionary psychologists such as Cosmides and Tooby from the University of California at Santa Barbara suggest that early humans filled what they refer to as "the cognitive niche," the use of reasoning including use of methods of adaptation

such as weapons, traps, coordinated driving of game, and ability to use abstract concepts and metaphors to enhance reasoning about the world (Pinker, 2010). These abilities depend upon the properties of an evolving brain, specifically on the mechanisms involved in an emerging intelligence in early humans. Brain mechanisms involved are discussed in the chapter in this text on Intelligence, Cognition and Language.

18.5.8 Summary

The development of stone tools undergoes several stages and early stages preceded emergence of *Homo sapiens*. The discovery of stone tools gives clues about the cognitive abilities of hominins including Neanderthal and early humans. Creation of sophisticated stone tools required conceptualization of the tool before its manufacture and suggests emerging abilities to imagine and to plan (see Chapter 14 on Intelligence, Coginition, and Language).

18.5.9 References

1. Harmand S, Lewis JE, Feibel CS, Lepre CJ, Prat S, Lenoble A, Boës X, Quinn RL, Brenet M, Arroyo A, Taylor N, Clément S, Daver G, Brugal JP, Leakey L, Mortlock RA, Wright JD, Lokorodi S, Kirwa C, Kent DV, Roche H. 2015. 3.3-million-year-old tools from Lomekwi 3, West Turkana, Kenya. Nature 521 (May 21): 310-315. Available from: http://www.nature.com.offcampus.lib....ture14464.html. doi:10.1038/nature14464.
2. Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences*, *107* (Supplement 2), 8993-8999.
3. Price, T. Douglas and Gary M. Feinman. 2010. *Images of the Past*, 6th edition. New York: McGraw Hill.
4. Scarre, Chris. 2005. *The Human Past*. London: Thames & Hudson Ltd.
5. Straus, Lawrence Guy. 2005. The Upper Paleolithic of Cantabrian Spain. *Evolutionary Anthropology* 14: 145-158.
6. Villmoare B, Kimbel WH, Seyoum C, Campisano CJ, DiMaggio EN, Rowan J, Braun DR, Arrowsmith JR, Reed KE. 2015. Early Homo at 2.8 MA from Ledi-Geraru, Afar, Ethiopia. Science [Internet]. [cited 2015 Aug 16]; 347(6228): 1352-1355. Available from: http://www.sciencemag.org.offcampus....6228/1352.full. doi: 10.1126/science.aaa1343

18.5.10 Attributions

Text and figures adapted by Kenneth A. Koenigshofer, PhD, from Biological Anthropology (Saneda and Field), chapter: Material Culture via LibreTexts by Tori Saneda & Michelle Field, Professors (Anthropology) at Cascadia Community College (via Wikieducator).

This page titled 18.5: Chapter 3- Material Culture is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

- **3.4: Material Culture** by Tori Saneda & Michelle Field is licensed CC BY-SA 4.0. Original source: https://wikieducator.org/Biological_Anthropology.
- **3.3: Homo Genus** by Tori Saneda & Michelle Field is licensed CC BY-SA 4.0. Original source: https://wikieducator.org/Biological_Anthropology.

18.6: Chapter 10- Computer Models and Learning in Artificial Neural Networks

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

18.6.1 Learning Objectives

1. Explain Hebb's rule and how it is applied in connectionist networks
2. Describe connection weights and how their modification by inputs to them create "learning" in a connectionist network
3. Describe the structure of a simple three-layer network
4. Explain what is meant by "coarse coding" and give an example
5. Explain Hebb's concept of "cell assemblies"
6. Discuss how the graded potentials (EPSPs and IPSPs) in biological neurons undergo a nonlinear transformation during neural communication permitting emergent properties

18.6.2 Overview

Artificial neural networks, or connectionist networks, are computer generated models using principles of computing based loosely on how biological neurons and biological neuron networks operate. Processing units in connectionist networks are frequently arranged in interconnected layers. The connections between processing units in connectionist networks are modifiable and can therefore become stronger or weaker, similar to the way that synapses in the brain are presumed to be modified in strength by learning and experience. Cognitive processes can be studied using artificial neural networks providing insights into how biological brains may code and process information.

18.6.3 Hebbian Synapses, Modification of Connection Weights, and Learning in Artificial Neural Networks

Hebbian Learning, also known as Hebb's Rule or Cell Assembly Theory, attempts to connect the psychological and neurological underpinnings of learning. As discussed in an earlier section of this chapter, the basis of **Hebb's theory** is when our brains learn something new, neurons are activated and connected with other neurons, forming a neural network or cell assembly (a group of synaptically connected neurons) which can represent cognitive structures such as perceptions, concepts, thinking, or memories. How large groups of neurons generate such cognitive structures and processes is unknown. However, in recent years, researchers from neuroscience and computer science have made advances in our understanding of how neurons work together by using computer modeling of brain activity. Not only has computer modeling of brain processes provided new insights about brain mechanisms in learning, memory, and other cognitive functions, but artificial intelligence (AI) research also has powerful practical applications in many areas of human endeavor.

The influence of Hebb's book, *The Organization of Behavior* (1949), has spread beyond psychology and neuroscience to the field of artificial intelligence research using artificial neural networks. An **artificial neural network** is a computer simulation of a "brain-like" system of interconnected processing units, similar to neurons. These processing units, sometimes referred to as "neurodes," mimic many of the properties of biological neurons. For example, "neurodes" in artificial neural networks have connections among them similar to the synapses in biological brains. Just like the strengths of conductivity of synapses between biological neurons can be modified by experience (i.e. learning), the **synaptic "weights"** at connections between "neurodes" can also be **modified by inputs** to the system, permitting **artificial neural networks** to **learn**. The basic idea is that changes in "connection weights" by inputs to an artifical network are analogous to the changes in synaptic strengths between neurons in the brain during learning. Recall from a prior section that experience can strengthen the **associative link between neurons when the activity in those neurons occurs together (i.e. Hebb's Rule**). In artificial neural networks, also called **connectionist networks**, when connection weights between processing units are altered by inputs, the output of the network changes--the artificial neural network learns.

Like biological synapses, the **connections** between processing units in an artificial neural network can be inhibitory as well as excitatory. "The synaptic weight in an artificial neural network is a number that defines the nature and strength of the connection.

For example, **inhibitory connections have negative weights, and excitatory connections have positive weights**. Strong connections have strong weights. The signal sent by one processor to another is a number that is transmitted through the weighted connection or "synapse." The "synaptic" connection and its corresponding "weight" serves as a communication channel between processing units (neurodes) that **amplifies or diminishes signals** being sent through it; the "synaptic" connection amplifies or diminishes the signal because these signals are multiplied by the weight (either positive or negative and of varying magnitude) associated with the "synaptic" connection. Strong connections, either positive or negative, have large (in terms of absolute value disregarding sign) weights and large effects on the signal, while weak connections have near-zero weights" (Libretext, Mind, Body, World: Foundations of Cognitive Science (Dawson), 4.2, Nature vs Nurture). **Changes in synaptic "weights" change** the **output of the neural network** constituting **learning** by the network. Notice the clear **similarity to Hebb's conception of learning in biological brains.**

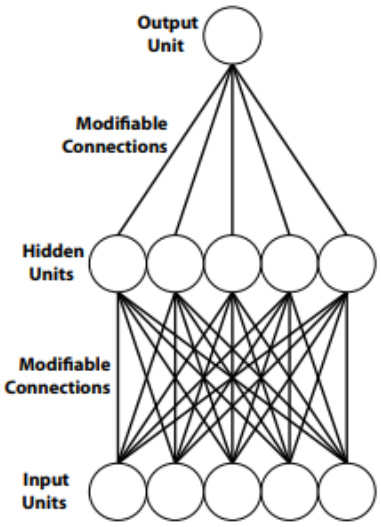


Figure 18.6.1: A multilayer neural network with one layer of hidden units between input and output units. (Image from Libretext, Mind, Body, World: Foundations of Cognitive Science (Dawson), 4.5, The Connectionist Sandwich; https://socialsci.libretexts.org/Boo...onist_Sandwich)

Additional hidden layers increase the complexity of the processing that a network can perform. Artificial neural networks with multiple layers have been shown to be capable of quite complex representations and sophisticated learning and problem-solving. This approach to modeling learning, memory, and cognition is known as "connectionism" (see *Matter and Consciousness* by Paul Churchland, MIT Press, 2013, for an excellent and very readable introduction to the field) and is the basis for much of the research in the growing field of artificial intelligence, including voice recognition and machine vision.

Recall that Hebb (1949) proposed that **if two interconnected neurons are both "on" at the same time, then the synaptic weight between them should be increased**. Thus, Hebbian learning is **a form of activity-dependent synaptic plasticity where correlated activation of pre- and postsynaptic neurons leads to the strengthening of the connection between the two neurons**. "The Hebbian learning rule is one of the earliest and the simplest learning rules for neural networks," "and is based in large part on the dynamics of biological systems. A synapse between two neurons is strengthened when the neurons on either side of the synapse (pre- and post-synaptic) have highly correlated outputs (recall the sections of this chapter on LTP). In essence, when an input neuron fires, if it frequently leads to the firing of the output neuron, the synapse is strengthened." Similarly, in an artificial neural network, the "synaptic" weight or strength between units in an artificial neural network "is increased with high correlation between the firing of the pre-synaptic and post-synaptic" units (adapted from Wikibooks, Artificial Neural Networks/Hebbian Learning, https://en.wikibooks.org/wiki/Artifi...bbian_Learning; retrieved 8/13/2021). Artificial neural networks are not hard wired. Instead, they learn from experience to set the values of their connection weights, just as might occur in biological brains when they learn. Changing connection weights within the network, changes the network's output, its "behavior." This is the physical basis of learning in an artificial neural network (a connectionist network) and is analogous to the changes in the brain believed to be the physical basis of learning in animals and humans.

More complex artificial neural networks of several different types using different learning rules or paradigms have been developed, but regardless of the learning rule employed, each uses modification of synaptic weights (synaptic strengths) to modify outputs as learning by the connectionist network proceeds--a fundamentally Hebbian idea.

In some systems, called self-organizing networks, experience shapes connectivity via "unsupervised learning" (Carpenter & Grossberg, 1992). When learning is unsupervised, networks are only provided with input patterns. **Networks whose connection weights are modified via unsupervised learning develop sensitivity to statistical regularities** in the inputs and organize **their output units** to **reflect these regularities**. In this case, the **connectionist** artificial neural **network can discover new patterns** in data not previously recognized by humans.

In cognitive science, **most networks** reported in the literature are not self-organizing and are not structured via unsupervised learning. Instead, they are networks that are instructed to mediate a desired input-output mapping. This is accomplished via **supervised learning**. In supervised learning, it is assumed that the network has an **external "teacher."** The network is presented with an input pattern and produces a response to it. The teacher compares the response generated by the network to the desired response, usually by calculating the amount of error associated with each output unit. The "teacher" then provides the error as feedback to the network. **A learning rule (such as the delta rule) uses feedback about error to modify connection weights** in the network in such a way that the next time this pattern is presented to the network, the amount of **error** that it produces will be **smaller**. Technically, this learning paradigm is **called "back-propagation of error"** and may be analogous to feedback learning in animals and humans learning to reach a goal (think of operant conditioning, specifically operant shaping).

18.6.4 Representation and Processing of Information in Artifical Networks Gives Clues About Representation and Processing in Biological Brains

One of the assumed advantages of connectionist cognitive science (a branch of psychology, neuroscience, and computer science that attempts to use connectionist systems to model human psychological processes) is that it can inspire alternative notions of representation of information in the brain. How do neurons, populations of neurons, and their operations represent perceptions, mental images, memories, concepts, categories, ideas and other cognitive structures? Coarse coding may be one answer.

A **coarse code** is one in which an individual processing unit is very broadly tuned, sensitive to either a wide range of features or at least to a wide range of values for an individual feature (Churchland & Sejnowski, 1992; Hinton, McClelland, & Rumelhart, 1986). In other words, individual processors are themselves very inaccurate devices for measuring or detecting a feature of the world, such as color. The accurate representation of a feature can become possible, though, by pooling or combining the responses of many such inaccurate detectors, particularly if their perspectives are slightly different (e.g., if they are sensitive to different ranges of features, or if they detect features from different input locations). Here, it is useful to think of auditory receptors, for example, each of which is sensitive to a *range* of frequencies of sound waves. Precision in discrimination of frequency, experienced as pitch, occurs by patterns of activity in auditory receptor cells ("hair cells" on the basilar membrane inside the cochlea of the inner ear) which are sensitive to overlapping ranges of frequencies (auditory receptive fields). Relative firing rates in a group of auditory receptors with overlapping ranges of sensitivity gives a precise neural coding of each particular frequency.

Another familiar example of **coarse coding** is provided by the nineteenth century trichromatic theory of color perception (Helmholtz, 1968; Wasserman, 1978). According to this theory, color perception is mediated by three types of retinal cone receptors. One is maximally sensitive to short (blue) wavelengths of light, another is maximally sensitive to medium (green) wavelengths, and the third is maximally sensitive to long (red) wavelengths. Thus none of these types of receptors are capable of representing, by themselves, the rich rainbow of perceptible hues.

However, these receptors are broadly tuned and have overlapping sensitivities. As a result, most light will activate all three channels simultaneously, but to different degrees. Actual colored light does not produce sensations of absolutely pure color; that red fire engine, for instance, even when completely freed from all admixture of white light, still does not excite those nervous fibers which alone are sensitive to impressions of red, but also, to a very slight degree, those which are sensitive to green, and perhaps to a still smaller extent those which are sensitive to violet rays (Helmholtz, 1968, p. 97).

The study of artificial neural networks has provided biological psychologists and cognitive neuroscientists new insights about how the brain might process information in ways that generate perception, learning and memory, and various forms of cognition including visual recognition and neural representation of spatial maps in the brain (one function of the mammalian hippocampus).

Even critics of connectionism admit that “the study of connectionist machines has led to a number of striking and unanticipated findings; it’s surprising how much computing can be done with a uniform network of simple interconnected elements” (Fodor & Pylyshyn, 1988, p. 6). Hillis (1988, p. 176) has noted that artificial neural networks allow “for the possibility of constructing intelligence without first understanding it.” As one researcher put it, “The major lesson of neural network research, I believe, has been to thus expand our vision of the ways a physical system like the brain might encode and exploit information and knowledge” (Clark, 1997, p. 58).

The **key idea in connectionism is association**: different ideas can be linked together, so that if one arises, then the association between them causes the other to arise as well. Note the similarity to Hebb's concept of **cell assemblies**, formed by associations mediated by changes in synapses, and the association of cell assemblies with one another via the same mechanism. Once again, Hebb's influence is evident in connectionism's artificial neural networks--even the name, connectionism, reflects Hebb's ideas about how cognitive structures and learning and memory are formed in the brain--by changes in connections between neurons. As previously discussed, **long-term potentiation (LTP)** has been extensively studied by biological psychologists and neuroscientists as a biologically plausible mechanism for the synaptic alterations by experience proposed in Hebb’s theory of learning and memory in biological brains.

While association is a fundamental notion in connectionist models, other notions are required by modern connectionist cognitive science in the effort to model complex properties of the mind. One of these additional ideas is **nonlinear processing**. If a system is linear, then its whole behavior is exactly equal to the sum of the behaviors of its parts. **Emergent properties**, however, where the properties of a whole (i.e., a complex idea) are more than the sum of the properties of the parts (for example, a set of associated simpler ideas), **require nonlinear processing**. Complex **cognition, perception, learning and memory may be emergent properties** arising from the activities of the **brain's neurons and networks**. This observation suggests that **nonlinear processing** is a key feature of **how information** is **represented and processed in the brain.** We can see the presence of nonlinear processing at the level of individual neurons.

Neurons demonstrate one powerful type of **nonlinear processing involving action potentials**. As discussed in chapter 5, the inputs to a neuron are weak electrical signals, called **graded potentials** (**EPSPs and IPSPs**; see chapter 5 on Communication within the Nervous System), which stimulate and travel through the dendrites of the receiving neuron. If enough of these weak graded potentials arrive at the neuron’s soma at roughly the same time, then they summate and if their cumulative effect reaches the neuron’s "trigger threshold," a massive depolarization of the membrane of the neuron’s axon, the action potential, occurs. The action potential is a signal of constant intensity that travels along the axon to eventually stimulate some other neuron (see chapter 5). A crucial property of the **action potential** is that it is an **all-or-none** phenomenon, representing **a nonlinear transformation of the summed graded potentials**. The neuron converts continuously varying inputs (EPSPs and IPSPs) into a response that is either on (action potential generated) or off (action potential not generated). This has been called the all-or-none law which states that once an action potential is generated it is always full size, minimizing the possibility that information will be lost as the action potential is conducted along the length of axon. The all-or-none output of neurons is a nonlinear transformation of summed, continuously varying input, and it is the reason that the brain can be described as digital in nature (von Neumann, 1958).

Artificial neural networks have been used to model a dizzying variety of phenomena including animal learning (Enquist & Ghirlanda, 2005), cognitive development (Elman et al., 1996), expert systems (Gallant, 1993), language (Mammone, 1993), pattern recognition and visual perception (Ripley, 1996), and musical cognition (Griffith & Todd, 1999). One neural network even discovered a mathematical proof that human mathematicians had been unable to solve (Churchland, 2013).

Researchers at the University of California at San Diego, and elsewhere in labs around the world, are producing "deep learning" using artificial networks with 5 or more layers that can perform tasks such as facial recognition, object categorization, or speech recognition (see https://cseweb.ucsd.edu//groups/guru/). Although psychologists, PDP researchers, and neuroscientists have a long way to go, modeling of learning and memory, perception, and cognition in artificial (PDP) neural networks has contributed to our understanding of how neurons in the brain's networks may represent, encode and process information during the psychological states that we experience daily, including learning and memory. For example, "The simulation of brain processing by artificial networks suggests that multiple memories can be encoded within a single neural network, by different patterns of synaptic connections. Conversely, a single memory may involve simultaneously activating several different groups of neurons in different parts of the brain" (from The Brain from Top to Bottom; https://thebrain.mcgill.ca/flash/a/a...07_cl_tra.html). As we have seen

earlier in this chapter, neural activity from different regions of cerebral cortex converges onto the hippocampus where processing may combine memory traces from different modalities into unified multi-modal memories.

18.6.5 References

Bliss, T. V., & Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *The Journal of physiology*, *232*(2), 331-356.

Churchland, P. M. (2013). *Matter and consciousness*. Cambridge, MA. MIT press.

Churchland, P.S. & Sejnowski, T.J. (1992). *The Computational Brain*. Cambridge, MA. MIT Press.

Clark, A. (1997). *Being there: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.

Elman, J. L., Bates, E. A., & Johnson, M. H. (1996). *Rethinking innateness: A connectionist perspective on development* (Vol. 10). MIT press.

Enquist, M., & Ghirlanda, S. (2005). *Neural Networks and Animal Behavior* (Vol. 33). Princeton University Press.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1-2), 3-71.

Gallant, S. I., & Gallant, S. I. (1993). *Neural network learning and expert systems*. MIT press.

Griffith, N., Todd, P. M., & Todd, P. M. (Eds.). (1999). *Musical networks: Parallel distributed perception and performance*. MIT Press.

Hayashi-Takagi, A., Yagishita, S., Nakamura, M., Shirai, F., Wu, Y.I., Loshbaugh, A.L., Kuhlman, B., Hahn, KM., Kasai, H. (2015). Labelling and optical erasure of synaptic memory traces in the motor cortex. *Nature 525*, 333–338.

Hebb, D.O. (1949). *The Organization of Behavior*. New York. Wiley.

Helmholtz, H. von, (1968). The recent progress of the theory of vision. *Helmholtz on Perception: Its Physiology and Development, Edited by Richard Warren and Roslyn Warren, John Wiley and Sons, New York, London, Sydney*, 125.

Henry, F. E., Hockeimer, W., Chen, A., Mysore, S. P., & Sutton, M. A. (2017). Mechanistic target of rapamycin is necessary for changes in dendritic spine morphology associated with long-term potentiation. *Molecular Brain*, *10* (1), 1-17.

Hillis, W. D. (1988). Intelligence as an emergent behavior; or, the songs of eden. *Daedalus*, 175-189.

Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Parallel distributed processing.

Kandel, E. (1976). *Cellular Basis of Behavior*. San Francisco. W.H. Freeman and Company.

Kandel, E. R., & Schwartz, J. H. (1982). Molecular biology of learning: Modulation of transmitter release. *Science*, 218 (4571), 433–443

Kolb, B. & Whishaw, I.Q. (2001). *In Introduction to Brain and Behavior*. New York. Worth Publishers.

Kozorovitskiy Y, et al. (2005). Experience induces structural and biochemical changes in the adult primate brain. Proc Natl Acad Sci USA.102:17478. [PMC free article] [PubMed]

Mammone, R. J. (Ed.). (1993). *Artificial neural networks for speech and vision* (Vol. 4). Kluwer Academic Publishers.

Miller, R. R., & Marlin, N. A. (2014). Amnesia following electroconvulsive shock. *Functional disorders of memory*, 143-178. New York. Psychology Press.

Pinel, P.J. & Barnes, S. (2021). *Biopsychology* (11th Edition). Boston. Pearson Education.

Renner MJ, & Rosenzweig MR. *Enriched and Impoverished Environments: Effects on Brain and Behavior*. Springer Verlag; New York: 1987.

Ripley, B.D. (1996) Pattern Recognition and Neural Networks. Cambridge University Press, Cambridge.

Rosenzweig, M. R. (2007). Modification of brain circuits through experience. In *Neural Plasticity and Memory: From Genes to Brain Imaging*. CRC Press/Taylor & Francis, Boca Raton (FL); 2007. PMID: 21204433.

Von Neumann, J. (1958). *The computer and the brain*. New Haven, Conn.: Yale University Press.

Wasserman, G. S. (1978). *Color vision: An historical introduction*. John Wiley & Sons.

18.6.6 Attributions

1. Section 10.9: "Computer Models and Learning in Artificial Neural Networks" by Kenneth A. Koenigshofer, PhD, Chaffey College, licensed CC BY 4.0

2. Parts of the section on "Learning in Artificial Neural Networks," is adapted by Kenneth A. Koenigshofer, PhD, from Libretext, *Mind, Body, World: Foundations of Cognitive Science*, 4.2, Nature vs Nurture; 4.3, Associations; 4.4 Non-linear transformations; 4.5 The Connectionist Sandwich; 4.19 What is Connectionist Cognitive Science; by Michael R. W. Dawson, LibreTexts; licensed under CC BY-NC-ND.

This page titled 18.6: Chapter 10- Computer Models and Learning in Artificial Neural Networks is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

18.7: Chapter 10- Acetylcholine, Enriched Experience, and Memory

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Overview

In this supplementary section, we briefly discuss evidence for the role of the neurotransmitter acetylcholine (ACh) in learning and memory. Some of this evidence involves selective breeding of rats for ability to learn mazes. Rats were tested for their ability to learn mazes. Good maze learners were bred with other good maze learners and poor maze learners were bred with other poor maze learners, generation after generation. After a number of generations of this type of selective breeding, two separate strains of rats were established, one consisting of rats that were good maze learners and a second strain of poor maze learners. Researchers found more whole brain concentrations of ACh, and more ACh metabolism in the cerebral cortex, in rats selectively bred to be good maze learners compared to the rats bred to be poor maze learners. These relationships were interpreted as evidence of an involvement of ACh in learning and memory. Another line of research involved "enriched environments" and their effects on the brain in a variety of species. Small lab animals such as rats and mice are often kept in small wire cages, two animals per cage. In studies using "enriched environments" animals may be kept in large group cages housing together many animals with "toys" such as running wheels, blocks, and other novel or interesting objects. These "enriched environments" not only provide the animals with many things to do, to experience, and to learn about, but these group environments allow much more social interaction than can occur in small cages which house only two animals together. Sometimes, animals are raised in "impoverished environments" in which animals kept by themselves, one animal per cage. Animals raised in "enriched environments" compared to those raised in "impoverished environments" developed thicker and heavier cerebral cortex, greater dendritic branching, more dendritic spines, and other anatomical changes, as well as changes in brain chemistry. Recent research clarifies the mechanisms involved in the effects of ACh on learning and memory. ACh increases the excitability of dendrites of post-synaptic neurons in the CA3 region of the hippocampus involved in memory.

Differences in brain ACh metabolism induced by selective breeding for learning ability

The amnesia and dementia associated with Alzheimer's Disease involve damage to the frontal lobes, the medial temporal lobes, and involve depletion of acetylcholine (ACh) neurotransmitter due to damage in the basal forebrain, located just above the hypothalamus. The basal forebrain is the main source of ACh in the brain (Pinel and Barnes, 2021). Research in animals had long suggested that ACh might be involved in memory.

One way to test for an involvement of ACh in memory is to use selective breeding in animals. If differences in learning ability could be associated with specific differences in brain function, such as ACh activity, these differences might give clues about the physical bases of learning and memory and might contribute to our understanding of some forms of amnesia and dementia.

Rosenzweig and colleagues found chemical differences in the brains of rats selectively bred for maze learning ability (Rosenzweig, 2007). Early experiments focused on the neurotransmitter acetylcholine (ACh). Activity of acetylcholinesterase (AChE), an enzyme which inactivates ACh following its stimulation of cholinergic receptors (post-synaptic receptor sites that receive ACh), was used an indicator of acetylcholine (ACh) metabolism. Comparisons between AChE activity in cerebral cortex in maze-bright and maze-dull rats (rats selectively bred over multiple generations either for superior or inferior maze learning) showed higher cortical AChE activity indicating higher ACh metabolism in the maze-bright strain compared to the maze-dull strain, and within each strain there were significant correlations between behavioral performance and AChE activity in cerebral cortex. In addition, whole brain ACh concentrations were higher in the maze-bright strain than in the maze-dull, suggesting a possible role of ACh synapses in learning and memory. In addition, Rosenzweig and his research group found that AChE activity increases with age in rats up to about 100 days and then declines.

Anatomical and biochemical changes in the brain induced by enriched experience

Additional experiments by Rozensweig and his research group tested the effects of experience on the brain. Rats raised from an early age in an enriched environment, with stimulus objects in a group cage, not only showed differences in AChE activity and increased RNA and protein, but also showed increased weight of cortex and about a 5% increase in thickness of cerebral cortex

compared to rats raised by themselves in individual cages in an impoverished environment. This unexpected finding was early evidence that experience (resulting in learning and memory) can cause anatomical changes in the brain, as well as changes in brain chemistry.

Similar results were found even when older rats were exposed to an enriched environment for as little as 30 days, and several additional studies showed that exposure to enriched environments for as little as 4 days was sufficient to induce cortical weight changes and increased dendritic branching. Later studies in occipital cortex found a 14% increase in glial cells, increased numbers of pyramidal cell bodies, and increased sizes of synaptic junctions, all induced by enriched experience, suggesting possible effects of learning and memory in the brain. These studies also showed increased numbers of dendritic spines in the rats exposed to enriched experience compared to rats in the impoverished environmental condition. As you have already learned, these small spiny structures on dendrites, associated with synapses, undergo transient structural change followed by sustained change in spine volume lasting about 30 minutes or more after stimulation. Renner and Rozensweig (1987) reported that cerebral effects of experience occur in a wide range of species tested, including rats, mice, squirrels, cats, monkeys, fish, and birds. Rosenzweig (2007) reports that Kozorovitskiy, et al. (2005) found that exposure of adult marmosets to an enriched environment in group cages, for just 30 days, "resulted in increases in dendritic spine density, dendritic length, and dendritic complexity of neurons in the hippocampus and the prefrontal cortex, and it raised the expression levels of several synaptic proteins in the same regions."

Acetylcholine Increases Dendritic Excitability in Hippocampus

More recent research is helping to clarify the role of acetylcholine in learning and memory. Humphries, et al. (2022) reported that by inhibiting potassium (K+) ion channels, acetylcholine enhances the excitability of dendrites in the CA3 regions of the hippocampus reducing the number of NMDA receptors that needed to be excited in order to trigger spike activity at NMDA synapses, known to be involved in long-term potentiation (LTP) and long-term memory (see Section 10.4). Thus, these authors propose that "acetylcholine facilitates dendritic integration and NMDA spike generation in selected CA3 dendrites which could strengthen connections between specific CA3 neurons to form memory ensembles" (Humphries, et al., 2022, p. 69).

References

Humphries, R., Mellor, J. R., & O'Donnell, C. (2022). Acetylcholine boosts dendritic NMDA spikes in a CA3 pyramidal neuron model. *Neuroscience*, *489*, 69-83.

Kozorovitskiy, Y., Gross, C. G., Kopil, C., Battaglia, L., McBreen, M., Stranahan, A. M., & Gould, E. (2005). Experience induces structural and biochemical changes in the adult primate brain. *Proceedings of the National Academy of Sciences*, *102*(48), 17478-17482.

Renner MJ, & Rosenzweig MR. *Enriched and Impoverished Environments: Effects on Brain and Behavior*. Springer Verlag; New York: 1987.

Rosenzweig, M. R. (2007). Modification of brain circuits through experience. In *Neural Plasticity and Memory: From Genes to Brain Imaging*. CRC Press/Taylor & Francis, Boca Raton (FL); 2007. PMID: 21204433.

18.7.1 Attributions

Section 10.9, "Supplement 2: Acetylcholine, Enriched Experience, and Memory," written by Kenneth A. Koenigshofer, PhD., Chaffey College, licensed under CC BY 4.0

This page titled 18.7: Chapter 10- Acetylcholine, Enriched Experience, and Memory is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

18.8: Chapter 10- Memory Types and Stages

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

18.8.1 Learning Objectives

1. Summarize main features of the case of Kim Peek and speculate on what might be taking place in his brain to account for his extraordinary memory
2. Compare and contrast explicit and implicit memory, identifying the features that define each.
3. Explain the stages and types of memory and the characteristics of each
4. Summarize the capacities of short-term memory and explain how working memory is used to process information in it
5. Discuss categories, prototypes and schemas, and their relationship to memory
6. Describe counterfactual thinking

18.8.2 Overview

We begin this supplementary section with an example of extraordinary memory in a person with autism. Memory which is so unusual inspires questions about what features of the brain of this individual make his memory so exceptional.

Memory and cognition are the two major interests of cognitive psychologists. The cognitive school was influenced in large part by the development of the electronic computer. Psychologists of this tradition conceptualize memory in terms of types, stages, and processes.

Two types of memory are explicit and implicit. Explicit memory is assessed using measures in which the individual being tested must consciously attempt to remember the information. Explicit memory includes semantic and episodic memory. Explicit memory tests include recall memory tests, recognition memory tests, and measures of relearning (also known as savings).

Implicit memory refers to the influence of experience on behavior, even if the individual is not aware of those influences. Implicit memory is made up of procedural memory, classical conditioning effects, and priming. Priming refers both to the activation of knowledge and to the influence of that activation on behavior. An important characteristic of implicit memories is that they are frequently formed and used automatically, without much effort or awareness on our part.

Stages of memory are sensory, short-term or working memory, and long-term memory. Sensory memory, including iconic and echoic memory, is a memory buffer that lasts only very briefly and then is forgotten, unless it is attended to and passed on for more processing in short-term memory.

Information that we turn our attention to may move into short-term memory (STM). STM is limited in both the length and the amount of information it can hold. Working memory is a set of memory procedures or operations that operates on the information in STM. Working memory’s "central executive" directs the strategies used to keep information in STM, such as maintenance rehearsal, visualization, and chunking.

Long-term memory (LTM) is memory storage that can hold information for days, months, and years. The information that we want to remember in LTM must be encoded and stored, and then retrieved. Some strategies for improving LTM include elaborative encoding, relating information to the self, making use of the forgetting curve and the spacing effect, overlearning, and being aware of context- and state-dependent retrieval effects.

Memories that are stored in LTM are not isolated but rather are linked together into categories and schemas. Schemas are important in part because they help us encode and retrieve information by providing an organizational structure for it.

The ability to maintain information in LTM involves a gradual strengthening of the connections among the neurons in the brain, known as long-term potentiation (LTP). The hippocampus is important in explicit memory, the cerebellum is important in implicit memory, and the amygdala is important in emotional memory. A number of neurotransmitters are important in consolidation and memory. Evidence for the role of different brain structures in different types of memories comes in part from case studies of patients who suffer from amnesia.

Extraordinary Memory

Our memories allow us to do relatively simple things, such as remembering where we parked our car or the name of the current president of the United States, but also allow us to form complex memories, such as how to ride a bicycle or to write a computer program. Moreover, our memories define us as individuals—they are our experiences, our relationships, our successes, and our failures. Without our memories, we would not have a life.

At least for some things, our memory is very good (Bahrick, 2000). Once we learn a face, we can recognize that face many years later. We know the lyrics of many songs by heart, and we can give definitions for tens of thousands of words. Mitchell (2006) contacted participants 17 years after they had been briefly exposed to some line drawings in a lab and found that they still could identify the images significantly better than participants who had never seen them.

For some people, memory is truly amazing. Consider, for instance, the case of Kim Peek, who was the inspiration for the film *Rain Man* (Note 10.1 "Video Clip: Kim Peek"). Although Peek's IQ was only 87, significantly below the average of 100, it is estimated that he memorized more than 10,000 books in his lifetime (Wisconsin Medical Society, n.d.; "Kim Peek," 2004).



Figure 18.8.1: Kim Peek, the subject of the movie Rain Man, was believed to have memorized the contents of more than 10,000 books. He could read a book in about an hour. Source: Photo courtesy of Darold A. Treffert, MD, and the Wisconsin Medical Society, *http://commons.wikimedia.org/wiki/File:Peek1.jpg*.

Video Clip: Kim Peek

(click to see video) *You can view an interview with Kim Peek and see some of his amazing memory abilities at this link.*

In this chapter we will see how psychologists use behavioral responses (such as memory tests and reaction times) to draw inferences about what and how people remember. And we will see that although we have very good memory for some things, our memories are far from perfect (Schacter, 1996). The errors that we make are due to the fact that our memories are not simply recording devices that input, store, and retrieve the world around us. Rather, we actively process and interpret information as we remember and recollect it, and these cognitive processes influence what we remember and how we remember it. Because memories are constructed, not recorded, when we remember events we don't reproduce exact replicas of those events (Bartlett, 1932). People who read the words "*dream*, *sheets*, *rest*, *snore*, *blanket*, *tired*, and *bed*" and then are asked to remember the words often think that they saw the word *sleep* even though that word was not in the list (Roediger & McDermott, 1995). Our cognitive processes influence the accuracy and inaccuracy of our memories and our judgments, and they lead us to be vulnerable to the types of errors that eyewitnesses may make.

In other sections of this chapter we examine the biology of memory including brain areas involved in memory and changes in the brain when a memory is formed. But first we explore the types, processes, and stages of memory.

18.8.3 Memories as Types and Stages

As you can see in Table 10.10.1 "Memory Conceptualized in Terms of Types, Stages, and Processes", psychologists conceptualize memory in terms of *types*, in terms of *stages*, and in terms of *processes*. In this section we will consider the two types of memory, *explicit memory* and *implicit memory*, and then the three major memory stages: *sensory*, *short-term*, and *long-term* (Atkinson &

Shiffrin, 1968). Then, in the next section, we will consider the nature of long-term memory, with a particular emphasis on the three processes that are central to long-term memory: *encoding*, *storage*, and *retrieval*.

As types	Explicit memory
	Implicit memory
As stages	Sensory memory
	Short-term memory
	Long-term memory
As processes	Encoding
	Storage
	Retrieval

Table 10.10.1 Memory Conceptualized in Terms of Types, Stages, and Processes

18.8.4 Explicit Memory

When we assess memory by asking a person to consciously remember things, we are measuring *explicit memory*. **Explicit memory** refers to *knowledge or experiences that can be consciously remembered*. As you can see in Figure 10.4.2, "Types of Memory," there are two types of explicit memory: *episodic* and *semantic*. **Episodic memory** refers to memories of *the firsthand experiences that we have had* (i.e. memories of episodes in one's life; for example, recollections of our high school graduation day or of the fantastic dinner you had in New York last year). **Semantic memory** refers to *our knowledge of facts and concepts about the world* (e.g., that the absolute value of −90 is greater than the absolute value of 9 and that one definition of the word “affect” is “the experience of feeling or emotion”).

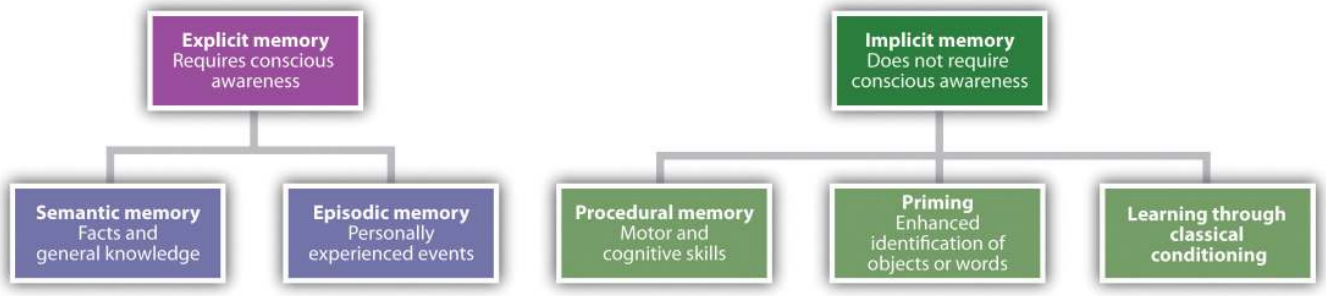


Figure 18.8.2: Types of Memory. Explicit memory is divided into semantic and episodic. Implicit memory includes procedural memory for motor and cognitive skills, priming, and classical conditioning.

Explicit memory is assessed using measures in which the individual being tested must consciously attempt to remember the information. A **recall memory test** is *a measure of explicit memory that involves bringing from memory information that has previously been remembered*. We rely on our recall memory when we take an essay test, because the test requires us to generate previously remembered information. A multiple-choice test is an example of a **recognition memory test**, *a measure of explicit memory that involves determining whether information has been seen or learned before*.

Your own experiences taking tests will probably lead you to agree with the scientific research finding that recall is more difficult than recognition. Recall, such as required on essay tests, involves two steps: first generating an answer and then determining whether it seems to be the correct one. Recognition, as on multiple-choice test, only involves determining which item from a list seems most correct (Haist, Shimamura, & Squire, 1992). Although they involve different processes, recall and recognition memory

measures tend to be correlated. Students who do better on a multiple-choice exam will also, by and large, do better on an essay exam (Bridgeman & Morgan, 1996).

A third way of measuring memory is known as *relearning* (Nelson, 1985). Measures of **relearning (or savings)** *assess how much more quickly information is processed or learned when it is studied again after it has already been learned but then forgotten*. If you have taken some French courses in the past, for instance, you might have forgotten most of the vocabulary you learned. But if you were to work on your French again, you'd learn the vocabulary much faster the second time around. Relearning can be a more sensitive measure of memory than either recall or recognition because it allows assessing memory in terms of "how much" or "how fast" rather than simply "correct" versus "incorrect" responses. Relearning also allows us to measure memory for procedures like driving a car or playing a piano piece, as well as memory for facts and figures.

18.8.5 Implicit Memory

While explicit memory consists of the things that we can consciously report that we know, implicit memory refers to knowledge that we cannot consciously access. However, implicit memory is nevertheless exceedingly important to us because it has a direct effect on our behavior. **Implicit memory** refers to *the influence of experience on behavior, even if the individual is not aware of those influences*. As you can see in Figure 10.4.2, "Types of Memory," there are three general types of implicit memory: procedural (motor) memory, classical conditioning effects, and priming.

Procedural memory refers to *our often unexplainable knowledge of how to do things*. When we walk from one place to another, speak to another person in English, dial a cell phone, or play a video game, we are using procedural memory. Procedural memory allows us to perform complex tasks, even though we may not be able to explain to others how we do them. There is no way to tell someone how to ride a bicycle; a person has to learn by doing it. The idea of implicit memory helps explain how infants are able to learn. The ability to crawl, walk, and talk are procedures, and these motor skills are easily and efficiently developed while we are children despite the fact that as adults we have no conscious memory of having learned them.

A second type of implicit memory is **classical conditioning** effects, in which we learn, often without effort or awareness, to associate neutral stimuli (such as a sound or a light) with another stimulus (such as food), which creates a naturally occurring response, such as enjoyment or salivation. The memory for the association is demonstrated when the conditioned stimulus (the sound of the bell) begins to create the same response as the unconditioned stimulus (the food) did before the **learning of the predictive relation between bell and food, the CS and US.**

The final type of implicit memory is known as **priming**, or *changes in behavior as a result of experiences that have happened frequently or recently*. Priming refers both to the activation of knowledge (e.g., we can prime the concept of "kindness" by presenting people with words related to kindness) and to the influence of that activation on behavior (people who are primed with the concept of kindness may act more kindly).

One measure of the influence of priming on implicit memory is the *word fragment test*, in which a person is asked to fill in missing letters to make words. You can try this yourself: First, try to complete the following word fragments, but work on each one for only three or four seconds. Do any words pop into mind quickly?

_ i b _ a _ y

_ h _ s _ _ i _ n

_ o _ k

_ h _ i s _

Now read the following sentence carefully:

"He got his materials from the shelves, checked them out, and then left the building."

Then try again to make words out of the word fragments.

I think you might find that it is easier to complete fragments 1 and 3 as "library" and "book," respectively, after you read the sentence than it was before you read it. However, reading the sentence didn't really help you to complete fragments 2 and 4 as "physician" and "chaise." This difference in implicit memory probably occurred because as you read the sentence, the concept of

"library" (and perhaps "book") was primed, even though they were never mentioned explicitly. Once a concept is primed it influences our behaviors, for instance, on word fragment tests.

Our everyday behaviors are influenced by priming in a wide variety of situations. Seeing an advertisement for cigarettes may make us start smoking, seeing the flag of our home country may arouse our patriotism, and seeing a student from a rival school may arouse our competitive spirit. And these influences on our behaviors may occur without our being aware of them.

18.8.6 Stages of Memory: Sensory, Short-Term, and Long-Term Memory

Another way of understanding memory is to think about it in terms of stages that describe the length of time that information remains available to us. According to this approach (see Figure 10.4.3), information begins in *sensory memory*, moves to *short-term memory*, and eventually moves to *long-term memory*. But not all information makes it through all three stages; most of it is forgotten. Whether the information moves from shorter-duration memory into longer-duration memory or whether it is lost from memory entirely depends on how the information is attended to and processed.

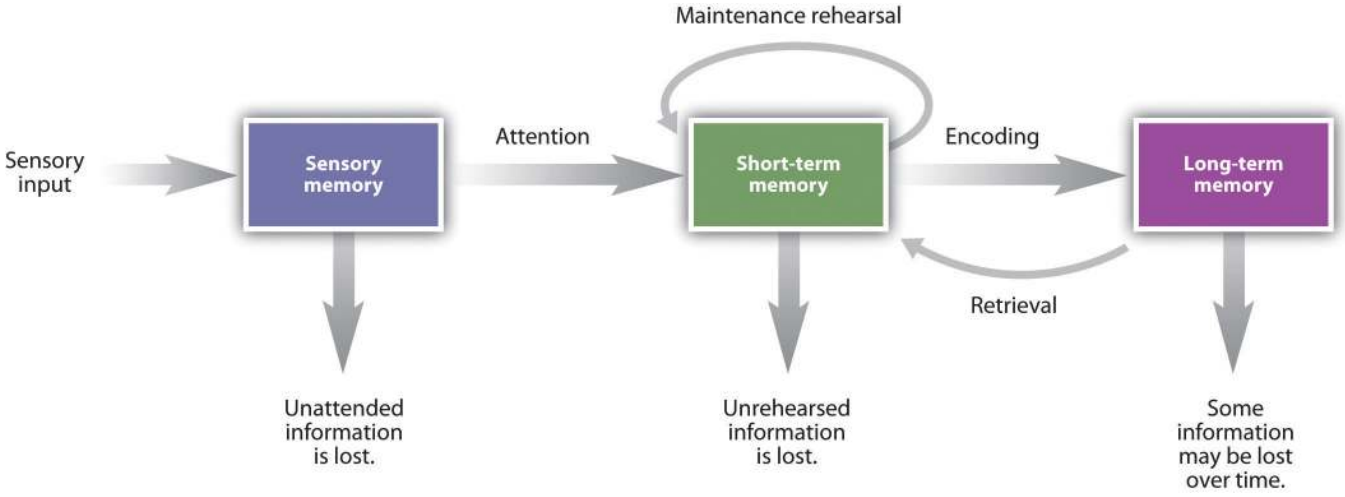


Figure 18.8.3: Memory Duration. Memory can characterized in terms of stages—the length of time that information remains available to us.

Source: Adapted from Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. Spence (Ed.), The psychology of learning and motivation (Vol. 2). Oxford, England: Academic Press.

18.8.7 Sensory Memory

Sensory memory refers to *the brief storage of sensory information*. Sensory memory is a memory buffer that lasts only very briefly and then, unless it is attended to and passed on for more processing, is forgotten. The purpose of sensory memory is to give the brain some time to process the incoming sensations, and to allow us to see the world as an unbroken stream of events rather than as individual pieces.

Visual sensory memory is known as **iconic memory**. Iconic memory was first studied by the psychologist George Sperling (1960). In his research, Sperling showed participants a display of letters in rows, similar to that shown in Figure 10.4.3, below. However, the display lasted only about 50 milliseconds (1/20 of a second). Then, Sperling gave his participants a recall test in which they were asked to name all the letters that they could remember. On average, the participants could remember only about one-quarter of the letters that they had seen.

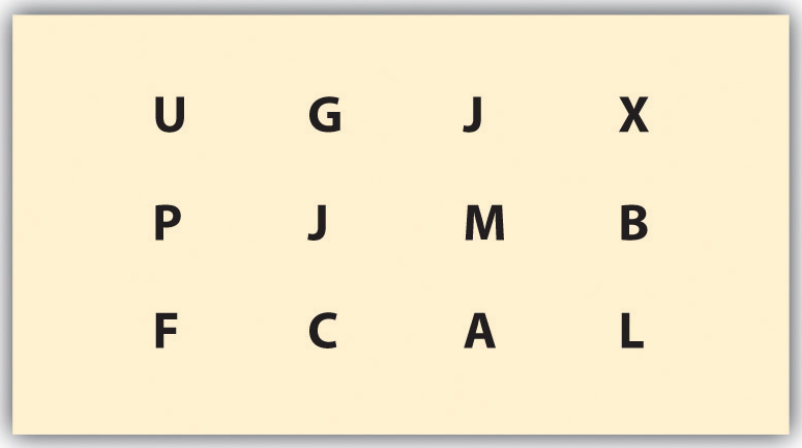


Figure 18.8.4: Measuring Iconic Memory.

Adapted from Sperling, G. (1960). The information available in brief visual presentation. Psychological Monographs, 74(11), 1–29.

Sperling (1960) showed his participants displays such as this one for only 1/20th of a second. He found that when he cued the participants to report one of the three rows of letters, they could do it, even if the cue was given shortly after the display had been removed. The research demonstrated the existence of iconic memory.

Sperling reasoned that the participants had seen all the letters but could remember them only very briefly, making it impossible for them to report them all. To test this idea, in his next experiment he first showed the same letters, but then *after the display had been removed*, he signaled to the participants to report the letters from either the first, second, or third row. In this condition, the participants now reported almost all the letters in that row. This finding confirmed Sperling's hunch: Participants had access to all of the letters in their iconic memories, and if the task was short enough, they were able to report on the part of the display he asked them to. The "short enough" is the length of iconic memory, which turns out to be about 250 milliseconds (¼ of a second).

Auditory sensory memory is known as **echoic memory**. In contrast to iconic memories, which decay very rapidly, echoic memories can last as long as 4 seconds (Cowan, Lichty, & Grove, 1990). This is convenient as it allows you—among other things—to remember the words that you said at the beginning of a long sentence when you get to the end of it, and to take notes on your psychology professor's most recent statement even after he or she has finished saying it.

In some people iconic memory seems to last longer, a phenomenon known as *eidetic imagery* (or "photographic memory") in which people can report details of an image over long periods of time. These people, who often suffer from psychological disorders such as autism, claim that they can "see" an image long after it has been presented, and can often report accurately on that image. There is also some evidence for eidetic memories in hearing; some people report that their echoic memories persist for unusually long periods of time. The composer Wolfgang Amadeus Mozart may have possessed eidetic memory for music, because even when he was very young and had not yet had a great deal of musical training, he could listen to long compositions and then play them back almost perfectly (Solomon, 1995).

18.8.8 Short-Term Memory

Most of the information that gets into sensory memory is forgotten, but information that we turn our attention to, with the goal of remembering it, may pass into *short-term memory*. **Short-term memory (STM)** is *the place where small amounts of information can be temporarily kept for more than a few seconds but usually for less than one minute* (Baddeley, Vallar, & Shallice, 1990). Information in short-term memory is not stored permanently. Although often conceptualized as an intermediate stage between snesory memory and long-term memory, it is also used when we are working on a task, such as doing a long division problem, where the task requires that we keep some itmes of information in mind. For this reason, it is often called **working memory**.

Although it is called "memory," working memory is not a store of memory like STM but rather a set of memory procedures or operations. Imagine, for instance, that you are asked to participate in a task such as this one, which is a measure of working

memory (Unsworth & Engle, 2007). Each of the following questions appears individually on a computer screen and then disappears after you answer the question:

Is 10 × 2 − 5 = 15? (Answer YES OR NO) Then remember “S”
Is 12 ÷ 6 − 2 = 1? (Answer YES OR NO) Then remember “R”
Is 10 × 2 = 5? (Answer YES OR NO) Then remember “P”
Is 8 ÷ 2 − 1 = 1? (Answer YES OR NO) Then remember “T”
Is 6 × 2 − 1 = 8? (Answer YES OR NO) Then remember “U”
Is 2 × 3 − 3 = 0? (Answer YES OR NO) Then remember “Q”

To successfully accomplish the task, you have to answer each of the math problems correctly and at the same time remember the letter that follows the task. Then, after the six questions, you must list the letters that appeared in each of the trials in the correct order (in this case S, R, P, T, U, Q).

To accomplish this difficult task you need to use a variety of skills. You clearly need to use STM, as you must keep the letters in storage until you are asked to list them. But you also need a way to make the best use of your available attention and processing. For instance, you might decide to use a strategy of “repeat the letters twice, then quickly solve the next problem, and then repeat the letters twice again including the new one.” Keeping this strategy (or others like it) going is the role of working memory’s *central executive*—the part of working memory that directs attention and processing. The central executive will make use of whatever strategies seem to be best for the given task. For instance, the central executive will direct the rehearsal process, and at the same time direct the visual cortex to form an image of the list of letters in memory. You can see that although STM is involved, the processes that we use to operate on the material in memory are also critical.

Short-term memory is limited in both the length and the amount of information it can hold. Peterson and Peterson (1959) found that when people were asked to remember a list of three-letter strings and then were immediately asked to perform a distracting task (counting backward by threes), the material was quickly forgotten (see Figure 10.4.4), such that by 18 seconds it was virtually gone.

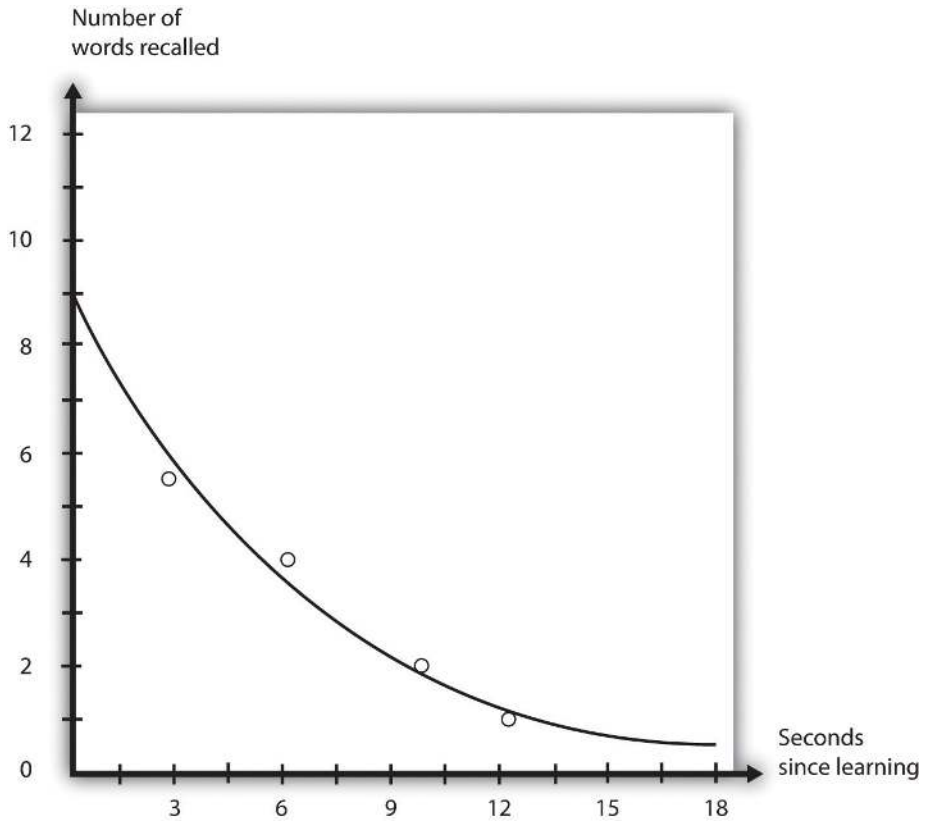


Figure 18.8.5: . STM Decay. Peterson and Peterson (1959) found that information that was not rehearsed decayed quickly from memory. Source: Adapted from Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. Journal of Experimental Psychology, 58(3), 193–198.

One way to prevent the decay of information from short-term memory is to use working memory to rehearse it. **Maintenance rehearsal** is *the process of repeating information mentally or out loud with the goal of keeping it in memory*. We engage in maintenance rehearsal to keep a something that we want to remember (e.g., a person's name, e-mail address, or phone number) in mind long enough to write it down, use it, or potentially transfer it to long-term memory.

If we continue to rehearse information it will stay in STM until we stop rehearsing it, but there is also a capacity limit to STM. Try reading each of the following rows of numbers, one row at a time, at a rate of about one number each second. Then when you have finished each row, close your eyes and write down as many of the numbers as you can remember.

019

3586

10295

861059

1029384

75674834

657874104

6550423897

If you are like the average person, you will have found that on this test of working memory, known as a *digit span test*, you did pretty well up to about the fourth line, and then you started having trouble. I bet you missed some of the numbers in the last three rows, and did pretty poorly on the last one.

The digit span of most adults is between five and nine digits, with an average of about seven. The cognitive psychologist George Miller (1956) referred to “seven plus or minus two” pieces of information as the “magic number” in short-term memory. But if we can only hold a maximum of about nine digits in short-term memory, then how can we remember larger amounts of information than this? For instance, how can we ever remember a 10-digit phone number long enough to dial it?

One way we are able to expand our ability to remember things in STM is by using a memory technique called *chunking*. **Chunking** is *the process of organizing information into smaller groupings (chunks), thereby increasing the number of items that can be held in STM*. For instance, try to remember this string of 12 letters:

XOFCBANNCVTM

You probably won’t do that well because the number of letters is more than the magic number of seven.

Now try again with this one:

MTVCNNABCFOX

Would it help you if I pointed out that the material in this string could be chunked into four sets of three letters each? I think it would, because then rather than remembering 12 letters, you would only have to remember the names of four television stations. In this case, chunking changes the number of items you have to remember from 12 to only four. This works because you are already very familiar with the letter triplets that are used to refer to these well known TV networks. That knowledge which is already fixed in your memory can be brought to the task at hand to make the memory task easier.

Experts rely on chunking to help them process complex information. Herbert Simon and William Chase (1973) showed chess masters and chess novices various positions of pieces on a chessboard for a few seconds each. The experts did a lot better than the novices in remembering the positions because they were able to see the “big picture.” They didn’t have to remember the position of each of the pieces individually, but chunked the pieces into several larger layouts. But when the researchers showed both groups random chess positions—positions that would be very unlikely to occur in real games—both groups did equally poorly, because in this situation the experts lost their ability to organize the layouts (see Figure 10.4.6. "Possible and Impossible Chess Positions," below). The same occurs for basketball. Basketball players recall actual basketball positions much better than do nonplayers, but only when the positions make sense in terms of what is happening on the court, or what is likely to happen in the near future, and thus can be chunked into bigger units (Didierjean & Marmèche, 2005).

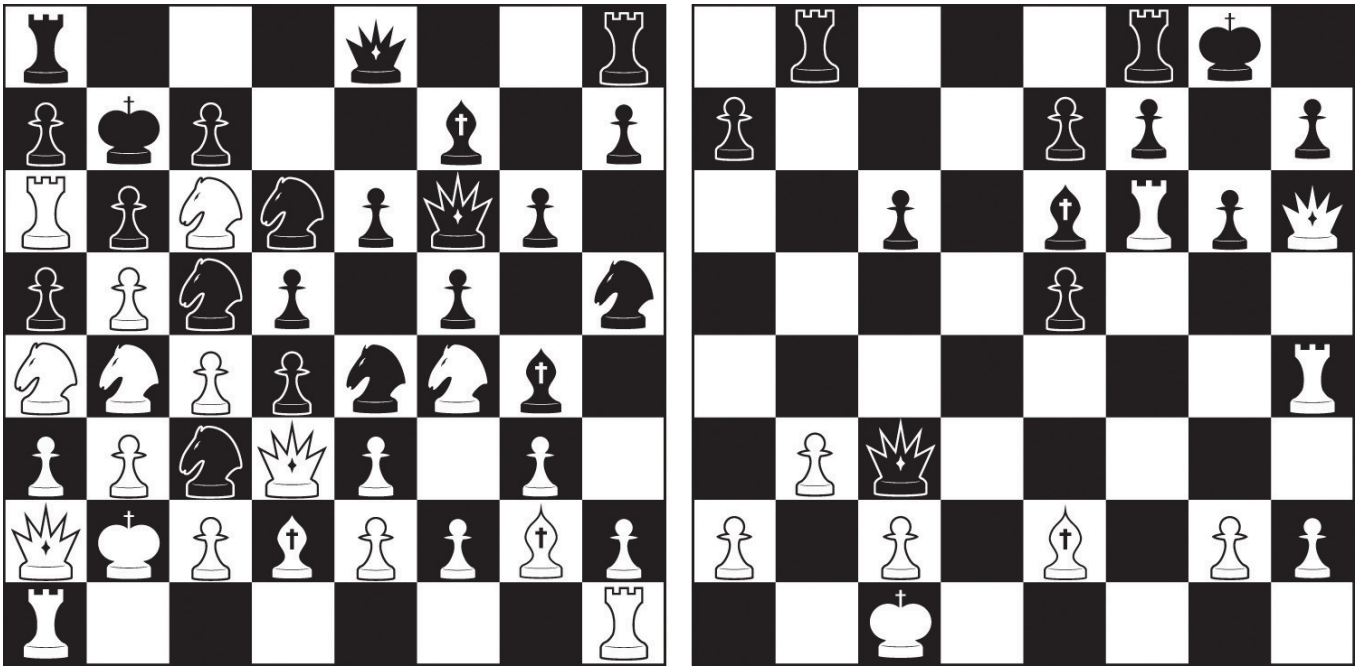


Figure 18.8.6: Possible and Impossible Chess Positions. Experience matters: Experienced chess players are able to recall the positions of the game on the right much better than are those who are chess novices. But the experts do no better than the novices in remembering the positions on the left, which cannot occur in a real game.

18.8.9 Long-Term Memory

If information makes it past short term-memory it may enter **long-term memory (LTM)**, *memory storage that can hold information for days, months, and years*. The capacity of long-term memory is large, and there is no known limit to what we can remember (Wang, Liu, & Wang, 2003). Although we may forget at least some information after we learn it, other things will stay with us forever. In the next section we will discuss the principles of long-term memory.

18.8.9.1 KEY TAKEAWAYS

- Memory refers to the ability to store and retrieve information over time.
- For some things our memory is very good, but our active cognitive processing of information assures that memory is never an exact replica of what we have experienced.
- Explicit memory refers to experiences that can be intentionally and consciously remembered, and it is measured using recall, recognition, and relearning. Explicit memory includes episodic and semantic memories.
- Measures of relearning (also known as savings) assess how much more quickly information is learned when it is studied again after it has already been learned but then forgotten.
- Implicit memory refers to the influence of experience on behavior, even if the individual is not aware of those influences. The three types of implicit memory are procedural memory, classical conditioning, and priming.
- Information processing begins in sensory memory, moves to short-term memory, and eventually moves to long-term memory.
- Maintenance rehearsal and chunking are used to keep information in short-term memory.
- The capacity of long-term memory is large, and there is no known limit to what we can remember.

18.8.9.2 EXERCISES AND CRITICAL THINKING

1. List some situations in which sensory memory is useful for you. What do you think your experience of the stimuli would be like if you had no sensory memory?
2. Describe a situation in which you need to use working memory to perform a task or solve a problem. How do your working memory skills help you?

18.8.10 References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. Spence (Ed.), *The psychology of learning and motivation* (Vol. 2). Oxford, England: Academic Press.

Bahrick, H. P. (2000). Long-term maintenance of knowledge. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 347–362). New York, NY: Oxford University Press.

Bartlett, F. C. (1932). *Remembering*. Cambridge, MA: Cambridge University Press.

Luria, A. (2004). *The mind of a mnemonist: A little book about a vast memory.* Cambridge, MA: Harvard University Press.

Mitchell, D. B. (2006). Nonconscious priming after 17 years: Invulnerable implicit memory? *Psychological Science, 17* (11), 925–928.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803–814.

Schacter, D. L. (1996). *Searching for memory: The brain, the mind, and the past* (1st ed.). New York, NY: Basic Books.

Wisconsin Medical Society. (n.d.). Retrieved from http://www.wisconsinmedicalsociety.org/_SAVANT/_PROFILES/kim_peek/_media/video/expedition/video.html

18.8.11 Additional References

Baddeley, A. D., Vallar, G., & Shallice, T. (1990). The development of the concept of working memory: Implications and contributions of neuropsychology. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory* (pp. 54–73). New York, NY: Cambridge University Press.

Bridgeman, B., & Morgan, R. (1996). Success in college for students with discrepancies between performance on multiple-choice and essay tests. *Journal of Educational Psychology, 88* (2), 333–340.

Cowan, N., Lichty, W., & Grove, T. R. (1990). Properties of memory for unattended spoken syllables. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 258–268.

Didierjean, A., & Marmèche, E. (2005). Anticipatory representation of visual basketball scenes by novice and expert players. *Visual Cognition*, *12*(2), 265–283.

Haist, F., Shimamura, A. P., & Squire, L. R. (1992). On the relationship between recall and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18* (4), 691–702.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97.

Nelson, T. O. (1985). Ebbinghaus's contribution to the measurement of retention: Savings during relearning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11* (3), 472–478.

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*(3), 193–198.

Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist*, *61* (4), 394–403.

Solomon, M. (1995). *Mozart: A life*. New York, NY: Harper Perennial.

Sperling, G. (1960). The information available in brief visual presentation. *Psychological Monographs*, *74* (11), 1–29. I

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133* (6), 1038–1066.

Wang, Y., Liu, D., & Wang, Y. (2003). Discovering the capacity of human memory. *Brain & Mind*, *4*(2), 189–198.

18.8.11.1 The Structure of LTM: Categories, Prototypes, and Schemas

Memories that are stored in LTM are not isolated but rather are linked together into **categories**—*networks of associated memories that have features in common with each other*. Forming categories, and using categories to guide behavior, is a fundamental part of human nature. Associated concepts within a category are connected through *spreading activation*, which occurs when activating one element of a category activates other associated elements. For instance, because tools are associated in a category, reminding people of the word "screwdriver" will help them remember the word "wrench." And, when people have learned lists of words that come from different categories, they do not recall the information haphazardly. If they have just remembered the word "wrench," they are more likely to remember the word "screwdriver" next than they are to remember the word "dahlia," because the words are organized in memory by category and because "dahlia" is activated by spreading activation from "wrench" (Srull & Wyer, 1989).

Some categories have *defining features* that must be true of all members of the category. For instance, all members of the category "triangles" have three sides, and all members of the category "birds" lay eggs. But most categories are not so well-defined; the members of the category share some common features, but it is impossible to define which are or are not members of the category. For instance, there is no clear definition of the category "tool." Some examples of the category, such as a hammer and a wrench, are clearly and easily identified as category members, whereas other members are not so obvious. Is an ironing board a tool? What about a car?

Members of categories (even those with defining features) can be compared to the category **prototype**, which is *the member of the category that is most average or typical of the category*. Some category members are more prototypical of, or similar to, the category than others. For instance, some category members (robins and sparrows) are highly prototypical of the category "birds," whereas other category members (penguins and ostriches) are less prototypical. We retrieve information that is prototypical of a category faster than we retrieve information that is less prototypical (Rosch, 1975).

Mental categories are sometimes referred to as **schemas**—*patterns of knowledge in long-term memory that help us organize information*. We have schemas about objects (that a triangle has three sides and may take on different angles), about people (that Sam is friendly, likes to golf, and always wears sandals), about events (the particular steps involved in ordering a meal at a restaurant), and about social groups (we call these group schemas *stereotypes*).

Schemas are important in part because they help us remember new information by providing an organizational structure for it. Read the following paragraph (Bransford & Johnson, 1972) and then try to write down everything you can remember.

> The procedure is actually quite simple. First you arrange things into different groups. Of course, one pile may be sufficient depending on how much there is to do. If you have to go somewhere else due to lack of facilities, that is the next step; otherwise you are pretty well set. It is important not to overdo things. That is, it is better to do too few things at once than too many. In the short run this may not seem important, but complications can easily arise. A mistake can be expensive as well. At first the whole procedure will seem complicated. Soon, however, it will become just another facet of life. It is difficult to foresee any end to the necessity for this task in the immediate future, but then one never can tell. After the procedure is completed, one arranges the materials into different groups again. Then they can be put into their appropriate places. Eventually they will be used once more and the whole cycle will then have to be repeated. However, that is part of life.

It turns out that people's memory for this information is quite poor, unless they have been told ahead of time that the information describes "doing the laundry," in which case their memory for the material is much better. This demonstration of the role of schemas in memory shows how our existing knowledge can help us organize new information, and how this organization can improve encoding, storage, and retrieval.

18.8.11.2 References

Bransford, J. D. & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning & Verbal Behavior, 11* (6), 717–726.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104* (3), 192–233.

Srull, T. & Wyer, R. (1989). Person memory and judgment. *Psychological Review, 96* (1), 58–83.

18.8.12 Counterfactual Thinking

In addition to influencing our judgments about ourselves and others, the ease with which we can retrieve potential experiences from memory can have an important effect on our own emotions. If we can easily imagine an outcome that is better than what actually happened, then we may experience sadness and disappointment; on the other hand, if we can easily imagine that a result might have been worse than what actually happened, we may be more likely to experience happiness and satisfaction. *The tendency to think about and experience events according to "what might have been"* is known as **counterfactual thinking** (Kahneman & Miller, 1986; Roese, 2005).

Imagine, for instance, that you were participating in an important contest, and you won the silver (second-place) medal. How would you feel? Certainly you would be happy that you won the silver medal, but wouldn't you also be thinking about what might have happened if you had been just a little bit better—you might have won the gold medal! On the other hand, how might you feel if you won the bronze (third-place) medal? If you were thinking about the *counterfactuals* (the "what might have beens") perhaps the idea of not getting any medal at all would have been highly accessible; you'd be happy that you got the medal that you did get, rather than coming in fourth.



Figure 18.8.7: Counterfactual Thinking. Does the bronze medalist look happier to you than the silver medalist? Medvec, V. H., Madey, S. F., & Gilovich, T. (1995). When less is more: Counterfactual thinking and satisfaction among Olympic medalists. Journal of Personality & Social Psychology, 69(4), 603–610. found that, on average, bronze medalists were happier. Source: Photo courtesy of kinnigurl, http://commons.wikimedia.org/wiki/File:2010_Winter_Olympic_Men%27s_Snowboard_Cross_medalists.jpg.

Tom Gilovich and his colleagues (Medvec, Madey, & Gilovich, 1995) investigated this idea by videotaping the responses of athletes who won medals in the 1992 Summer Olympic Games. They videotaped the athletes both as they learned that they had won a silver or a bronze medal and again as they were awarded the medal. Then the researchers showed these videos, without any sound, to raters who did not know which medal which athlete had won. The raters were asked to indicate how they thought the athlete was feeling, using a range of feelings from “agony” to “ecstasy.” The results showed that the bronze medalists were, on average, rated as happier than were the silver medalists. In a follow-up study, raters watched interviews with many of these same athletes as they talked about their performance. The raters indicated what we would expect on the basis of counterfactual thinking—the silver medalists talked about their disappointments in having finished second rather than first, whereas the bronze medalists focused on how happy they were to have finished third rather than fourth.

You might have experienced counterfactual thinking in other situations. Once I was driving across country, and my car was having some engine trouble. I really wanted to make it home when I got near the end of my journey; I would have been extremely disappointed if the car broke down only a few miles from my home. Perhaps you have noticed that once you get close to finishing something, you feel like you really need to get it done. Counterfactual thinking has even been observed in juries. Jurors who were asked to award monetary damages to others who had been in an accident offered them substantially more in compensation if they barely avoided injury than they offered if the accident seemed inevitable (Miller, Turnbull, & McFarland, 1988).

18.8.13 References

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*, 136–153; Roese, N. (2005). *If only: How to turn regret into opportunity*. New York, NY: Broadway Books.

Medvec, V. H., Madey, S. F., & Gilovich, T. (1995). When less is more: Counterfactual thinking and satisfaction among Olympic medalists. *Journal of Personality & Social Psychology, 69* (4), 603–610.

Miller, D. T., Turnbull, W., & McFarland, C. (1988). Particularistic and universalistic evaluation in the social comparison process. *Journal of Personality and Social Psychology, 55*, 908–917.

18.8.14 Attributions

18.8.14.1 Licensing Information

This module, 10.4, was adapted by Saylor Academy under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License without attribution as requested by the work's original creator or licensor.



Saylor Academy would like to thank Andy Schmitz for his work in maintaining and improving the HTML versions of these textbooks. This textbook is adapted from his HTML version, and his project can be found here.

How to cite this work:

- Publisher: Saylor Academy
- Year Published: 2012

This page titled 18.8: Chapter 10- Memory Types and Stages is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

18.9: Chapter 10- The Psychology of Memory Encoding, Storage, and Retrieval

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

18.9.0.1 Memory (Encoding, Storage, Retrieval)

By Kathleen B. McDermott and Henry L. Roediger III, Washington University in St. Louis

(Adapted and Modified by Kenneth A. Koenigshofer, PhD., Chaffey College)

“Memory” is a single term that reflects a number of different abilities: holding information briefly while working with it (working memory), remembering episodes of one’s life (episodic memory), and our general knowledge of facts of the world (semantic memory), among other types. Remembering episodes involves three processes: encoding information (learning it, by perceiving it and relating it to past knowledge), storing it (maintaining it over time), and then retrieving it (accessing the information when needed). Failures can occur at any stage, leading to forgetting or to having false memories. The key to improving one’s memory is to improve processes of encoding and to use techniques that guarantee effective retrieval. Good encoding techniques include relating new information to what one already knows, forming mental images, and creating associations among information that needs to be remembered. The key to good retrieval is developing effective cues that will lead the rememberer back to the encoded information. Classic mnemonic systems, known since the time of the ancient Greeks and still used by some today, can greatly improve one’s memory abilities.

Learning Objectives

1. Define and note differences between the following forms of memory: working memory, episodic memory, semantic memory, collective memory.
2. Describe the three stages in the process of learning and remembering.
3. Describe strategies that can be used to enhance the original learning or encoding of information.
4. Describe strategies that can improve the process of retrieval.
5. Describe why the classic mnemonic device, the method of loci, works so well.

18.9.1 Introduction

In 2013, Simon Reinhard sat in front of 60 people in a room at Washington University, where he memorized an increasingly long series of digits. On the first round, a computer generated 10 random digits—6 1 9 4 8 5 6 3 7 1—on a screen for 10 seconds. After the series disappeared, Simon typed them into his computer. His recollection was perfect. In the next phase, 20 digits appeared on the screen for 20 seconds. Again, Simon got them all correct. No one in the audience (mostly professors, graduate students, and undergraduate students) could recall the 20 digits perfectly. Then came 30 digits, studied for 30 seconds; once again, Simon didn’t misplace even a single digit. For a final trial, 50 digits appeared on the screen for 50 seconds, and again, Simon got them all right. In fact, Simon would have been happy to keep going. His record in this task—called “forward digit span”—is 240 digits!



Figure **18.9.1**: In some ways memory is like file drawers where you store mental information. Memory is also a series of processes: how does that information get filed to begin with and how does it get retrieved when needed? [Image: M Cruz, https://goo.gl/DhOMgp, CC BY-SA 4.0, https://goo.gl/SWjq94]

When most of us witness a performance like that of Simon Reinhard, we think one of two things: First, maybe he's cheating somehow. (No, he is not.) Second, Simon must have abilities more advanced than the rest of humankind. After all, psychologists established many years ago that the normal memory span for adults is about 7 digits, with some of us able to recall a few more and others a few less (Miller, 1956). That is why the first phone numbers were limited to 7 digits—psychologists determined that many errors occurred (costing the phone company money) when the number was increased to even 8 digits. But in normal testing, no one gets 50 digits correct in a row, much less 240. So, does Simon Reinhard simply have a photographic memory? He does not. Instead, Simon has taught himself simple strategies for remembering that have greatly increased his capacity for remembering virtually any type of material—digits, words, faces and names, poetry, historical dates, and so on. Twelve years earlier, before he started training his memory abilities, he had a digit span of 7, just like most of us. Simon has been training his abilities for about 10 years as of this writing, and has risen to be in the top two of "memory athletes." In 2012, he came in second place in the World Memory Championships (composed of 11 tasks), held in London. He currently ranks second in the world, behind another German competitor, Johannes Mallow. In this module, we reveal what psychologists and others have learned about memory, and we also explain the general principles by which you can improve your own memory for factual material.

18.9.2 Varieties of Memory



Figure **18.9.2**: To be a good chess player you have to learn to increase working memory so you can plan ahead for several offensive moves while simultaneously anticipating - through use of memory - how the other player could counter each of your planned moves. [Image: karpidis, https://goo.gl/EhzMKM, CC BY-SA 2.0, https://goo.gl/jSSrcO]

For most of us, remembering digits relies on *short-term memory*, or *working memory*—the ability to hold information in our minds for a brief time and work with it (e.g., multiplying 24 x 17 without using paper would rely on working memory). Another type of memory is **episodic memory**—the ability to remember the episodes of our lives. If you were given the task of recalling everything you did 2 days ago, that would be a test of episodic memory; you would be required to mentally travel through the day in your mind and note the main events. **Semantic memory** is our storehouse of more-or-less permanent knowledge, such as the meanings of words in a language (e.g., the meaning of "parasol") and the huge collection of facts about the world (e.g., there are 196 countries in the world, and 206 bones in your body). *Collective memory* refers to the kind of memory that people in a group share (whether family, community, schoolmates, or citizens of a state or a country). For example, residents of small towns often strongly identify with those towns, remembering the local customs and historical events in a unique way. That is, the community's collective memory passes stories and recollections between neighbors and to future generations, forming a memory system unto itself.

Psychologists continue to debate the classification of types of memory, as well as which types rely on others (Tulving, 2007), but for this module we will focus on episodic memory. Episodic memory is usually what people think of when they hear the word "memory." For example, when people say that an older relative is "losing her memory" due to Alzheimer's disease, the type of memory-loss they are referring to is the inability to recall events, or episodic memory. (Semantic memory is actually preserved in early-stage Alzheimer's disease.) Although remembering specific events that have happened over the course of one's entire life (e.g., your experiences in sixth grade) can be referred to as **autobiographical memory**, we will focus primarily on the episodic memories of more recent events.

18.9.3 Three Stages of the Learning/Memory Process

Psychologists distinguish between three necessary stages in the learning and memory process: **encoding**, **storage**, and **retrieval** (Melton, 1963). Encoding is defined as the initial learning of information; storage refers to maintaining information over time; retrieval is the ability to access information when you need it. If you meet someone for the first time at a party, you need to encode

her name (Lyn Goff) while you associate her name with her face. Then you need to maintain the information over time. If you see her a week later, you need to recognize her face and have it serve as a cue to retrieve her name. Any successful act of remembering requires that all three stages be intact. However, two types of errors can also occur. Forgetting is one type: you see the person you met at the party and you cannot recall her name. The other error is misremembering (false recall or false recognition): you see someone who looks like Lyn Goff and call the person by that name (false recognition of the face). Or, you might see the real Lyn Goff, recognize her face, but then call her by the name of another woman you met at the party (misrecall of her name).

Whenever forgetting or misremembering occurs, we can ask, at which stage in the learning/memory process was there a failure?—though it is often difficult to answer this question with precision. One reason for this inaccuracy is that the three stages are not as discrete as our description implies. Rather, all three stages depend on one another. How we encode information determines how it will be stored and what cues will be effective when we try to retrieve it. And too, the act of retrieval itself also changes the way information is subsequently remembered, usually aiding later recall of the retrieved information. The central point for now is that the three stages—encoding, storage, and retrieval—affect one another, and are inextricably bound together.

18.9.4 Encoding

Encoding refers to the initial experience of perceiving and learning information. Psychologists often study recall by having participants study a list of pictures or words. Encoding in these situations is fairly straightforward. However, “real life” encoding is much more challenging. When you walk across campus, for example, you encounter countless sights and sounds—friends passing by, people playing Frisbee, music in the air. The physical and mental environments are much too rich for you to encode all the happenings around you or the internal thoughts you have in response to them. So, an important first principle of encoding is that it is selective: we attend to some events in our environment and we ignore others. A second point about encoding is that it is prolific; we are always encoding the events of our lives—attending to the world, trying to understand it. Normally this presents no problem, as our days are filled with routine occurrences, so we don’t need to pay attention to everything. But if something does happen that seems strange—during your daily walk across campus, you see a giraffe—then we pay close attention and try to understand why we are seeing what we are seeing.



Figure **18.9.3**: A giraffe in the context of a zoo or its natural habitat may register as nothing more than ordinary, but put it in another setting - in the middle of a campus or a busy city - and its level of distinctiveness increases dramatically. Distinctiveness is a key attribute to remembering events. [Image: Colin J Babb, https://goo.gl/Cci2yl, CC BY-SA 2.0, https://goo.gl/jSSrcO]

Right after your typical walk across campus (one without the appearance of a giraffe), you would be able to remember the events reasonably well if you were asked. You could say whom you bumped into, what song was playing from a radio, and so on. However, suppose someone asked you to recall the same walk a month later. You wouldn’t stand a chance. You would likely be able to recount the basics of a typical walk across campus, but not the precise details of that particular walk. Yet, if you had seen a giraffe during that walk, the event would have been fixed in your mind for a long time, probably for the rest of your life. You would tell your friends about it, and, on later occasions when you saw a giraffe, you might be reminded of the day you saw one on

campus. Psychologists have long pinpointed **distinctiveness**—having an event stand out as quite different from a background of similar events—as a key to remembering events (Hunt, 2003).

In addition, when vivid memories are tinged with strong emotional content, they often seem to leave a permanent mark on us. Public tragedies, such as terrorist attacks, often create vivid memories in those who witnessed them. But even those of us not directly involved in such events may have vivid memories of them, including memories of first hearing about them. For example, many people are able to recall their exact physical location when they first learned about the assassination or accidental death of a national figure. The term **flashbulb memory** was originally coined by Brown and Kulik (1977) to describe this sort of vivid memory of finding out an important piece of news. The name refers to how some memories seem to be captured in the mind like a flash photograph; because of the distinctiveness and emotionality of the news, they seem to become permanently etched in the mind with exceptional clarity compared to other memories.

Take a moment and think back on your own life. Is there a particular memory that seems sharper than others? A memory where you can recall unusual details, like the colors of mundane things around you, or the exact positions of surrounding objects? Although people have great confidence in flashbulb memories like these, the truth is, our objective accuracy with them is far from perfect (Talarico & Rubin, 2003). That is, even though people may have great confidence in what they recall, their memories are not as accurate (e.g., what the actual colors were; where objects were truly placed) as they tend to imagine. Nonetheless, all other things being equal, distinctive and emotional events are well-remembered.

Details do not leap perfectly from the world into a person's mind. We might say that we went to a party and remember it, but what we remember is (at best) what we encoded. As noted above, the process of encoding is selective, and in complex situations, relatively few of many possible details are noticed and encoded. The process of encoding always involves **recoding**—that is, taking the information from the form it is delivered to us and then converting it in a way that we can make sense of it. For example, you might try to remember the colors of a rainbow by using the acronym ROY G BIV (red, orange, yellow, green, blue, indigo, violet). The process of recoding the colors into a name can help us to remember. However, recoding can also introduce errors—when we accidentally add information during encoding, then remember that *new* material as if it had been part of the actual experience (as discussed below).

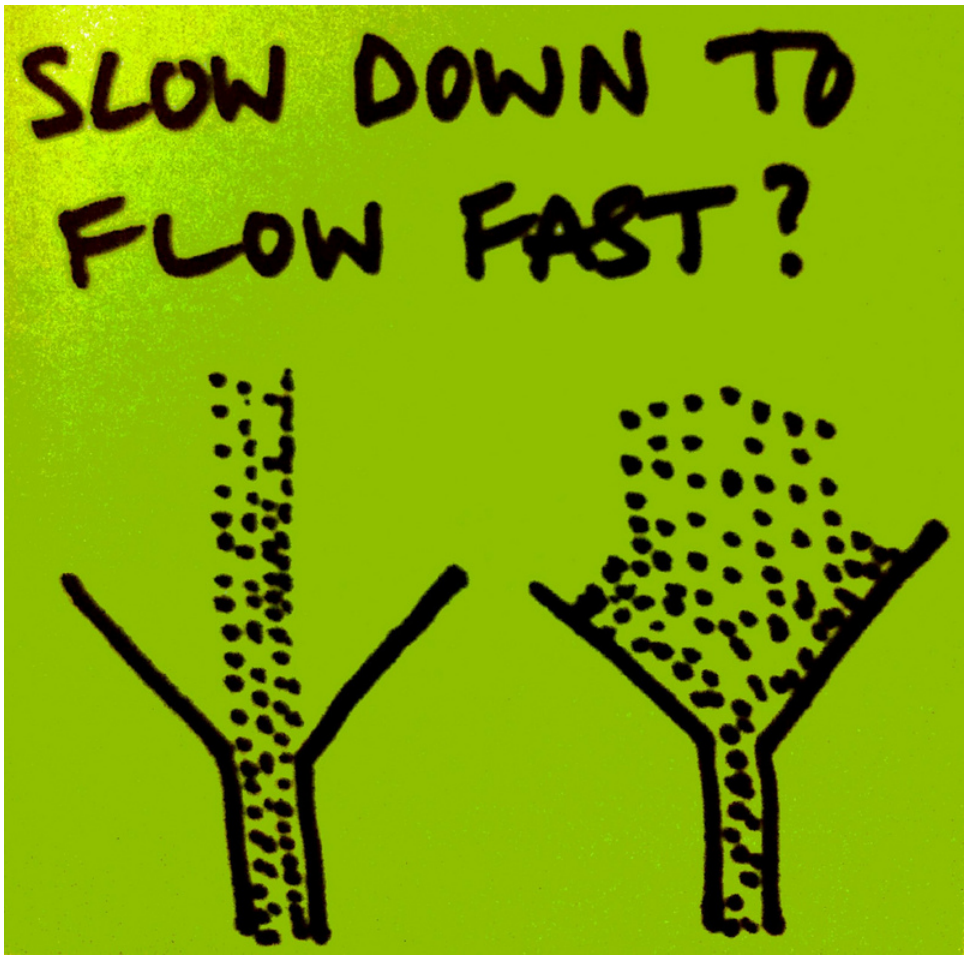


Figure **18.9.4**: Although it requires more effort, using images and associations can improve the process of recoding. [Image: psd, https://goo.gl/9xjcDe, CC BY 2.0, https://goo.gl/9uSnqN]

Psychologists have studied many recoding strategies that can be used during study to improve retention. First, research advises that, as we study, we should think of the meaning of the events (Craik & Lockhart, 1972), and we should try to relate new events to information we already know. This helps us form associations that we can use to retrieve information later. Second, imagining events also makes them more memorable; creating vivid images out of information (even verbal information) can greatly improve later recall (Bower & Reitman, 1972). Creating imagery is part of the technique Simon Reinhard uses to remember huge numbers of digits, but we can all use images to encode information more effectively. The basic concept behind good encoding strategies is to form distinctive memories (ones that stand out), and to form links or associations among memories to help later retrieval (Hunt & McDaniel, 1993). Using study strategies such as the ones described here is challenging, but the effort is well worth the benefits of enhanced learning and retention.

We emphasized earlier that encoding is selective: people cannot encode all information they are exposed to. However, recoding can add information that was not even seen or heard during the initial encoding phase. Several of the recoding processes, like forming associations between memories, can happen without our awareness. This is one reason people can sometimes remember events that did not actually happen—because during the process of recoding, details got added. One common way of inducing false memories in the laboratory employs a word-list technique (Deese, 1959; Roediger & McDermott, 1995). Participants hear lists of 15 words, like *door, glass, pane, shade, ledge, sill, house, open, curtain, frame, view, breeze, sash, screen,* and *shutter.* Later, participants are given a test in which they are shown a list of words and asked to pick out the ones they'd heard earlier. This second list contains some words from the first list (e.g., *door, pane, frame*) and some words not from the list (e.g., *arm, phone, bottle*). In this example, one of the words on the test is *window*, which—importantly—does not appear in the first list, but which is related to other words in that list. When subjects were tested, they were reasonably accurate with the studied words (*door*, etc.), recognizing them 72% of the time. However, when *window* was on the test, they falsely recognized it as having been on the list 84% of the time (Stadler, Roediger, & McDermott, 1999). The same thing happened with many other lists the authors used. This phenomenon is referred to

as the DRM (for Deese-Roediger-McDermott) effect. One explanation for such results is that, while students listened to items in the list, the words triggered the students to think about *window*, even though *window* was never presented. In this way, people seem to encode events that are not actually part of their experience.

Because humans are creative, we are always going beyond the information we are given: we automatically make associations and infer from them what is happening. But, as with the word association mix-up above, sometimes we make false memories from our inferences—remembering the inferences themselves as if they were actual experiences. To illustrate this, Brewer (1977) gave people sentences to remember that were designed to elicit *pragmatic inferences*. Inferences, in general, refer to instances when something is not explicitly stated, but we are still able to guess the undisclosed intention. For example, if your friend told you that she didn't want to go out to eat, you may infer that she doesn't have the money to go out, or that she's too tired. With *pragmatic* inferences, there is usually *one* particular inference you're likely to make. Consider the statement Brewer (1977) gave her participants: "The karate champion hit the cinder block." After hearing or seeing this sentence, participants who were given a memory test tended to remember the statement as having been, "The karate champion *broke* the cinder block." This remembered statement is not necessarily a *logical* inference (i.e., it is perfectly reasonable that a karate champion could hit a cinder block without breaking it). Nevertheless, the *pragmatic* conclusion from hearing such a sentence is that the block was likely broken. The participants remembered this inference they made while hearing the sentence in place of the actual words that were in the sentence (see also McDermott & Chan, 2006).

Encoding—the initial registration of information—is essential in the learning and memory process. Unless an event is encoded in some fashion, it will not be successfully remembered later. However, just because an event is encoded (even if it is encoded well), there's no guarantee that it will be remembered later.

18.9.5 Storage



Figure 18.9.5: Memory traces, or engrams, are NOT perfectly preserved recordings of past experiences. The traces are combined with current knowledge to reconstruct what we think happened in the past. [Simon Bierdwald, https://goo.gl/JDhdCE, CC BY-NC-SA 2.0, https://goo.gl/jSSrcO]

Every experience we have changes our brains. That may seem like a bold, even strange, claim at first, but it's true. We encode each of our experiences within the structures of the nervous system, making new impressions in the process—and each of those impressions involves changes in the brain. Psychologists (and neurobiologists) say that experiences leave **memory traces**, or **engrams** (the two terms are synonyms). Memories have to be stored somewhere in the brain, so in order to do so, the brain biochemically alters itself and its neural tissue. Just like you might write yourself a note to remind you of something, the brain "writes" a memory trace, changing its own physical composition to do so. The basic idea is that events (occurrences in our environment) create engrams through a process of **consolidation**: the neural changes that occur after learning to create the memory trace of an experience. Although neurobiologists are concerned with exactly what neural processes change when memories are

created, for psychologists, the term *memory trace* simply refers to the physical change in the nervous system (whatever that may be, exactly) that represents our experience.

Although the concept of engram or memory trace is extremely useful, we shouldn't take the term too literally. It is important to understand that memory traces are not perfect little packets of information that lie dormant in the brain, waiting to be called forward to give an accurate report of past experience. Memory traces are not like video or audio recordings, capturing experience with great accuracy; as discussed earlier, we often have errors in our memory, which would not exist if memory traces were perfect packets of information. Thus, it is wrong to think that remembering involves simply "reading out" a faithful record of past experience. Rather, when we remember past events, we reconstruct them with the aid of our memory traces—but also with our current belief of what happened. For example, if you were trying to recall for the police who started a fight at a bar, you may not have a memory trace of who pushed whom first. However, let's say you remember that one of the guys held the door open for you. When thinking back to the start of the fight, this knowledge (of how one guy was friendly to you) may unconsciously influence your memory of what happened in favor of the nice guy. Thus, memory is a construction of what you actually recall and what you believe happened. In a phrase, remembering is reconstructive (we reconstruct our past with the aid of memory traces) not reproductive (a perfect reproduction or recreation of the past).

Psychologists refer to the time between learning and testing as the retention interval. Memories can consolidate during that time, aiding retention. However, experiences can also occur that undermine the memory. For example, think of what you had for lunch yesterday—a pretty easy task. However, if you had to recall what you had for lunch 17 days ago, you may well fail (assuming you don't eat the same thing every day). The 16 lunches you've had since that one have created **retroactive interference**. Retroactive interference refers to new activities (i.e., the subsequent lunches) during the retention interval (i.e., the time between the lunch 17 days ago and now) that interfere with retrieving the specific, older memory (i.e., the lunch details from 17 days ago). But just as newer things can interfere with remembering older things, so can the opposite happen. *Proactive interference* is when past memories interfere with the encoding of new ones. For example, if you have ever studied a second language, often times the grammar and vocabulary of your native language will pop into your head, impairing your fluency in the foreign language.

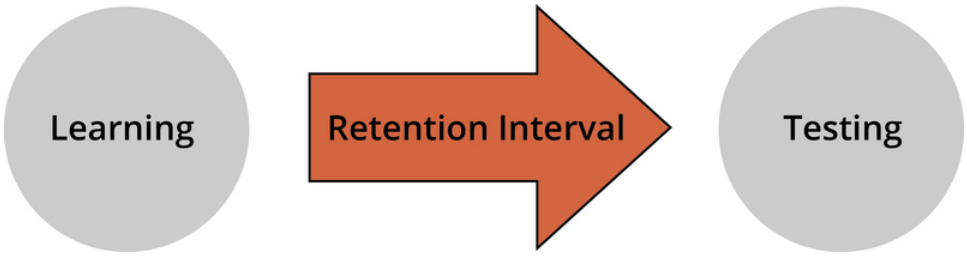


Figure 18.9.6:

Retroactive interference is one of the main causes of forgetting (McGeoch, 1932). In the module *Eyewitness Testimony and Memory Biases* http://noba.to/uy49tm37 Elizabeth Loftus describes her fascinating work on eyewitness memory, in which she shows how memory for an event can be changed via misinformation supplied during the retention interval. For example, if you witnessed a car crash but subsequently heard people describing it from their own perspective, this new information may interfere with or disrupt your own personal recollection of the crash. In fact, you may even come to remember the event happening exactly as the others described it! This **misinformation effect** in eyewitness memory represents a type of retroactive interference that can occur during the retention interval (see Loftus [2005] for a review). Of course, if correct information is given during the retention interval, the witness's memory will usually be improved.

Although interference may arise between the occurrence of an event and the attempt to recall it, *the effect itself is always expressed when we retrieve memories*, the topic to which we turn next.

18.9.6 Retrieval

Endel Tulving argued that "the key process in memory is retrieval" (1991, p. 91). Why should retrieval be given more prominence than encoding or storage? For one thing, if information were encoded and stored but could not be retrieved, it would be useless. As discussed previously in this module, we encode and store thousands of events—conversations, sights and sounds—every day, creating memory traces. However, we later access only a tiny portion of what we've taken in. Most of our memories will never be

used—in the sense of being brought back to mind, consciously. This fact seems so obvious that we rarely reflect on it. All those events that happened to you in the fourth grade that seemed so important then? Now, many years later, you would struggle to remember even a few. You may wonder if the traces of those memories still exist in some latent form. Unfortunately, with currently available methods, it is impossible to know.

Psychologists distinguish information that is available in memory from that which is accessible (Tulving & Pearlstone, 1966). *Available* information is the information that is stored in memory—but precisely how much and what types are stored cannot be known. That is, all we can know is what information we can retrieve—*accessible* information. The assumption is that accessible information represents only a tiny slice of the information available in our brains. Most of us have had the experience of trying to remember some fact or event, giving up, and then—all of a sudden!—it comes to us at a later time, even after we've stopped trying to remember it. Similarly, we all know the experience of failing to recall a fact, but then, if we are given several choices (as in a multiple-choice test), we are easily able to recognize it.



Figure **18.9.7**: We can't know the entirety of what is in our memory, but only that portion we can actually retrieve. Something that cannot be retrieved now and which is seemingly gone from memory may, with different cues applied, reemerge. [Image: Ores2k, https://goo.gl/1du8Qe, CC BY-NC-SA 2.0, https://goo.gl/jSSrcO]

What factors determine what information can be retrieved from memory? One critical factor is the type of hints, or *cues*, in the environment. You may hear a song on the radio that suddenly evokes memories of an earlier time in your life, even if you were not trying to remember it when the song came on. Nevertheless, the song is closely associated with that time, so it brings the experience to mind.

The general principle that underlies the effectiveness of retrieval cues is the **encoding specificity principle** (Tulving & Thomson, 1973): when people encode information, they do so in specific ways. For example, take the song on the radio: perhaps you heard it while you were at a terrific party, having a great, philosophical conversation with a friend. Thus, the song became part of that whole complex experience. Years later, even though you haven't thought about that party in ages, when you hear the song on the radio, the whole experience rushes back to you. In general, the encoding specificity principle states that, to the extent a retrieval cue (the song) matches or overlaps the memory trace of an experience (the party, the conversation), it will be effective in evoking the memory. A classic experiment on the encoding specificity principle had participants memorize a set of words in a unique setting. Later, the participants were tested on the word sets, either in the same location they learned the words or a different one. As a result

of encoding specificity, the students who took the test in the same place they learned the words were actually able to recall more words (Godden & Baddeley, 1975) than the students who took the test in a new setting.

One caution with this principle, though, is that, for the cue to work, it can't match too many other experiences (Nairne, 2002; Watkins, 1975). Consider a lab experiment. Suppose you study 100 items; 99 are words, and one is a picture—of a penguin, item 50 in the list. Afterwards, the cue "recall the picture" would evoke "penguin" perfectly. No one would miss it. However, if the *word* "penguin" were placed in the same spot among the other 99 words, its memorability would be exceptionally worse. This outcome shows the power of distinctiveness that we discussed in the section on encoding: one picture is perfectly recalled from among 99 words because it stands out. Now consider what would happen if the experiment were repeated, but there were 25 pictures distributed within the 100-item list. Although the picture of the penguin would still be there, the probability that the cue "recall the picture" (at item 50) would be useful for the penguin would drop correspondingly. Watkins (1975) referred to this outcome as demonstrating the **cue overload principle**. That is, to be effective, a retrieval cue cannot be overloaded with too many memories. For the cue "recall the picture" to be effective, it should only match one item in the target set (as in the one-picture, 99-word case).

To sum up how memory cues function: for a retrieval cue to be effective, a match must exist between the cue and the desired target memory; furthermore, to produce the best retrieval, the cue-target relationship should be distinctive. Next, we will see how the encoding specificity principle can work in practice.

Psychologists measure memory performance by using production tests (involving recall) or recognition tests (involving the selection of correct from incorrect information, e.g., a multiple-choice test). For example, with our list of 100 words, one group of people might be asked to recall the list in any order (a free recall test), while a different group might be asked to circle the 100 studied words out of a mix with another 100, unstudied words (a recognition test). In this situation, the recognition test would likely produce better performance from participants than the recall test.

We usually think of recognition tests as being quite easy, because the cue for retrieval is a copy of the actual event that was presented for study. After all, what could be a better cue than the exact target (memory) the person is trying to access? In most cases, this line of reasoning is true; nevertheless, recognition tests do not provide perfect indexes of what is stored in memory. That is, you can fail to recognize a target staring you right in the face, yet be able to recall it later with a different set of cues (Watkins & Tulving, 1975). For example, suppose you had the task of recognizing the surnames of famous authors. At first, you might think that being given the actual last name would always be the best cue. However, research has shown this not necessarily to be true (Muter, 1984). When given names such as Tolstoy, Shaw, Shakespeare, and Lee, subjects might well say that Tolstoy and Shakespeare are famous authors, whereas Shaw and Lee are not. But, when given a cued recall test using first names, people often recall items (produce them) that they had failed to recognize before. For example, in this instance, a cue like *George Bernard* _____ often leads to a recall of "Shaw," even though people initially failed to recognize *Shaw* as a famous author's name. Yet, when given the cue "William," people may not come up with Shakespeare, because William is a common name that matches many people (the cue overload principle at work). This strange fact—that recall can sometimes lead to better performance than recognition—can be explained by the encoding specificity principle. As a cue, *George Bernard* _____ matches the way the famous writer is stored in memory better than does his surname, Shaw, does (even though it is the target). Further, the match is quite distinctive with *George Bernard* _____, but the cue *William* _____ is much more overloaded (Prince William, William Yeats, William Faulkner, will.i.am).

The phenomenon we have been describing is called the *recognition failure of recallable words*, which highlights the point that a cue will be most effective depending on how the information has been encoded (Tulving & Thomson, 1973). The point is, the cues that work best to evoke retrieval are those that recreate the event or name to be remembered, whereas sometimes even the target itself, such as *Shaw* in the above example, is not the best cue. Which cue will be most effective depends on how the information has been encoded.

Whenever we think about our past, we engage in the act of retrieval. We usually think that retrieval is an objective act because we tend to imagine that retrieving a memory is like pulling a book from a shelf, and after we are done with it, we return the book to the shelf just as it was. However, research shows this assumption to be false; far from being a static repository of data, the memory is constantly changing. In fact, every time we retrieve a memory, it is altered. For example, the act of retrieval itself (of a fact, concept, or event) makes the retrieved memory much more likely to be retrieved again, a phenomenon called the *testing effect* or the *retrieval practice effect* (Pyc & Rawson, 2009; Roediger & Karpicke, 2006). However, retrieving some information can actually cause us to forget other information related to it, a phenomenon called *retrieval-induced forgetting* (Anderson, Bjork, &

Bjork, 1994). Thus the act of retrieval can be a double-edged sword—strengthening the memory just retrieved (usually by a large amount) but harming related information (though this effect is often relatively small).

As discussed earlier, retrieval of distant memories is reconstructive. We weave the concrete bits and pieces of events in with assumptions and preferences to form a coherent story (Bartlett, 1932). For example, if during your 10th birthday, your dog got to your cake before you did, you would likely tell that story for years afterward. Say, then, in later years you misremember where the dog actually found the cake, but repeat that error over and over during subsequent retellings of the story. Over time, that inaccuracy would become a basic fact of the event in your mind. Just as retrieval practice (repetition) enhances accurate memories, so will it strengthen errors or false memories (McDermott, 2006). Sometimes memories can even be manufactured just from hearing a vivid story. Example of a mneumonic system created by a student to study cranial nerves. [Image: Kelidimari, https://goo.gl/kiA1kP, CC BY-SA 3.0, https://goo.gl/SCkRfm]

How did Simon Reinhard remember those digits? He uses “memory palaces” (elaborate scenes with discrete places) combined with huge sets of images for digits. For example, imagine mentally walking through the home where you grew up and identifying as many distinct areas and objects as possible. Simon has hundreds of such memory palaces that he uses. Next, for remembering digits, he has memorized a set of 10,000 images. Every four-digit number for him immediately brings forth a mental image. So, for example, 6187 might recall Michael Jackson. When Simon hears all the numbers coming at him, he places an image for every four digits into locations in his memory palace. He can do this at an incredibly rapid rate, faster than 4 digits per 4 seconds when they are flashed visually, as in the demonstration at the beginning of the module. As noted, his record is 240 digits, recalled in exact order. Simon also holds the world record in an event called “speed cards,” which involves memorizing the precise order of a shuffled deck of cards. Simon was able to do this in 21.19 seconds! Again, he uses his memory palaces, and he encodes groups of cards as single images.

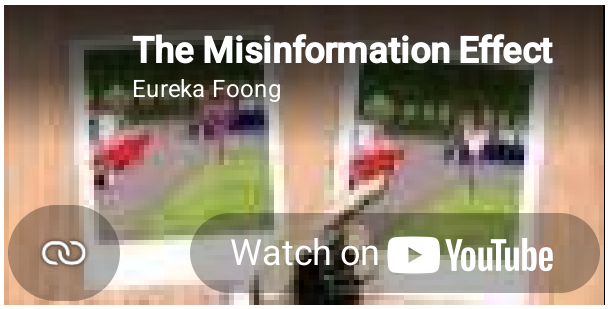
Many books exist on how to improve memory using mnemonic devices, but all involve forming distinctive encoding operations and then having an infallible set of memory cues. We should add that to develop and use these memory systems takes a great amount of time and concentration. The World Memory Championships are held every year and the records keep improving. However, for most common purposes, just keep in mind that to remember well you need to encode information in a distinctive way and to have good cues for retrieval. You can adapt a system that will meet most any purpose.

18.9.6.1 Outside Resources

Book: Brown, P.C., Roediger, H. L. & McDaniel, M. A. (2014). *Make it stick: The science of successful learning.* Cambridge, MA: Harvard University Press.

https://www.amazon.com/Make-Stick-Science-Successful-Learning/dp/0674729013

Student Video 1: Eureka Foong\\\\'s - The Misinformation Effect. This is a student-made video illustrating this phenomenon of altered memory. It was one of the winning entries in the 2014 Noba Student Video Award.

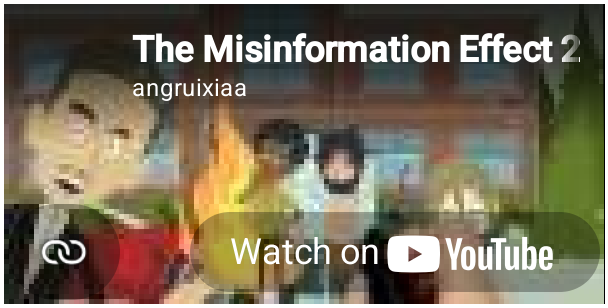




Student Video 2: Kara McCord\\\\'s - Flashbulb Memories. This is a student-made video illustrating this phenomenon of autobiographical memory. It was one of the winning entries in the 2014 Noba Student Video Award.



Stu Rui Xia & Ong Jun Hao\\\\'s - The Misinformation Effect. Another student-made video exploring the mis . Also an award winner from 2014.





Video: Simon Reinhard breaking the world record in speedcards.



Web: Retrieval Practice, a website with research, resources, and tips for both educators and learners around the memory-strengthening skill of retrieval practice.

http://www.retrievalpractice.org/

18.9.6.2 Discussion Questions

1. Mnemonists like Simon Reinhard develop mental "journeys," which enable them to use the method of loci. Develop your own journey, which contains 20 places, in order, that you know well. One example might be: the front walkway to your parents' apartment; their doorbell; the couch in their living room; etc. Be sure to use a set of places that you know well and that have a natural order to them (e.g., the walkway comes before the doorbell). Now you are more than halfway toward being able to memorize a set of 20 nouns, in order, rather quickly. As an optional second step, have a friend make a list of 20 such nouns and read them to you, slowly (e.g., one every 5 seconds). Use the method to attempt to remember the 20 items.
2. Recall a recent argument or misunderstanding you have had about memory (e.g., a debate over whether your girlfriend/boyfriend had agreed to something). In light of what you have just learned about memory, how do you think about it? Is it possible that the disagreement can be understood by one of you making a pragmatic inference?
3. Think about what you've learned in this module and about how you study for tests. On the basis of what you have learned, is there something you want to try that might help your study habits?

18.9.6.3 Vocabulary

Autobiographical memory

Memory for the events of one's life.

Consolidation

The process occurring after encoding that is believed to stabilize memory traces.

Cue overload principle

The principle stating that the more memories that are associated to a particular retrieval cue, the less effective the cue will be in prompting retrieval of any one memory.

Distinctiveness

The principle that unusual events (in a context of similar events) will be recalled and recognized better than uniform (nondistinctive) events.

Encoding

The initial experience of perceiving and learning events.

Encoding specificity principle

The hypothesis that a retrieval cue will be effective to the extent that information encoded from the cue overlaps or matches information in the engram or memory trace.

Engram

the physical change in the nervous system representing memory of an event or fact; also, memory trace.

Episodic memory

Memory for events in a particular time and place.

Flashbulb memory

Vivid personal memories of receiving the news of some momentous (and usually emotional) event.

Memory traces

the change in the nervous system representing memory an event or fact; also engram.

Misinformation effect

When erroneous information occurring after an event is remembered as having been part of the original event.

Mnemonic devices

A strategy for remembering large amounts of information, usually involving imaging events occurring on a journey or with some other set of memorized cues.

Recoding

The ubiquitous process during learning of taking information in one form and converting it to another form, usually one more easily remembered.

Retrieval

The process of accessing stored information.

Retroactive interference

The phenomenon whereby events that occur after some particular event of interest will usually cause forgetting of the original event.

Semantic memory

The more or less permanent store of knowledge that people have.

Storage

The stage in the learning/memory process that bridges encoding and retrieval; the persistence of memory over time.

18.9.6.4 References

Anderson, M. C., Bjork, R., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology-Learning Memory and Cognition, 20*, 1063–1087.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge: Cambridge University Press.

Bower, G. H., & Reitman, J. S. (1972). Mnemonic elaboration in multilist learning. *Journal of Verbal Learning and Verbal Behavior, 11*, 478–485.

Brewer, W. F. (1977). Memory for the pragmatic implications of sentences. *Memory & Cognition, 5(6)*, 673–678.

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition, 5*, 73–99.

Chan, J.C.K. & McDermott, K.B. (2006). Remembering pragmatic inferences. *Applied Cognitive Psychology, 20*, 633-639.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*,66 (3), 325-331

Hunt, R. (2003). Two contributions of distinctive processing to accurate memory. *Journal of Memory and Language, 48*, 811–825.

Hunt, R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32*, 421-445.

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12*, 361–366.

McDermott, K. B. (2006). Paradoxical effects of testing: Repeated retrieval attempts enhance the likelihood of later accurate and false recall. *Memory & Cognition, 34*, 261–267.

McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review, 39(4)*, 352.

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior, 2*, 1–21.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Muter, P. (1984). Recognition and recall of words with a single meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 198–202.

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory, 10*, 389–395.

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition, 25*, 838–848.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60*, 437–447.

Roediger, H. L. (1980). The effectiveness of four mnemonics in ordering recall. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 558.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249–255.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology-Learning Memory and Cognition, 21*, 803–814.

Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition, 27*, 494–500.

Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science, 14*, 455–461.

Tulving, E. (2007). Are there 256 different kinds of memory? In J.S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger, III* (pp. 39–52). New York: Psychology Press.

Tulving, E. (1991). Interview. *Journal of Cognitive Neuroscience, 3*, 89–94

Tulving, E., & Bower, G. H. (1975). The logic of memory representations. *The psychology of learning and motivation, 8*, 265-301.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5*, 381–391.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352–373.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352–373.

Watkins, M. J. (1975). Inhibition in recall with extralist “cues.” *Journal of Verbal Learning and Verbal Behavior, 14*, 294–303.

Watkins, M. J., & Tulving, E. (1975). Episodic memory: When recognition fails. *Journal of Experimental Psychology: General, 104*, 5–29.

18.9.6.5 Attributions

Adapted by Kenneth A. Koenigshofer, Ph.D. from McDermott, K. B. & Roediger, H. L. (2021). Memory (encoding, storage, retrieval). In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology.* Champaign, IL: DEF publishers. Retrieved from http://noba.to/bdc4uger

Authors:

- 

  Kathleen B. McDermott is Professor of Psychology and Radiology at Washington University in St. Louis. She studies remembering using both behavioral and neuroimaging techniques. She received the Shahin Hashtroudi Memorial Prize for Researcher in Memory from the Association for Psychological Science and the James S. McGuigan Young Investigator Prize from the American Psychological Foundation. She is a Fellow of the Association for Psychological Science.
- 

  Henry L. Roediger, III is the James S. McDonnell Distinguished University Professor at Washington University in St. Louis who has spent his career studying learning and memory. He has received the Howard Warren Crosby Medal from the Society of Experimental Psychologists and the William James Award for Lifetime Achievements in Psychology from the Association of Psychological Science. He also served as President of APS.

18.9.6.6 Creative Commons License

Memory (Encoding, Storage, Retrieval) by Kathleen B. McDermott and Henry L. Roediger III is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Permissions beyond the scope of this license may be available in our Licensing Agreement

This page titled 18.9: Chapter 10- The Psychology of Memory Encoding, Storage, and Retrieval is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

18.10: Chapter 10- Psychological Factors Affecting Learning, Memory, and Forgetting

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

18.10.1 **Forgetting**

Learning Objectives

1. Identify five reasons we forget and give examples of each.
2. Describe how forgetting can be viewed as an adaptive process.
3. Explain the difference between anterograde and retrograde amnesia.

18.10.2 Overview

This section explores the causes of everyday forgetting. Forgetting is viewed as an adaptive process that allows us to be efficient in terms of the information we retain.

18.10.3 **Introduction**

Chances are that you have experienced memory lapses and been frustrated by them. You may have had trouble remembering the definition of a key term on an exam or found yourself unable to recall the name of an actor from one of your favorite TV shows. Maybe you forgot to call your aunt on her birthday or you routinely forget where you put your cell phone. Oftentimes, the bit of information we are searching for comes back to us, but sometimes it does not. Clearly, forgetting seems to be a natural part of life. Why do we forget? And is forgetting always a bad thing?



Figure **18.10.1**: Forgetting can often be obnoxious or even embarrassing. But as we explore this module, you'll learn that forgetting is important and necessary for everyday functionality. [Image: jazbeck, https://goo.gl/nkRrJy, CC BY 2.0, https://goo.gl/BRvSA7]

18.10.4 Causes of Forgetting

One very common and obvious reason why you cannot remember a piece of information is because you did not learn it in the first place. If you fail to encode information into memory, you are not going to remember it later on. Usually, **encoding** failures occur because we are distracted or are not paying attention to specific details. For example, people have a lot of trouble recognizing an actual penny out of a set of drawings of very similar pennies, or lures, even though most of us have had a lifetime of experience handling pennies (Nickerson & Adams, 1979). However, few of us have studied the features of a penny in great detail, and since

we have not attended to those details, we fail to recognize them later. Similarly, it has been well documented that distraction during learning impairs later memory (e.g., Craik, Govoni, Naveh-Benjamin, & Anderson, 1996). Most of the time this is not problematic, but in certain situations, such as when you are studying for an exam, failures to encode due to distraction can have serious repercussions.

Another proposed reason why we forget is that memories fade, or **decay**, over time. It has been known since the pioneering work of Hermann Ebbinghaus (1885/1913) that as time passes, memories get harder to recall. Ebbinghaus created more than 2,000 nonsense syllables, such as *dax*, *bap*, and *rif*, and studied his own memory for them, learning as many as 420 lists of 16 nonsense syllables for one experiment. He found that his memories diminished as time passed, with the most forgetting happening early on after learning. His observations and subsequent research suggested that if we do not rehearse a memory and the neural representation of that memory is not reactivated over a long period of time, the memory representation (the engram) may disappear entirely or fade to the point where it can no longer be accessed. As you might imagine, it is hard to definitively prove that a memory has decayed as opposed to it being inaccessible for another reason. Critics argued that forgetting must be due to processes other than simply the passage of time, since disuse of a memory does not always guarantee forgetting (McGeoch, 1932). More recently, some memory theorists have proposed that recent memory traces may be degraded or disrupted by new experiences (Wixted, 2004).

Memory traces need to be **consolidated**, or transferred from the **hippocampus** to more durable representations in the **cortex,** in order for them to last (McGaugh, 2000). When the consolidation process is interrupted by the encoding of other experiences, the memory trace for the original experience does not get fully developed and thus is forgotten.



Figure **18.10.2**: At times, we will completely blank on something we're certain we've learned - people we went to school with years ago for example. However, once we get the right retrieval cue (a name perhaps), the memory (faces or experiences) rushes back to us like it was there all along. [Image: sbhsclass84, https://goo.gl/sHZyQI, CC BY-SA 2.0, https://goo.gl/rxiUsF]

Both encoding failures and decay account for more permanent forms of forgetting, in which the memory trace does not exist, but forgetting may also occur when a memory exists yet we temporarily cannot access it. This type of forgetting, failure of retrieval, may occur when we lack the appropriate **retrieval** cues for bringing the memory to mind. You have probably had the frustrating experience of forgetting your password for an online site. Usually, the password has not been permanently forgotten; instead, you just need the right reminder to remember what it is. For example, if your password was "pizza0525," and you received the password hints "favorite food" and "Mom's birthday," you would easily be able to retrieve it. Retrieval hints can bring back to mind seemingly forgotten memories (Tulving & Pearlstone, 1966). One real-life illustration of the importance of retrieval cues comes from a study showing that whereas people have difficulty recalling the names of high school classmates years after graduation, they are easily able to recognize the names and match them to the appropriate faces (Bahrick, Bahrick, & Wittinger, 1975). The names are powerful enough retrieval cues that they bring back the memories of the faces that went with them. The fact that the presence of the right retrieval cues is critical for remembering adds to the difficulty in proving that a memory is permanently forgotten as opposed to temporarily unavailable.

Retrieval failures can also occur because other memories are blocking or getting in the way of recalling the desired memory. This blocking is referred to as **interference**. For example, you may fail to remember the name of a town you visited with your family on summer vacation because the names of other towns you visited on that trip or on other trips come to mind instead. Those memories then prevent the desired memory from being retrieved. Interference is also relevant to the example of forgetting a password: passwords that we have used for other websites may come to mind and interfere with our ability to retrieve the desired password. Interference can be either proactive, in which old memories block the learning of new related memories, or retroactive, in which new memories block the retrieval of old related memories. For both types of interference, competition between memories seems to be key (Mensink & Raaijmakers, 1988). Your memory for a town you visited on vacation is unlikely to interfere with your ability to remember an Internet password, but it is likely to interfere with your ability to remember a different town's name. Competition between memories can also lead to forgetting in a different way. Recalling a desired memory in the face of competition may result in the inhibition of related, competing memories (Levy & Anderson, 2002). You may have difficulty recalling the name of Kennebunkport, Maine, because other Maine towns, such as Bar Harbor, Winterport, and Camden, come to mind instead. However, if you are able to recall Kennebunkport despite strong competition from the other towns, this may actually change the competitive landscape, weakening memory for those other towns' names, leading to forgetting of them instead.

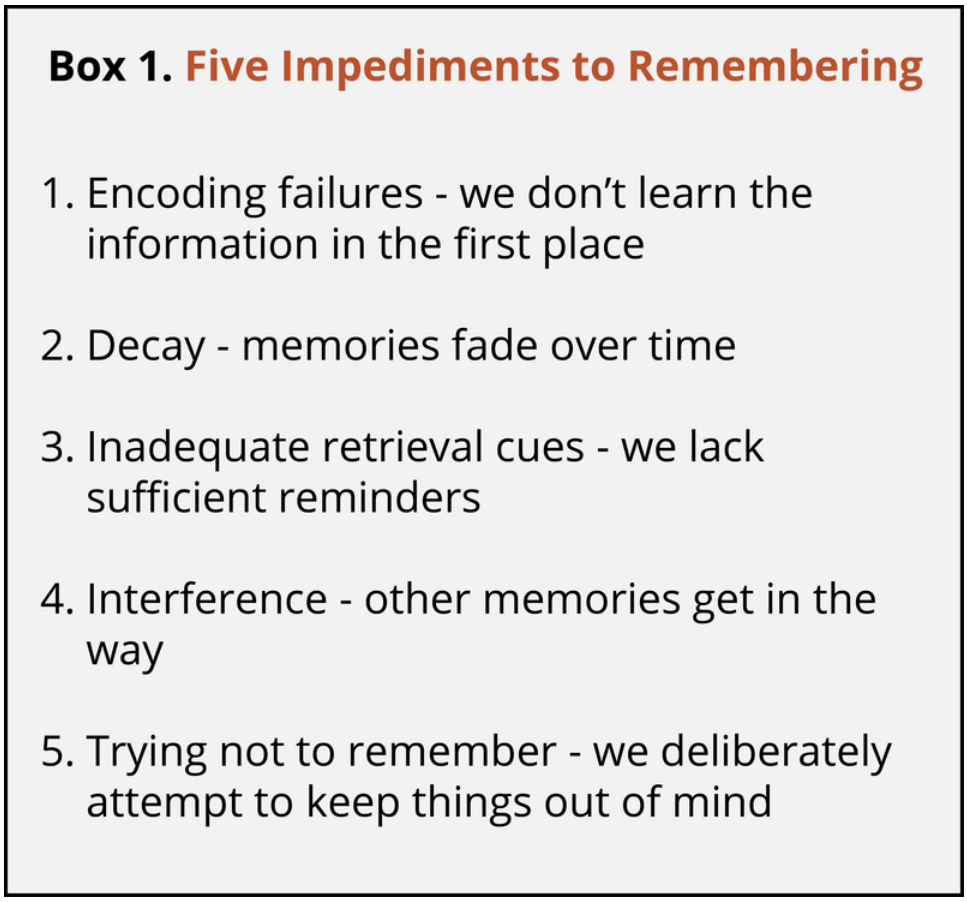


Figure 18.10.3: Causes of forgetting.

Finally, some memories may be forgotten because *we deliberately attempt to keep them out of mind*. Over time, by actively trying not to remember an event, we can sometimes successfully keep the undesirable memory from being retrieved either by inhibiting the undesirable memory or generating diversionary thoughts (Anderson & Green, 2001). Imagine that you slipped and fell in your high school cafeteria during lunch time, and everyone at the surrounding tables laughed at you. You would likely wish to avoid thinking about that event and might try to prevent it from coming to mind. One way that you could accomplish this is by thinking of other, more positive, events that are associated with the cafeteria. Eventually, this memory may be suppressed to the point that it would only be retrieved with great difficulty (Hertel & Calcaterra, 2005).

18.10.5 Adaptive Forgetting

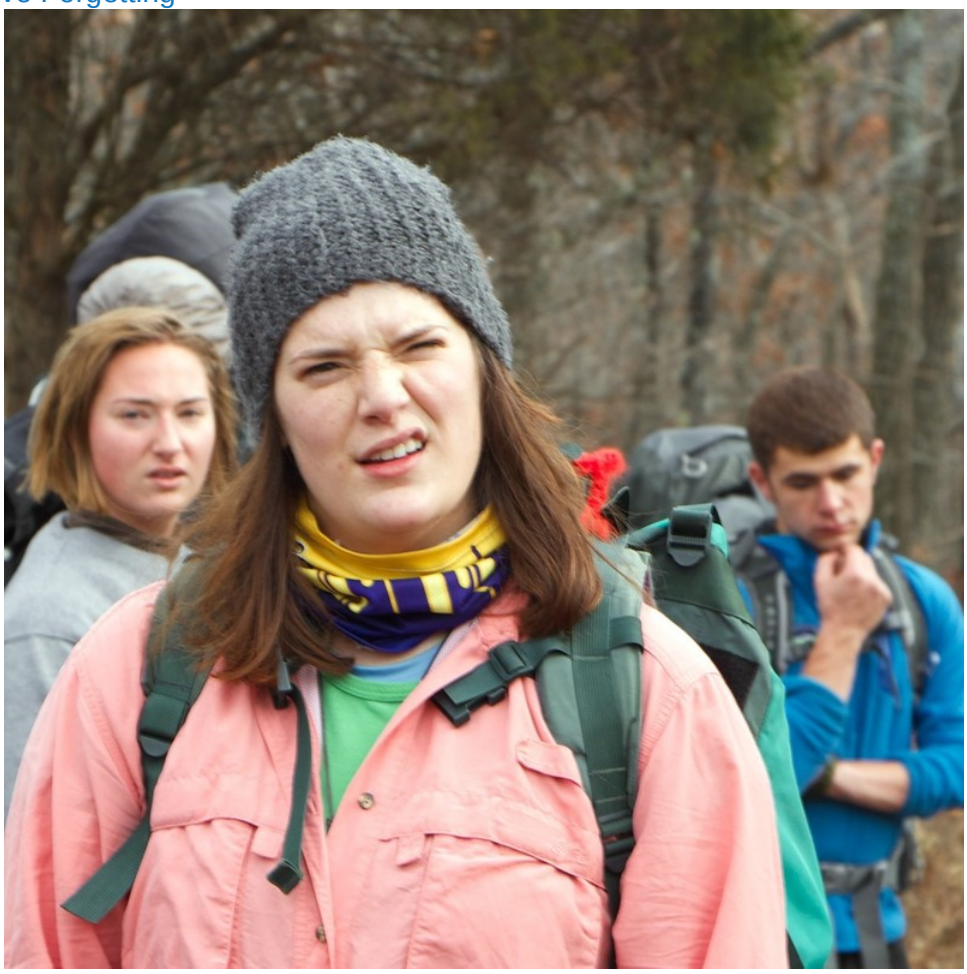


Figure **18.10.4**: Could you imagine being unable to forget every path you have taken while hiking? Each new trip, you would be walking around the forest for days, incapable of distinguishing today's path from the prior ones. [Image: Dan Trew, https://goo.gl/8fJWWE, CC BY-SA 2.0, https://goo.gl/rxiUsF]

We have explored five different causes of forgetting. Together they can account for the day-to-day episodes of forgetting that each of us experience. Typically, we think of these episodes in a negative light and view forgetting as a memory failure. Is forgetting ever good? Most people would reason that forgetting that occurs in response to a deliberate attempt to keep an event out of mind is a good thing. No one wants to be constantly reminded of falling on their face in front of all of their friends. However, beyond that, it can be argued that forgetting is adaptive, allowing us to be efficient and hold onto only the most relevant memories (Bjork, 1989; Anderson & Milson, 1989). Shereshevsky, or "S," the mnemonist studied by Alexander Luria (1968), was a man who almost never forgot. His memory appeared to be virtually limitless. He could memorize a table of 50 numbers in under 3 minutes and recall the numbers in rows, columns, or diagonals with ease. He could recall lists of words and passages that he had memorized over a decade before. Yet Shereshevsky found it difficult to function in his everyday life because he was constantly distracted by a flood of details and associations that sprung to mind. His case history suggests that remembering everything is not always a good thing. You may occasionally have trouble remembering where you parked your car, but imagine if every time you had to find your car, every single former parking space came to mind. The task would become impossibly difficult to sort through all of those irrelevant memories. Thus, forgetting is adaptive in that it makes us more efficient. The price of that efficiency is those moments when our memories seem to fail us (Schacter, 1999).

18.10.6 Conclusion

Just as the case study of the mnemonist Shereshevsky illustrates what a life with a near perfect memory would be like, amnesiac patients show us what a life without memory would be like. Each of the mechanisms we discussed that explain everyday forgetting

—encoding failures, decay, insufficient retrieval cues, interference, and intentional attempts to forget—help to keep us highly efficient, retaining the important information and for the most part, forgetting the unimportant.

18.10.6.1 Outside Resources

Web: Brain Case Study: Patient HM

https://bigpictureeducation.com/brain-case-study-patient-hm

Web: Self-experiment, Penny demo

http://www.indiana.edu/~p1013447/dictionary/penny.htm

Web: The Man Who Couldn't Remember

http://www.pbs.org/wgbh/nova/body/corkin-hm-memory.html

18.10.6.2 Discussion Questions

1. Is forgetting good or bad? Do you agree with the authors that forgetting is an adaptive process? Why or why not?
2. Can we ever prove that something is forgotten? Why or why not?
3. Which of the five reasons for forgetting do you think explains the majority of incidences of forgetting?

18.10.6.3 Terms with Links

- Anterograde amnesia
- Consolidation
- Decay
- Declarative memory
- Dissociative amnesia
- Encoding
- Interference
- Medial temporal lobes
- Retrieval
- Retrograde amnesia
- Temporally graded retrograde amnesia

18.10.6.4 Vocabulary

Decay

18.10.6.5 References

The fading of memories with the passage of time.

Declarative memory

Conscious memories for facts and events.

Dissociative amnesia

Loss of autobiographical memories from a period in the past in the absence of brain injury or disease.

Encoding

Process by which information gets into memory.

Interference

Other memories get in the way of retrieving a desired memory

Retrieval

Process by which information is accessed from memory and utilized.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, 96, 703–719.

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, 410, 366–369.

Bahrick, H. P., Bahrick, P. O., & Wittinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General*, 104, 54–75.

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger, III, & F. I. M. Craik (Eds.), *Varieties of Memory and Consciousness* (pp. 309– 330). Hillsdale, NJ: Erlbaum.

Corkin, S. (2002). What's new with the amnesic patient H. M.? *Nature Reviews Neuroscience*, 3, 153–160.

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, 125, 159–180.

Ebbinghaus, H. (1913). *Memory. A contribution to experimental psychology*. New York: Teachers College/Columbia University (Engl. ed.). (Original work published in 1885.)

Hertel, P. T., & Calcaterra, G. (2005). Intentional forgetting benefits from thought substitution. *Psychonomic Bulletin & Review*, 12, 484–489.

Hodges, J. R. (1994). Retrograde amnesia. In A. Baddeley, B. A. Wilson, & F. Watts (Eds.), *Handbook of Memory Disorders* (pp. 81–107). New York: Wiley.

Kihlstrom, J. F. (2005). Dissociative disorders. Annual *Review of Clinical Psychology*, 1, 227– 253.

Kopelman, M. (2000). Focal retrograde amnesia and the attribution of causality: An exceptionally critical review. *Cognitive Neuropsychology*, 17, 585–621.

Levy, B. J., & Anderson, M. C. (2002). Inhibitory processes and the control of memory retrieval. *Trends in Cognitive Sciences*, 6, 299–305.

Luria, A. R. (1968). *The mind of a mnemonist: A little book about a vast memory* (L. Solataroff, Trans.). New York: Basic Books.

McGaugh, J. L. (2000). Memory: A century of consolidation. *Science*, 287, 248–251.

McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Reviews*, 39, 352– 370.

Mensink, G., & Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychological Review*, 95, 434–455.

Nickerson, R. S., & Adams, M. J. (1979). Long-term memory for a common object. *Cognitive Psychology*, 11, 287–307.

Reed, J. M. & Squire, L. R. (1998). Retrograde amnesia for facts and events: Findings from four new cases. *Journal of Neuroscience*, 18, 3943–3954.

Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist*, 54, 182–203.

Scoville, W. B. & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology*, Neurosurgery, & Psychiatry, 20, 11–21.

Squire, L. R., & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: A neurobiological perspective. *Current Opinions in Neurobiology*, 5, 169–177.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 5, 381–391.

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Reviews of Psychology*, 55, 235–269.

18.10.6.6 Attributions

Adapted by Kenneth A. Koenigshofer, PhD, from Dudukovic, N. & Kuhl, B. (2021). Forgetting and amnesia. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology.* Champaign, IL: DEF publishers. Retrieved from http://noba.to/m38qbftg

18.10.6.7 Authors

- Nicole Dudukovic earned her Ph.D. in Psychology from Stanford University and has taught courses at Stanford, Trinity

College, and New York University. Her research explores interactions between attention and memory, and she is interested in applying memory research to other fields, particularly education.

- 

  Brice Kuhl earned his Ph.D. in Psychology from Stanford University and completed postdoctoral work at Yale University. He is currently an Assistant Professor of Psychology at New York University. His research explores the neural mechanisms of memory and causes of forgetting.

18.10.6.8 Creative Commons License

Forgetting and Amnesia by Nicole Dudukovic and Brice Kuhl is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Permissions beyond the scope of this license may be available in our Licensing Agreement.

18.10.7 **Factors Influencing Learning**

By Aaron Benjamin, University of Illinois at Urbana-Champaign

Modified by Kenneth A. Koenigshofer, PhD., Chaffey College

Learning is a complex process that defies easy definition and description. This module reviews some of the characteristics of learners and of encoding activities that seem to affect how well people can acquire new memories, knowledge, or skills.

Learning Objectives

- Consider what kinds of activities constitute learning.
- Name multiple forms of learning.
- List some individual differences that affect learning.
- Describe the effect of various encoding activities on learning.

18.10.8 Introduction

What do you do when studying for an exam? Do you read your class notes and textbook? Do you try to find a quiet place without distraction? Do you use flash cards to test your knowledge? The choices you make reveal your theory of learning, but there is no reason to limit yourself to your own intuitions. There is a vast science of learning, in which researchers from psychology, education, and neuroscience study basic principles of learning and memory.



Figure 18.10.5: When you study for a test, you incorporate your past knowledge into learning this new knowledge. That is, depending on your previous experiences, you will "learn" the material in different ways. [Image: UBC Learning Commons, https://goo.gl/eT0jvd, CC BY 2.0, https://goo.gl/BRvSA7]

In fact, learning is much broader than you might think. Consider: Is listening to music a form of learning? We know that your brain's response to auditory information changes with your experience with that information, a form of learning called auditory **perceptual learning** (Polley, Steinberg, & Merzenich, 2006). When we exhibit changes in behavior without having intended to learn something, that is called **implicit learning** (Seger, 1994), and when we exhibit changes in our behavior that reveal the influence of past experience even though we are not attempting to use that experience, that is called **implicit memory** (Richardson-Klavehn & Bjork, 1988).

Other well-studied forms of learning include the types of learning that are general across species. We can't ask a slug to learn a poem or a lemur to learn to bat left-handed, but we can assess learning in other ways. For example, we can look for a change in our responses to things when we are repeatedly stimulated. If you live in a house with a grandfather clock, you know that what was once an annoying and intrusive sound is now probably barely audible to you. Similarly, poking an earthworm again and again is likely to lead to a reduction in its retraction from your touch. These phenomena are forms of **nonassociative learning**, in which single repeated exposure leads to a change in behavior (Pinsker, Kupfermann, Castelluci, & Kandel, 1970). When our response lessens with exposure, it is called **habituation**, and when it increases (like it might with a particularly annoying laugh), it is called **sensitization**. Animals can also learn about relationships between things, such as when an alley cat learns that the sound of janitors working in a restaurant precedes the dumping of delicious new garbage (an example of stimulus-stimulus learning called **classical conditioning**), or when a dog learns to roll over to get a treat (a form of stimulus-response learning called **operant conditioning**). These forms of learning will be covered in the module on Conditioning and Learning (http://noba.to/ajxhcqdr).

Here, we'll review some of the conditions that affect learning, with a focus on the type of explicit learning we do when trying to learn something.

18.10.9 Learners

One well-studied and important variable is **working memory** capacity. Working memory describes the form of memory we use to hold onto information temporarily. Working memory is used, for example, to keep track of where we are in the course of a

complicated math problem, and what the relevant outcomes of prior steps in that problem are. Higher scores on working memory measures are predictive of better reasoning skills (Kyllonen & Christal, 1990), reading comprehension (Daneman & Carpenter, 1980), and even better control of attention (Kane, Conway, Hambrick, & Engle, 2008).



Figure **18.10.6**: Research attests that we can hold between 5 and 9 individual pieces of information in our working memory at once. This is partly why in the 1950s Bell Labs developed a 7-digit phone number system. [Image: Diamondmagna, https://goo.gl/xeUxfw, CC BY-SA 3.0, https://goo.gl/eLCn2O]

Anxiety also affects the quality of learning. For example, people with math anxiety have a smaller capacity for remembering math-related information in working memory, such as the results of carrying a digit in arithmetic (Ashcraft & Kirk, 2001). Having students write about their specific anxiety seems to reduce the worry associated with tests and increases performance on math tests (Ramirez & Beilock, 2011).

Another factor to consider is the role of expertise. Though there probably is a finite capacity on our ability to store information (Landauer, 1986), having more knowledge or expertise actually enhances our ability to learn new information. A classic example is comparing a chess master with a chess novice on their ability to learn and remember the positions of pieces on a chessboard (Chase & Simon, 1973). In that experiment, the master remembered the location of many more pieces than the novice, even after only a very short glance. Maybe chess masters are just smarter than the average chess beginner, and have better memory? No: The advantage the expert exhibited only was apparent when the pieces were arranged in a plausible format for an ongoing chess game; when the pieces were placed randomly, both groups did equally poorly. Expertise allowed the master to **chunk** (Simon, 1974) multiple pieces into a smaller number of pieces of information—but only when that information was structured in such a way so as to allow the application of that expertise by exploiting memory of familiar chunks (in this case, groups of chess pieces in familiar configurations).

18.10.10 Encoding Activities

What we do when we're learning is very important. We've all had the experience of reading something and suddenly coming to the realization that we don't remember a single thing, even the sentence that we just read. *How* we go about **encoding** information determines a lot about how much we remember.

Merely intending to learn something is not enough. When a learner actively processes the material, encoding and memory are improved; for example, reading words and evaluating their meaning leads to better learning than reading them and evaluating the way that the words look or sound (Craik & Lockhart, 1972). If you are trying to learn a list of words, just evaluating each word for its part of speech (i.e., noun, verb, adjective) helps you *recall* the words—that is, it helps you remember and write down more of the words later. But it actually impairs your ability to *recognize* the words—to judge on a later list which words are the ones that you studied (Eagle & Leiter, 1964). So this is a case in which **incidental learning**—that is, learning without the intention to learn—is better than **intentional learning**.

Such examples are not particularly rare and are not limited to recognition. Nairne, Pandeirada, and Thompson (2008) showed, for example, that survival processing—thinking about and rating each word in a list for its relevance in a survival scenario—led to much higher recall than just the intention to learn (and also higher, in fact, than other encoding activities that are also known to lead to high levels of recall). Interacting with the material to be learned, thinking about it and processing it in terms of its meaning for something important like survival, improves recall. To process the material you have to attend to it and you also form connections or associations which may imporve recall.

If you are studying for a final exam next week and plan to spend a total of five hours, what is the best way to distribute your study? The evidence is clear that *spacing* one's repetitions apart in time is superior than *massing* them all together (Baddeley & Longman, 1978; Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Melton, 1967).

A similar advantage is evident for the practice of interleaving multiple skills to be learned: For example, baseball batters improved more when they faced a mix of different types of pitches than when they faced the same pitches blocked by type (Hall, Domingues, & Cavazos, 1994). Students also showed better performance on a test when different types of mathematics problems were interleaved rather than grouped during learning (Taylor & Rohrer, 2010).

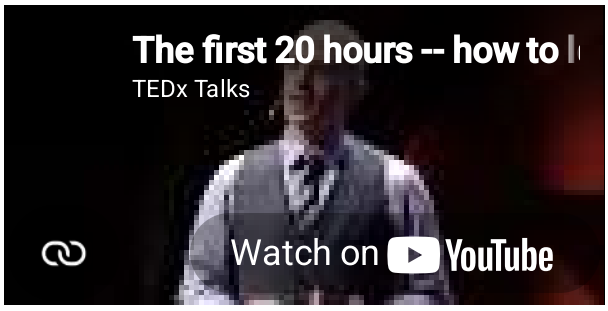
One final factor that merits discussion is the role of testing. Educators and students often think about testing as a way of assessing knowledge, and this is indeed an important use of tests. But tests themselves affect memory, because retrieval is one of the most powerful ways of enhancing learning (Roediger & Butler, 2013). Self-testing is an underutilized and potent means of making learning more durable.

18.10.11 Conclusion

To wrap things up, let's think back to the questions we began the module with. What might you now do differently when preparing for an exam? Hopefully, you will think about testing yourself frequently, developing an accurate sense of what you do and do not know, how you are likely to use the knowledge, and using the scheduling of study sessions to your advantage. If you are learning a new skill or new material, using the scientific study of learning as a basis for the study and practice decisions you make is a good bet.

18.10.11.1 Outside Resources

Video: The First 20 hours – How to Learn Anything - Watch a video by Josh Kaufman about how we can get really good at almost anything with 20 hours of efficient practice.





Video: The Learning Scientists - Terrific YouTube Channel with videos covering such important topics as interleaving, spaced repetition, and retrieval practice.

https://www.youtube.com/channel/UCjbAmxL6GZXiaoXuNE7cIYg

Video: What we learn before we're born - In this video, science writer Annie Murphy Paul answers the question "When does learning begin?" She covers through new research that shows how much we learn in the womb — from the lilt of our native language to our soon-to-be-favorite foods.

https://www.ted.com/talks/annie_murphy_paul_what_we_learn_before_we_re_born

Web: Neuroscience News - This is a science website dedicated to neuroscience research, with this page addressing fascinating new memory research.

http://neurosciencenews.com/neuroscience-terms/memory-research/

Web: The Learning Scientists - A websitecreated by three psychologists who wanted to make scientific research on learning more accessible to students, teachers, and other educators.

http://www.learningscientists.org/

18.10.11.2 Discussion Questions

1. How would you best design a computer program to help someone learn a new foreign language? Think about some of the principles of learning outlined in this module and how those principles could be instantiated in "rules" in a computer program.
2. In what kinds of situations not discussed here might you find a benefit of forgetting on learning?

18.10.11.3 Vocabulary

Chunk

The process of grouping information together using our knowledge.

Classical conditioning

Describes stimulus-stimulus associative learning.

Encoding

The pact of putting information into memory.

Habituation

Occurs when the response to a stimulus decreases with exposure.

Implicit learning

Occurs when we acquire information without intent that we cannot easily express.

Implicit memory

A type of long-term memory that does not require conscious thought to encode. It's the type of memory one makes without intent.

Incidental learning

Any type of learning that happens without the intention to learn.

Intentional learning

Any type of learning that happens when motivated by intention.

Metacognition

Describes the knowledge and skills people have in monitoring and controlling their own learning and memory.

Nonassociative learning

Occurs when a single repeated exposure leads to a change in behavior.

Operant conditioning

Describes stimulus-response associative learning.

Perceptual learning

Occurs when aspects of our perception changes as a function of experience.

Sensitization

Occurs when the response to a stimulus increases with exposure

Transfer-appropriate processing

A principle that states that memory performance is superior when a test taps the same cognitive processes as the original encoding activity.

Working memory

The form of memory we use to hold onto information temporarily, usually for the purposes of manipulation.

18.10.11.4 References

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology*: General, 130, 224–237.

Baddeley, A. D., & Longman, D. J. A. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics*, 21, 627–635.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. O. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, 4, 316–321.

Bjork, R. A. (2011). On the symbiosis of learning, remembering, and forgetting. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A Festschrift in honor of Robert A. Bjork* (pp. 1–22). London, UK: Psychology Press.

Castel, A. D., Benjamin, A. S., Craik, F. I. M., & Watkins, M. J. (2002). The effects of aging on selectivity and control in short-term recall. *Memory & Cognition*, 30, 1078–1085.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55–81.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671–684.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.

Eagle, M., & Leiter, E. (1964). Recall and recognition in intentional and incidental learning. *Journal of Experimental Psychology*, 68, 58–63.

Garavalia, L. S., & Gredler, M. E. (2002). Prior achievement, aptitude, and use of learning strategies as predictors of college student achievement. *College Student Journal*, 36, 616–626.

Hall, K. G., Domingues, D. A., & Cavazos, R. (1994). Contextual interference effects with skilled baseball players. *Perceptual and Motor Skills*, 78, 835–841.

Heyer, A. W., Jr., & O'Kelly, L. I. (1949). Studies in motivation and retention: II. Retention of nonsense syllables learned under different degrees of motivation. *Journal of Psychology: Interdisciplinary and Applied*, 27, 143–152.

Jenkins, J. J. (1979). Four points to remember: A tetrahedral model of memory experiments. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing and human memory* (pp. 429–446). Hillsdale, NJ: Erlbaum.

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2008). Variation in working memory capacity as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in Working Memory* (pp. 22–48). New York, NY: Oxford University Press.

Kimball, D. R., Smith, T. A., & Muntean, W. J. (2012). Does delaying judgments of learning really improve the efficacy of study decisions? Not so much. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 923–954.

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 32, 609–622.*

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity. *Intelligence*, 14, 389–433.

Landauer, T. K. (1986). How much do people remember? Some estimates of the quantity of learned information in long-term memory. *Cognitive Science, 10*, 477–493.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). London: Academic Press.

Melton, A. W. (1967). Repetition and retrieval from memory. *Science*, 158, 532.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 16, 519–533.

Mullin, P. A., Herrmann, D. J., & Searleman, A. (1993). Forgotten variables in memory theory and research. *Memory*, 1, 43–64.

Nairne, J. S., Pandeirada, J. N. S., & Thompson, S. R. (2008). Adaptive memory: the comparative value of survival processing. *Psychological Science*, 19, 176–180.

Pinsker, H., Kupfermann, I., Castelluci, V., & Kandel, E. (1970). Habituation and dishabituation of the gill-withdrawal reflex in Aplysia. *Science*, 167, 1740–1742.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, 95, 667–686.

Polley, D. B., Steinberg, E. E., & Merzenich, M. M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. *The Journal of Neuroscience*, 26, 4970–4982.

Ramirez, G., & Beilock, S. L. (2011). Writing about testing worries boosts exam performance in the classroom. *Science*, 331, 211–213.

Richardson-Klavehn, A. & Bjork, R.A. (1988). Measures of memory. *Annual Review of Psychology*, 39, 475–543.

Roediger, H. L., & Butler, A.C. (2013). Retrieval practice (testing) effect. In H. L. Pashler (Ed.), *Encyclopedia of the mind*. Los Angeles, CA: Sage Publishing Co.

Seger, C. A. (1994). Implicit learning. *Psychological Bulletin*, 115, 163–196.

Simon, H. A. (1974). How big is a chunk? *Science*, 4124, 482–488.

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, 24, 837–848.

Tullis, J. G., & Benjamin, A. S. (2012). Consequences of restudy choices in younger and older learners. *Psychonomic Bulletin & Review*, 19, 743–749.

Tullis, J. G., & Benjamin, A. S. (2011). On the effectiveness of self-paced learning. *Journal of Memory and Language*, 64, 109–118.

18.10.11.5 Attributions

Adapted by Kenneth A. Koenigshofer, PhD, from Benjamin, A. (2021). Factors influencing learning. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology.* Champaign, IL: DEF publishers. Retrieved from http://noba.to/rnxyg6wp

18.10.11.6 **Authors**



- Aaron Benjamin is Professor of Psychology at the University of Illinois at Urbana-Champaign. He is President of the International Association for Metacognition and a member of the Governing Board of the Psychonomic Society, and has been an Associate Editor of the Journal of Experimental Psychology: Learning, Memory, and Cognition. He conducts research on memory, metamemory, and decision-making.

18.10.11.7 Creative Commons License

Factors Influencing Learning by Aaron Benjamin is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Permissions beyond the scope of this license may be available in our Licensing Agreement

This page titled 18.10: Chapter 10- Psychological Factors Affecting Learning, Memory, and Forgetting is shared under a mixed license and was authored, remixed, and/or curated by Kenenth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

18.11: Chapter 14- Traditional Models of Human Intelligence

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Discuss two common tests for measuring intelligence.
2. Describe at least one “type” of intelligence.
3. Describe Carroll's three-stratum theory of intelligence and name the factor at the top level in this theory.
4. Discuss intelligence in simple terms
5. Name a brain network thought to be associated with intelligence

18.11.0.1 Overview

The development of tests to measure intelligence has had a major impact on the development of ideas about the nature and structure of human intelligence, and its biological basis in the brain. Most theories of human intelligence are based on data derived from intelligence tests which is analyzed using factor analysis, a mathematical method for analyzing patterns of correlations among different measures of mental abilities. In module 14.2, we have already discussed how this method, invented and used by Spearman (1904), revealed the "g" factor in human intelligence.

To understand current thinking and research about the biological basis of human intelligence, it is essential to gain at least a general familiarity with the major theoretical models of human intelligence psychologists have developed. The theories we examine in this module are based to a large extent on intelligence testing and factor analysis, while others are more intuitive. This section introduces key historical figures, major theories of intelligence, and common assessment strategies used to measure human intelligence.

In section 14.2, we discussed a number of enduring, across-generation, universal regularities of the environment which have been incorporated by evolution into brain organization and intelligence. As described in that section, these enduring facts about how the world works include innate, genetically internalized information about objects in three-dimensional space, the passage of time, daily cycles of light and dark, causality relations (forming basis for causal logic and inference), similarity relations (leading to category formation and categorical logic and inference), and predictive relations, based on covariation of events, allowing human and animal brains to mentally project the organism into future time. All of these invariant properties of the world must be included in the brain's neural models or cognitive maps of the world if the brain is to effectively guide adaptive behavior.

When we examine the traditional models of intelligence in this section, you will recognize that each focuses on only one, or a few of the facets of intelligence discussed in the evolutionary approach taken in section 14.2. In a sense, each theory discussed in this section is akin to the fable of the blind men trying to describe an elephant. Each blind man only knows that part of the elephant which he happens to feel and so each man has a different and incomplete understanding of the whole. Likewise, each theory of intelligence focuses on only part of the complex of processes that we collectively refer to as "intelligence." Nevertheless, each theory makes a contribution, and each, in one or more ways, is related to the evolutionary discussion in section 14.2.

For example, as you will see, emotional intelligence, including Gardner's intra- and inter-personal intelligence, is related to neural representations of the contingencies of the social environment--brain mechanisms for which are the focus of the new field of **social cognitive neuroscience.** Gardner's multiple intelligences include spatial intelligence related to representation of objects in three-dimensional space, abilities which require portions of **parietal cortex** and **hippocampus**. At Level III in Carroll's theory of intelligence is "g," general intelligence, related to representations of causal, similarity, and predictive relations, likely involving the **frontoparietal network** (Jung & Haier, 2007). Of these theories, the first, Carroll's three-stratum theory of human intelligence is by far the most widely accepted and most productive in terms of explanatory power and empirical evidence. With this background, we are better prepared to examine the traditional models of intelligence in this module and, perhaps more importantly for this course, we will be better prepared to understand the biological bases of intelligence and thinking, a primary focus of this chapter.

18.11.0.1 Introduction

Every year hundreds of grade school students converge on Washington, D.C., for the annual Scripps National Spelling Bee. The "bee" is an elite event in which children as young as 8 square off to spell words like "cymotrichous" and "appoggiatura." Most people who watch the bee think of these kids as being "smart" and you likely agree with this description.



Figure 18.11.1: A participant in the Scripps National Spelling Bee. (CC BY-NC 2.0; Scripps National Spelling Bee via Flicrk)

What makes a person intelligent? Is it heredity (two of the 2014 contestants in the bee have siblings who have previously won) (National Spelling Bee, 2014a)? Is it interest and motivation (the most frequently listed favorite subject among spelling bee competitors is math)(NSB, 2014b)? By the end of the module you should be able to define intelligence, discuss methods for measuring intelligence, and describe theories of intelligence. In addition, we will tackle the politically thorny issue of whether there are differences in intelligence between groups such as men and women. As you read through this module, note that we discuss possible links between each theory of intelligence and the material from the previous module on adaptation, evolution and the brain mechanisms of intelligence. Recall that all of the information processing done by the brain, including that involving what we term intelligence and thinking, can have no effect at all on the outside world unless that neural activity converges onto and acts on the motor neurons in the spinal cord and medulla which stimulate the muscles to generate movement, behavior. Intelligence and cognition, as discussed in the last module, are part of the elaborate control systems which guide movement into adaptive patterns of behavior.

18.11.0.1 Defining and Measuring Intelligence

When you think of "smart people" you likely have an intuitive sense of the qualities that make them intelligent. Maybe you think they have a good memory, or that they can think quickly, or that they simply know a whole lot of information. Indeed, people who exhibit such qualities appear very intelligent. That said, it seems that intelligence must be more than simply knowing facts and being able to remember them. One point in favor of this argument is the idea of animal intelligence. It will come as no surprise to you that a dog, which can learn commands and tricks seems smarter than a snake that cannot. In fact, researchers and lay people generally agree with one another that primates—monkeys and apes (including humans)—are among the most intelligent animals (see comparisons among species on cognitive abilities in module 10.1). Apes such as chimpanzees are capable of complex problem solving and sophisticated communication (Kohler, 1924).

Scientists point to the social nature of primates as one evolutionary source of their intelligence. Primates live together in troops or family groups and are, therefore, highly social creatures. As such, primates tend to have brains that are better developed for communication and long term thinking than most other animals. For instance, the complex social environment has led primates to develop deception, altruism, numerical concepts, and "theory of mind" (a sense of the self as a unique individual separate from others in the group and understanding that others have minds; Gallup, 1982; Hauser, MacNeilage & Ware, 1996). [Also see module on Theory of Mind later in this chapter and at http://noba.to/a8wpytg3]

The question of what constitutes human intelligence is one of the oldest inquiries in psychology. When we talk about intelligence we typically mean intellectual ability. This broadly encompasses the ability to learn, remember and use new information, to solve problems and to adapt to novel situations. As discussed in module 10.1, an early scholar of intelligence, Charles Spearman, proposed that intelligence was one thing, a "general factor" sometimes known as simply "g." He based this conclusion on the

observation that people who perform well in one intellectual area such as verbal ability also tend to perform well in other areas such as logic and reasoning (Spearman, 1904).

Francis Galton, a contemporary of Spearman and a cousin of Charles Darwin, was among those who pioneered psychological measurement (Hunt, 2009). Galton was particularly interested in intelligence, which he thought was heritable in much the same way that height and eye color are. He conceived of several rudimentary methods for assessing whether his hypothesis was true. For example, he carefully tracked the family tree of the top-scoring Cambridge students over the previous 40 years. Although he found specific families disproportionately produced top scholars, intellectual achievement could still be the product of economic status, family culture or other non-genetic factors. Galton was also, possibly, the first to popularize the idea that the heritability of psychological traits could be studied by looking at identical and fraternal twins. Although his methods were crude by modern standards, Galton established intelligence as a variable that could be measured (Hunt, 2009).



Figure 18.11.2: Alfred Binet (public domain)

The person best known for formally pioneering the measurement of intellectual ability is Alfred Binet. Like Galton, Binet was fascinated by individual differences in intelligence. For instance, he blindfolded chess players and saw that some of them had the ability to continue playing using only their memory (most likely involving use of a type of cognitive map, perhaps involving **parietal and prefrontal cortex**) to keep the many positions of the pieces in mind (Binet, 1894). Binet was particularly interested in the development of intelligence, a fascination that led him to observe children carefully in the classroom setting.

Along with his colleague Theodore Simon, Binet created a test of children's intellectual capacity. They created individual test items that should be answerable by children of given ages. For instance, a child who is three should be able to point to her mouth and eyes, a child who is nine should be able to name the months of the year in order, and a twelve year old ought to be able to name sixty words in three minutes. Their assessment became the first "IQ test."

Intelligence tests

Some examples of the types of items you might see on an intelligence test.

1. Which of the following is most similar to 1313323?
   A. ACACCBC
   B. CACAABC
   C. ABABBCA
   D. ACACCDC
2. Jenny has some chocolates. She eats two and gives half of the remainder to Lisa. If Lisa Has six chocolates how many does Jenny have in the beginning?
   A. 6
   B. 12
   C. 14
   D. 18
3. Which of the following items is not like the others in the list: duck, raft, canoe, stone, rubber ball

A. Duck
B. Canoe
C. Stone
D. Rubber ball

4. What do steam and ice have in common?

A. They can both harm skin
B. They are both made from water
C. They are both found in the kitchen
D. They are both the products of water at extreme temperatures

"IQ" or "intelligence quotient" is a name given to the score of the Binet-Simon test. The score is derived by dividing a child's mental age (the score from the test) by their chronological age to create an overall quotient. These days, the phrase "IQ" does not apply specifically to the Binet-Simon test and is used to generally denote intelligence or a score on any intelligence test. In the early 1900s the Binet-Simon test was adapted by a Stanford professor named Lewis Terman to create what is, perhaps, the most famous intelligence test in the world, the Stanford-Binet (Terman, 1916). The major advantage of this new test was that it was standardized. Based on a large sample of children Terman was able to plot the scores in a normal distribution, shaped like a "bell curve" (see Fig. 1). To understand a normal distribution think about the height of people. Most people are average in height with relatively fewer being tall or short, and fewer still being extremely tall or extremely short. Terman (1916) laid out intelligence scores in exactly the same way, allowing for easy and reliable categorizations and comparisons between individuals.

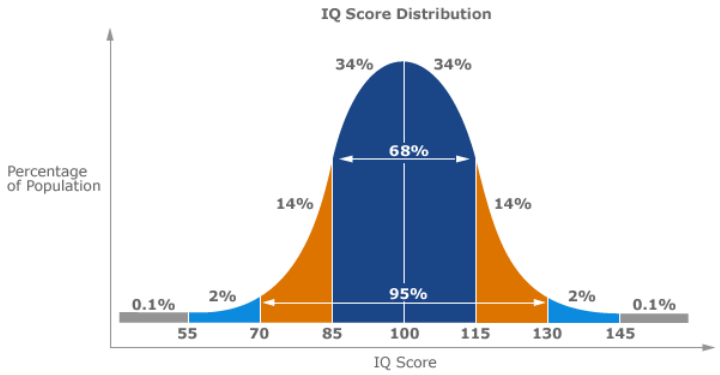


Figure 18.11.3: A graph of IQ Score Distribution: 68% of the population will score between 85 - 115, 95% will score between 70 - 130, 2% will score below 70, 2% will score above 130, 0.1% will score below 55 and 0.1% will score above 145.

Looking at another modern intelligence test—the Wechsler Adult Intelligence Scale (WAIS)—can provide clues to a definition of intelligence itself. Motivated by several criticisms of the Stanford-Binet test, psychologist David Wechsler sought to create a superior measure of intelligence. He was critical of the way that the Stanford-Binet relied so heavily on verbal ability and was also suspicious of using a single score to capture all of intelligence. To address these issues Wechsler created a test that tapped a wide range of intellectual abilities. This understanding of intelligence—that it is made up of a pool of specific abilities—is a notable departure from Spearman's concept of general intelligence. The WAIS assesses people's ability to remember, compute, understand language, reason well, and process information quickly (Wechsler, 1955). However, as we will see below, these two approaches were integrated by Carroll (1993) in his model of intelligence.

One interesting by-product of measuring intelligence for so many years is that we can chart changes over time. Over the last 80 years we have been measuring human intelligence, when new waves of people are asked to take older intelligence tests they tend to outperform the original sample from years ago on which the test was normed. This gain in measured average intelligence is known as the "Flynn Effect," named after James Flynn, the researcher who first identified it (Flynn, 1987). Several hypotheses have been put forth to explain the Flynn Effect including better nutrition (healthier brains!), greater familiarity with testing in general, and more exposure to visual stimuli. Today, there is no perfect agreement among psychological researchers about the causes of these increases in average scores on intelligence tests over the past 80 years. Keep in mind that these intelligence tests were originally designed to predict school performance. Could it be that improvements over the years in public education, or some other social variable, may, at least in part, account for the Flynn Effect?

18.11.1 Types of Intelligence

David Wechsler's approach to testing intellectual ability was based on the fundamental idea that there are many aspects to intelligence. Other scholars have echoed this idea by going so far as to suggest that there are actually even different types of intelligence. You likely have heard distinctions made between "street smarts" and "book learning." The former refers to practical wisdom accumulated through experience while the latter indicates formal education. A person high in street smarts might have a superior ability to catch a person in a lie, to persuade others, or to think quickly under pressure. A person high in book learning, by contrast, might have a large vocabulary and be able to remember a large number of facts. Although psychologists don't use street smarts or book smarts as professional terms some do believe that there are different types of intelligence.

Carroll's Three-Stratum Model

There are many ways to parse apart the concept of intelligence. Many scholars believe that a theory proposed by Carroll (1993) provides the best and most comprehensive model of human intelligence. Carroll divided human intelligence into three levels, or strata, descending from the most abstract down to the most specific (see Figures 10.2.4 and 10.2.5). Carroll called the highest level (stratum III) the general intelligence factor "g," following Spearman's (1904) original concept of a general intelligence factor, evolutionary origins of which were discussed in module 10.1. Below stratum III were more specific stratum II categories which are different subsets of "g" such as fluid intelligence (Gf), crystallized intelligence (Gc), broad visual perception (Gv), processing speed, and a number of others (see Figure 10.2.5). Each of these, in turn, can be sub-divided into very specific components such as spatial scanning, reaction time, and word fluency.

Thinking of intelligence as Carroll (1993) does, as a collection of specific mental abilities, has helped researchers conceptualize this topic in new ways. For example, Horn and Cattell (1966) were first to distinguish between "fluid" and "crystallized" intelligence, both of which are on stratum II of Carroll's model. Fluid intelligence refers to basic processes of reasoning and other mental activities that are only minimally dependent upon prior learning and experience (such as education). Fluid intelligence is the ability to think and reason flexibly and abstractly to solve problems and encompasses the ability to see complex relationships. This is closest to the concept of general intelligence discussed in module 10.1 and most likely involves the **parietofrontal network** described in that module.

Crystallized intelligence, on the other hand, refers to learned procedures and knowledge and includes the ability to use language, and skills and knowledge accumulated from experience (see chapter on learning and memory for discussion of neural mechanisms of learning). Crystallized intelligence is characterized as acquired knowledge and the ability to retrieve it. Fluid intelligence helps you tackle complex, abstract challenges in your daily life, whereas crystallized intelligence helps you overcome concrete, straightforward problems (Cattell, 1963). The latter increases with age. In general, older people have a relatively superior store of knowledge that can be put to use to solve problems.

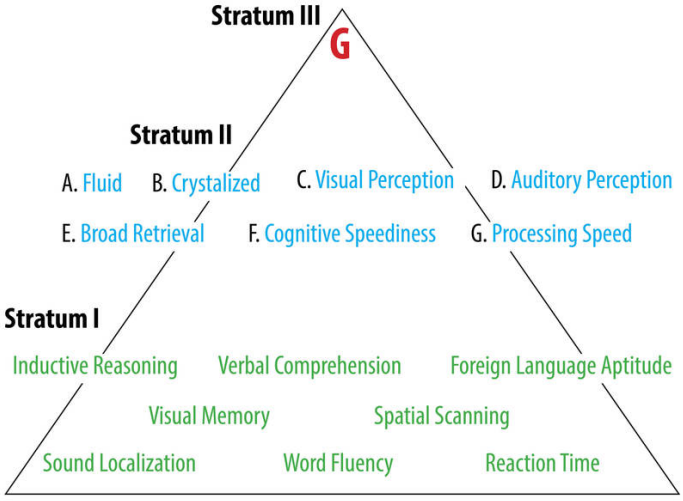


Figure **18.11.4**: A Simplified Summary of Carroll's Model of Intelligence. Stratum I: inductive reasoning, verbal comprehension, foreign languange aptitude, visual memory, spatial scanning, sound localization, word fluency and reaction time. Stratum II: fluid, crystallized, visual perception, auditory perception, broad retrieval, cognitive speediness, processing speed. Stratum III: G (i.e. g, general intelligence).

Carroll's **three-stratum theory** is based on a factor-analytic study of the correlation of individual-difference variables from data such as psychological tests, school grades, and competence ratings from more than 460 datasets. These analyses suggested a three-layered model where each layer accounts for the variations in the correlations within the previous layer.

The three layers (strata) are defined as representing narrow, broad, and general cognitive ability. The factors describe stable and observable differences among individuals in their performance of cognitive tasks. Carroll argues further that they are not mere artifacts of a mathematical process, but likely reflect **neurophysiological factors that explain the differences in ability (e.g., nerve firing rates, processing efficiency as proposed by Haier and discussed below, conduction velocity, etc).**

Carroll's taxonomy of intelligence distinguishes between level factors and speed factors. The tasks dependent upon level factors can be sorted by difficulty and individuals' scores are differentiated by whether they have acquired the skill to perform the tasks. Tasks that contribute to speed factors are distinguished by the relative speed with which individuals can complete them. Carroll suggests that the distinction between level and speed factors may be the broadest taxonomy of cognitive tasks that can be offered.

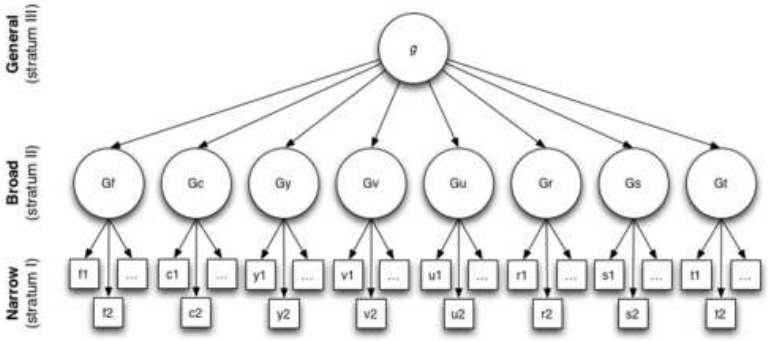


Figure 18.11.5: Carroll's three-stratum model. Key: fluid intelligence (Gf), crystallized intelligence (Gc), general memory and learning (Gy), broad visual perception (Gv), broad auditory perception (Gu), broad retrieval ability (Gr), broad cognitive speediness (Gs), and processing speed (Gt). Carroll regarded the broad abilities as different "flavors" of *g*.

Gardener's Multiple Intelligences Theory

Howard Gardner, a Harvard psychologist and former student of Erik Erikson, is another figure in psychology who is well-known for championing the notion that there are different types of intelligence. In Gardner's theory, each person possesses at least eight intelligences. Among these eight intelligences, a person typically excels in some and falters in others (Gardner, 1983). Gardner's theory is appropriately, called "multiple intelligences." Gardner's theory is based on the idea that people process information through different "channels" and these are relatively independent of one another. He has identified 8 common intelligences including 1) logic-math, 2) spatial, 3) music-rhythm, 4) verbal-linguistic, 5) bodily-kinesthetic, 6) interpersonal, 7) intrapersonal, and 8) naturalistic (Gardner, 1985). Many people are attracted to Gardner's theory because it suggests that people each learn in unique ways. There are now many Gardner-influenced schools in the world. Gardener's idea of different intelligences involving different "channels" suggests separate specialized brain mechanisms for these different cognitive abilities. This is consistent with what evolutionary psychologists refer to as a **modular model** of the mind, brain, and intelligence. On this view the mind/brain consists of a large collection of specialized information processing modules or mini-computers each evolved to process a particular kind of information needed to solve a particular category of adaptive problem (Ermer, et al., 2007). This model is consistent with the concept of **localization of function**, the hypothesis that different psychological functions are anatomically localized to different areas of the brain.

Gardner's Mulitiple Intelligences

Intelligence Type	Characteristics	Representative Career
Linguistic intelligence	Perceives different functions of language, different sounds and meanings of words, may easily learn multiple languages	Journalist, novelist, poet, teacher
Logical-mathematical intelligence	Capable of seeing numerical patterns, strong ability to use reason and logic	Scientist, mathematician
Musical intelligence	Understands and appreciates rhythm, pitch, and tone; may play multiple instruments or perform as a vocalist	Composer, performer
Bodily kinesthetic intelligence	High ability to control the movements of the body and use the body to perform various physical tasks	Dancer, athlete, athletic coach, yoga instructor
Spatial intelligence	Ability to perceive the relationship between objects and how they move in space	Choreographer, sculptor, architect, aviator, sailor
Interpersonal intelligence	Ability to understand and be sensitive to the various emotional states of others	Counselor, social worker, salesperson
Intrapersonal intelligence	Ability to access personal feelings and motivations, and use them to direct behavior and reach personal goals	Key component of personal success over time
Naturalist intelligence	High capacity to appreciate the natural world and interact with the species within it	Biologist, ecologist, environmentalist

Note that Gardener's Bodily kinesthetic intelligence and Spatial intelligence are likely to both involve the **parietal lobe**, while his Linguistic intelligence involves **Broca's and Wernicke's areas**, prominent language areas of the brain to be discussed later in this chapter. It is also likely that Gardener's Logical-mathematical intelligence requires processing in the **frontoparietal network** (Jung & Haier, 2007).

It has been suggested that Gardner simply relabeled what other theorists called “cognitive styles” as “intelligences” (Morgan, 1996). Furthermore, developing traditional measures of Gardner’s intelligences is extremely difficult (Furnham, 2009; Gardner & Moran, 2006; Klein, 1997).

Gardner’s inter- and intrapersonal intelligences are often combined into a single type: emotional intelligence.

Emotional Intelligence

Emotional intelligence encompasses the ability to understand the emotions of yourself and others, show empathy, understand social relationships and cues, and regulate your own emotions and respond in culturally appropriate ways (Parker, Saklofske, & Stough, 2009). People with high emotional intelligence typically have well-developed social skills and Goleman claims it can be a better predictor of success than traditional intelligence (Goleman, 1995). However, emotional intelligence is difficult to measure and study empirically, with some researchers pointing out inconsistencies in how it is defined and described (Locke, 2005; Mayer, Salovey, & Caruso, 2004).

Regardless of the specific definition of emotional intelligence, studies have shown a link between this concept and job performance (Lopes, Grewal, Kadis, Gall, & Salovey, 2006). In fact, emotional intelligence is similar to more traditional notions of cognitive intelligence with regards to workplace success.

Emotional intelligence, as defined by Parker, et al. (2009), likely involves the **anterior cingulate cortex (ACC)**, the **limbic system**, and the **prefrontal cortex** and the connections between them (Stevens, et al., 2011).

Sternberg's Triarchic Model of Human Intelligence

Robert Sternberg developed another theory of intelligence, which he titled the **triarchic theory of intelligence** because it sees intelligence as comprised of three parts (Sternberg, 1988): practical, creative, and analytical intelligence (Figure).

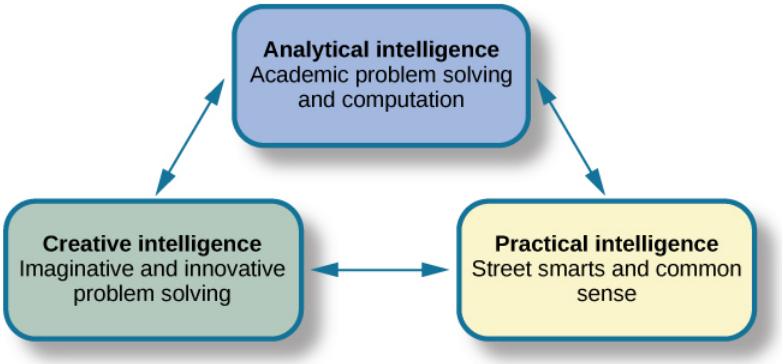


Figure 18.11.6: Sternberg's theory identifies three types of intelligence: practical (street smarts and common sense), creative (imaginative and innovative problem solving), and analytical (academic problem solving and computation). (What are Intelligence and Creativity? by OpenStax Colleg licensed CC BY-NC 4.0 via OER Commons)

Practical intelligence, as proposed by Sternberg, is sometimes compared to "street smarts." Being practical means you find solutions that work in your everyday life by applying knowledge based on your experiences. This type of intelligence appears to be separate from traditional understanding of IQ; individuals who score high in practical intelligence may or may not have comparable scores in creative and analytical intelligence (Sternberg, 1988).

Analytical intelligence is closely aligned with academic problem solving and computations. Sternberg says that analytical intelligence is demonstrated by an ability to analyze, evaluate, judge, compare, and contrast.

Creative intelligence is marked by inventing or imagining a solution to a problem or situation. Creativity in this realm can include finding a novel solution to an unexpected problem or producing a beautiful work of art or a well-developed short story. Imagine for a moment that you are camping in the woods with some friends and realize that you've forgotten your camp coffee pot. The person in your group who figures out a way to successfully brew coffee for everyone would be credited as having higher creative intelligence.

18.11.1.1 Intelligence and Creativity

Creativity is the ability to generate, create, or discover new ideas, solutions, and possibilities. Very creative people often have intense knowledge about something, work on it for years, look at novel solutions, seek out the advice and help of other experts, and take risks. Although creativity is often associated with the arts, it can be found in every area of life, from the way you decorate your residence to a new way of understanding how a cell works.

Dr. Tom Steitz, the Sterling Professor of Biochemistry and Biophysics at Yale University, has spent his career looking at the structure and specific aspects of RNA molecules and how their interactions could help produce antibiotics and ward off diseases. As a result of his lifetime of work, he won the Nobel Prize in Chemistry in 2009. He wrote, "Looking back over the development and progress of my career in science, I am reminded how vitally important good mentorship is in the early stages of one's career development and constant face-to-face conversations, debate and discussions with colleagues at all stages of research. Outstanding discoveries, insights and developments do not happen in a vacuum" (Steitz, 2010, para. 39). Based on Steitz's comment, it becomes clear that someone's creativity, although an individual strength, benefits from interactions with others.

Creativity is often assessed as a function of one's ability to engage in divergent thinking. Divergent thinking can be described as thinking "outside the box;" it allows an individual to arrive at unique, multiple solutions to a given problem. In contrast, convergent thinking describes the ability to provide a correct, well-established answer or standard solution to a problem (Cropley, 2006; Gilford, 1967).

18.11.1.2 Brain Correlates of Intelligence and Creativity

Jung and Haier (2013) report a number of brain correlates of intelligence and creativity. However, they argue against the idea of one brain area for one cognitive function. Instead, as discussed in the prior module, they argue that brain networks involving multiple brain areas are involved in cognition, especially in complex psychological processes such as intelligence and creativity. Nevertheless, they recognize that brain injury and lesion studies reveal brain structures that are necessary, though not sufficient, for certain psychological functions. They give three examples: 1) Phineas Gage, who survived an iron rod passing through his **frontal lobe** resulting in personality and emotional changes as well as impaired judgement and loss of many social inhibitions; 2) "Tan," whose brain damage led to identification of Broca's area for language expression; and 3) H.M., whose bilateral surgical removal of **temporal lobe** structures including **hippocampus** revealed the role of hippocampus and related structures in formation of new long-term explicit memories and their retrieval.

Within this context, Jung and Haier (2013) note some interesting observations from post-mortem examination of the brain of the famous theoretical physicist, Albert Einstein (whose work led to the equation, $E=mc^2$), and what it might suggest about brain mechanisms in creativity. Einstein's brain was unremarkable in many ways. Its size and weight were within the normal range for a man of his age, and frontal and temporal lobe morphology and corpus callosum area were no different from control brains. However, there was one pronounced difference. According to Jung and Haier, Einstein's brain was missing the parietal operculum, the typical location of the secondary somatosensory cortex, resulting in a larger inferior parietal lobule. In Einstein's brain, the **inferior parietal lobule** was approximately 15% wider than in the brains of normal controls. According to Jung and Haier, this region of brain is associated with "visuospatial cognition, mathematical reasoning, and imagery of movement . . . and its expansion was noted in other cases of prominent physicists and mathematicians." They add that further examination of this area of Einstein's brain revealed that rather than more neurons, this region of his brain had a much larger number of **glial cells**, which provide nutrition to neurons, perhaps indicating an unusually large amount of activity among neurons in this region of his brain.

Significantly, as described in the prior module, **parietal cortex** has strong linkages with **prefrontal cortex** forming a **frontoparietal network**: the **inferior parietal lobule** is primarily **connected with dorsolateral prefrontal cortex** (Bruner, 2010), associated, in part, with abilities for abstract thought, while **upper parietal regions**, according to Bruner, as discussed in module 14.2, are associated in the literature with functions such as abstract representation, internal mental images, “imagined world[s],. . . and thought experiment” (i.e., imagination). Jung and Haier detail another study of Einstein's **right prefrontal association cortex**, where researchers found **greater packing density of neurons** (same number of neurons in a smaller space), which was interpreted as **shorter conduction times between cortical neurons in Einstein's brain** compared to control brains. Jung and Haier conclude that **Einstein's brain differed from controls in the frontoparietal network**. These authors have proposed that the **frontoparietal network is crucial to human intelligence**; furthermore they hypothesized that differences among people in the efficiency of neural communication between the frontal and parietal regions of cortex accounts for differences in intelligence in humans (Jung & Haier, 2007). In part, this idea is based on their finding that **high IQ people** show **less activity** in these brain areas during a complex cognitive task, while lower IQ people show more brain activity, suggesting that high IQ is related to **efficiency** in neural information processing operations. Moreover, higher IQ and ability for abstraction are both inversely correlated with cerebral glucose metabolic rate (Haier et al., 1988, 1992, 2003, 2004), suggesting an efficiency model of individual differences in *g* in which superior **ability for abstraction increases processing efficiency**. In their **Parietal-Frontal Integration Theory (P-FIT)** of the neural basis of intelligence, after sensory processing, information "is then fed forward to the **angular, supramarginal, and inferior parietal cortices**, wherein structural symbolism and/or abstraction are generated and manipulated. The **parietal cortex** then **interacts with frontal regions** that serve to **hypothesis test various solutions** to a given problem." They add that "the **anterior cingulate** is involved in response selection as well as inhibition of competing responses. This process is critically dependent on the fidelity of underlying **white matter** needed to facilitate rapid and error-free transmission of data **between frontal and parietal lobes**" (Jung & Haier, 2013, p. 239). They also note that research in genetics shows that "intelligence and brain structure (i.e., gray and white matter) share common genes" (p. 240).

Regarding creativity specifically, these authors refer to a theory by Flaherty (2005) which proposes a **frontotemporal system** driven by **dopaminergic limbic activity** which provides the drive for creative expression whether art, music, writing, science, etc. and as measured by tests of divergent thinking. Jung and Haier (2013) explain that the **temporal lobe normally inhibits the frontal lobe** so that lesion or mild dysfunction of the temporal lobe releases activity from the frontal lobe by disinhibition causing increased interactions of frontal lobe with other brain regions, sometimes leading to increased creative outputs from neurological

patients with left side damage. They argue that this and other data from "three structural studies point to a decidedly left lateralized, frontosubcortical, and disinhibitory network of brain regions underlying creative cognition and achievement" (p. 244). They add that this model, which still requires much more empirical investigation, "appears to include the frontal and temporal lobes, with cortical “tone” being modulated via interactions between the frontal lobes, basal ganglia and thalamus (part of the dopamine system) through white-matter pathways" (p. 244). Although this model is speculative for such a complex form of cognition as creativity, it can guide continuing research into how humans develop creative intellectual and artistic products.

Inferior parietal lobule

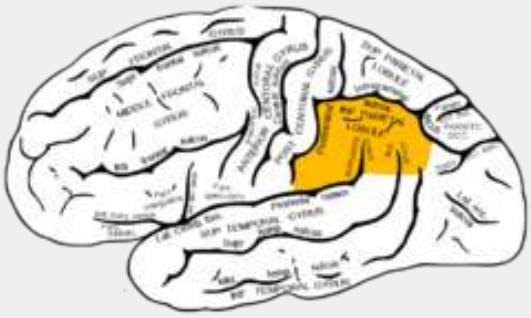


Figure **18.11.7**: Lateral surface of left cerebral hemisphere, viewed from the side. (Inferior parietal lobule is shown in orange.)

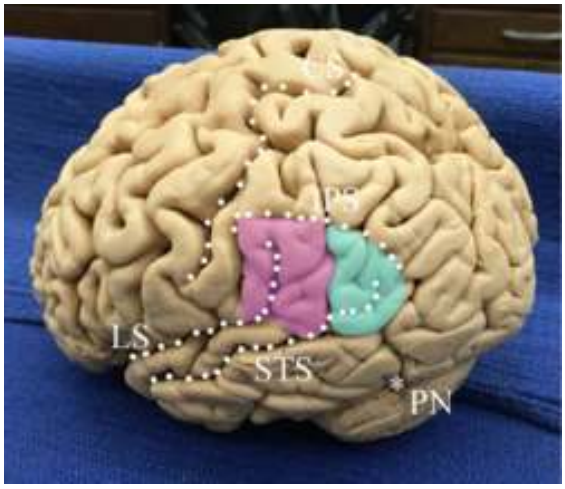


Figure **18.11.8**: Superficial anatomy of the inferior parietal lobule. (Images from Wikipedia, Inferior Parietal Lobule, retrieved 9/30/21). Purple: Supramarginal gyrus. Blue: Angular gyrus. **LS**: Lateral sulcus (Sylvian fissure). **CS**: Central sulcus. **IPS**: Intraparietal sulcus. **STS**:Superior temporal sulcus. **PN**: Preoccipital notch.

18.11.1.3 **Summary**

Intelligence is a complex concept involving multiple mental abilities, including, according to Carroll, Spearman's g factor, itself composed of a number of subtypes identified by mathematical analysis of patterns of correlations among scores on different cognitive tasks. Additional models of human intelligence include Gardner's multiple intelligences, emotional intelligence, and Sternberg's triarchic theory of intelligence, each consistent with the view that intelligence is comprised of many interacting factors. Creativity seems to be another facet of intelligence. It may be explained as arising from imagination involving visual-like mental manipulations to explore alternative actions and their probable outcomes (see discussion module 10.1). One way to understand why there is such diversity in conceptions and theories of intelligence among psychologists is that each is focusing on only one or a few aspects of the multiple processes the brain engages in when it generates neural models of the world to guide adaptive behavior. Recall that module 10.1 on Intelligence, Cognition, Language, and Adaptation included discussion of many complex properties of the social and physical environment that must be neurally modeled by the brain in order for human or animal to successfully navigate the social and physical environments. Each of the traditional theories of human intelligence discussed in this module involve representations of one or only a few of those properties of the social and physical environment. The evolutionary analysis

in module 10.1 can unify the divergent models of intelligence described in this module by showing how each focuses on different sets of components of intelligence required for the construction of accurate mental/neural models or "cognitive maps" (Behrens, et al., 2018; Tolman, 1948) of the biologically significant facts of the physical and social worlds in order to guide behavior toward successful adaptation.

18.11.1.4 Review Questions

1. Fluid intelligence is characterized by _____.
    1. being able to recall information
    2. being able to create artistic pieces
    3. being able to understand and communicate with different cultures
    4. being able to see complex relationships and solve problems
2. Which of the following is not one of Gardner's Multiple Intelligences?
    1. creative
    2. spatial
    3. linguistic
    4. musical
3. Which theorist put forth the triarchic theory of intelligence?
    1. Goleman
    2. Gardner
    3. Sternberg
    4. Steitz
4. When you are examining data to look for trends, which type of intelligence are you using most?
    1. practical
    2. analytical
    3. emotional
    4. creative

18.11.1.5 References

Bar-On, R. (2006). The Bar-On model of emotional-social intelligence (ESI). *Psicometha, 18* (Suppl.), 13–25.

Binet, A. (1894). *Psychologie des grands calculateurs et joueurs d'échecs*. Paris: Librairie Hachette.

Bouchard, T.J. (2004). Genetic influence on human psychological traits - A survey. *Current Directions in Psychological Science 13* (4), 148–151.

Bruner, E. (2010). Morphological differences in the parietal lobes within the human genus. *Current Anthropology*, *51* (S1), S77-S88.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, England:Cambridge University Press.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, England:Cambridge University Press.

Ermer, E., Cosmides, L, and Tooby J. (2007). Functional Specialization and the Adaptationist Program. In S.W. Gangestad and J. A. Simpson (Eds.), *Evolution of Mind: Fundamental Questions and Controversies.* New York: The Guilford Press

Flynn J. R. (1987). "Massive IQ gains in 14 nations: What IQ tests really measure". *Psychological Bulletin 101*, 171–191.

Gallup, G. G. (1982). Self-awareness and the emergence of mind in primates. *American Journal of Primatology, 2* (3), 237-248.

Gardner, H. (1985). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Haier, R. J., Siegel Jr, B. V., Nuechterlein, K. H., Hazlett, E., Wu, J. C., Paek, J., Browning, H.L. and Buchsbaum, M. S. (1988). Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with positron emission tomography.

Intelligence, *12* (2),199-217.

Haier, R. J., Siegel, B., Tang, C., Abel, L., and Buchsbaum, M. S. (1992). Intelligence and changes in regional cerebral glucose metabolic rate following learning. *Intelligence 16* (3), 415.

Haier, R. J., White, N. S., and Alkire, M. T. (2003). Individual differences in general intelligence correlate with brain function during nonreasoning tasks. *Intelligence*, *31*(5), 429-441.

Haier R.J, Jung R., Yeo R., Head K., Alkire M.T. (2004). Structural brain variation and general intelligence. *NeuroImage*, *23*(1), 425-433.

Halpern, D. F. (1997). Sex differences in intelligence: Implications for education. *American Psychologist*, *52*(10), 1091-1102.

Halpern, D. F. (1997). Sex differences in intelligence: Implications for education. *American Psychologist*, *52*(10), 1091-1102.

Hauser, M. D., MacNeilage, P., & Ware, M. (1996). Numerical representations in primates. *Proceedings of the National Academy of Sciences*, *93*(4), 1514-1517.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*(5), 253-270.

Hunt, M. (2009). *The story of psychology*. New York: Random House, LLC.

Hunt, M. (2009). *The story of psychology*. New York: Random House, LLC.

Jung, R. E., & Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence. *Behavioral and Brain Sciences*, *30*,
135–154.

Jung, R. E., & Haier, R. J. (2013). Creativity and Intelligence: Brain Networks That Link and Differentiate the Expression of Genius.

Kohler, W. (1924). *The mentality of apes*. Oxford: Harcourt, Brace.

Lopes, P. N., Grewal, D., Kadis, J., Gall, M., & Salovey, P. (2006). Evidence that emotional intelligence is related to job performance and affect and attitudes at work. *Psicothema*, *18* (Suppl.), 132–138.

Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology*, *42*(2), 236-243.

Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. J. Sluyter (Eds.), *Emotional development and emotional intelligence: Educational implications* (pp. 3–34). New York: Basic.

National Spelling Bee. (2014a). *Statistics*. Retrieved from: http://www.spellingbee.com/statistics

National Spelling Bee. (2014b). *Get to Know the Competition*. Retrieved from: http://www.spellingbee.com/UserFiles...od2341418.html

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274

Spearman, C. (1904). " General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, *15*(2), 201-292.

Stevens, F. L., Hurley, R. A., & Taber, K. H. (2011). Anterior cingulate cortex: unique role in cognition and emotion. *The Journal of neuropsychiatry and clinical neurosciences*, *23*(2), 121-125.

Terman, L. M. (1916). *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford revision and extension of the Binet-Simon Intelligence Scale*. Boston: Houghton Mifflin.

Terman, L. M. (1916). *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford revision and extension of the Binet-Simon Intelligence Scale*. Boston: Houghton Mifflin.

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale*. Oxford: Psychological Corporation.

18.11.1.6 **Attributions**

Adapted by Kenneth A. Koenigshofer, PhD., from Intelligence by Robert Biswas-Diener, licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License; Thinking and Intelligence by OpenStax licensed CC BY-NC 4.0 via OER Commons; from What are Intelligence and Creativity? by OpenStax licensed CC BY-NC 4.0 via OER Commons; Wikipedia, Three-stratum theory, retrieved 9/29/21; Wikipedia, Fluid and Crystallized Intelligence, retrieved 9/29/21.

"Overview," "Brain Correlates of Intelligence and Creativity," and "Summary" is original material written by Kenneth A. Koenigshofer, PhD, Chaffey College, licensed under CC BY 4.0

18.11.2 Outside Resources

Blog: Dr. Jonathan Wai has an excellent blog on Psychology Today discussing many of the most interesting issues related to intelligence.

Video: Hank Green gives a fun and interesting overview of the concept of intelligence in this installment of the Crash Course series.

18.11.2.1 Vocabulary

G (or g)

Short for “general factor” and is often used to be synonymous with intelligence itself.

IQ

Short for “intelligence quotient.” This is a score, typically obtained from a widely used measure of intelligence that is meant to rank a person’s intellectual ability against that of others.

Norm

Assessments are given to a representative sample of a population to determine the range of scores for that population. These “norms” are then used to place an individual who takes that assessment on a range of scores in which he or she is compared to the population at large.

Standardize

Assessments that are given in the exact same manner to all people . With regards to intelligence tests standardized scores are individual scores that are computed to be referenced against normative scores for a population (see “norm”).

This page titled 18.11: Chapter 14- Traditional Models of Human Intelligence is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

18.12: Chapter 14- Problem Solving, Categories and Concepts

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Define problem types
2. Describe problem solving strategies
3. Define algorithm and heuristic
4. Describe the role of insight in problem solving
5. Explain some common roadblocks to effective problem solving
6. What is meant by a search problem
7. Describe means-ends analysis
8. How do analogies and restructuring contribute to problem solving
9. Explain how experts solve problems and what gives them an advantage over non-experts
10. Describe the brain mechanisms in problem solving

18.12.0.1 Overview

In this section we examine problem-solving strategies. People face problems every day—usually, multiple problems throughout the day. Sometimes these problems are straightforward: To double a recipe for pizza dough, for example, all that is required is that each ingredient in the recipe be doubled. Sometimes, however, the problems we encounter are more complex. For example, say you have a work deadline, and you must mail a printed copy of a report to your supervisor by the end of the business day. The report is time-sensitive and must be sent overnight. You finished the report last night, but your printer will not work today. What should you do? First, you need to identify the problem and then apply a strategy, usually a set of steps, for solving the problem.

18.12.0.1 Defining Problems

We begin this module on Problem Solving by giving a short description of what psychologists regard as a problem. Afterwards we are going to present different approaches towards problem solving, starting with gestalt psychologists and ending with modern search strategies connected to artificial intelligence. In addition we will also consider how experts do solve problems and finally we will have a closer look at two topics: The neurophysiological background on the one hand and the question what kind of role can be assigned to evolution regarding problem solving on the other.

The most basic definition is “A problem is any given situation that differs from a desired goal”. This definition is very useful for discussing problem solving in terms of evolutionary adaptation, as it allows to understand every aspect of (human or animal) life as a problem. This includes issues like finding food in harsh winters, remembering where you left your provisions, making decisions about which way to go, repeating and varying all kinds of complex movements by learning, and so on. Though all these problems were of crucial importance during the evolutionary process that created us the way we are, they are by no means solved exclusively by humans. We find a most amazing variety of different solutions for these problems of adaptation in animals as well (just consider, e.g., by which means a bat hunts its prey, compared to a spider).

However, for this module, we will mainly focus on abstract problems that humans may encounter (e.g. playing chess or doing an assignment in college). Furthermore, we will not consider those situations as abstract problems that have an obvious solution: Imagine a college student, let's call him Knut. Knut decides to take a sip of coffee from the mug next to his right hand. He does not even have to think about how to do this. This is not because the situation itself is trivial (a robot capable of recognizing the mug, deciding whether it is full, then grabbing it and moving it to Knut’s mouth would be a highly complex machine) but because in the

context of all possible situations it is so trivial that it no longer is a problem our consciousness needs to be bothered with. The problems we will discuss in the following all need some conscious effort, though some seem to be solved without us being able to say how exactly we got to the solution. Still we will find that often the strategies we use to solve these problems are applicable to more basic problems, as well as the more abstract ones such as completing a reading or writing assignment for a college class.

Non-trivial, abstract problems can be divided into two groups:

18.12.0.0.1 Well-defined Problems

For many abstract problems it is possible to find an algorithmic solution. We call all those problems well-defined that can be properly formalized, which comes along with the following properties:

- The problem has a clearly defined given state. This might be the line-up of a chess game, a given formula you have to solve, or the set-up of the towers of Hanoi game (which we will discuss later).
- There is a finite set of operators, that is, of rules you may apply to the given state. For the chess game, e.g., these would be the rules that tell you which piece you may move to which position.
- Finally, the problem has a clear goal state: The equations is resolved to x, all discs are moved to the right stack, or the other player is in checkmate.

Not surprisingly, a problem that fulfills these requirements can be implemented algorithmically (also see convergent thinking). Therefore many well-defined problems can be very effectively solved by computers, like playing chess.

18.12.0.0.2 Ill-defined Problems

Though many problems can be properly formalized (sometimes only if we accept an enormous complexity) there are still others where this is not the case. Good examples for this are all kinds of tasks that involve creativity, and, generally speaking, all problems for which it is not possible to clearly define a given state and a goal state: Formalizing a problem of the kind “Please paint a beautiful picture” may be impossible. Still this is a problem most people would be able to access in one way or the other, even if the result may be totally different from person to person. And while Knut might judge that picture X is gorgeous, you might completely disagree.

Nevertheless ill-defined problems often involve sub-problems that can be totally well-defined. On the other hand, many every-day problems that seem to be completely well-defined involve a great deal of creativity and many ambiguities. For example, suppose Knut has to read some technical material and then write an essay about it.

If we think of Knut's fairly ill-defined task of writing an essay, he will not be able to complete this task without first understanding the text he has to write about. This step is the first sub-goal Knut has to solve. Interestingly, ill-defined problems often involve subproblems that are well-defined.

Knut’s situation could be explained as a classical example of problem solving: He needs to get from his present state – an unfinished assignment – to a goal state - a completed assignment - and has certain operators to achieve that goal. Both Knut’s short and long term memory are active. He needs his short term memory to integrate what he is reading with the information from earlier passages of the paper. His long term memory helps him remember what he learned in the lectures he took and what he read in other books. And of course Knut’s ability to comprehend language enables him to make sense of the letters printed on the paper and to relate the sentences in a proper way.

Same place, different day. Knut is sitting at his desk again, staring at a blank paper in front of him, while nervously playing with a pen in his right hand. Just a few hours left to hand in his essay and he has not written a word. All of a sudden he smashes his fist on the table and cries out: "I need a plan!

18.12.0.0.3 How is a problem represented in the mind?

Generally speaking, problem representations are models of the situation as experienced by the agent. Representing a problem means to analyze it and split it into separate components:

- objects, predicates
- state space
- operators

- selection criteria

Therefore the efficiency of Problem Solving depends on the underlying representations in a person's mind. Analyzing the problem domain according to different dimensions, i.e., changing from one representation to another, results in arriving at a new understanding of a problem. This is basically what is described as restructuring.

18.12.0.0.4 Insight

There are two very different ways of approaching a *goal-oriented situation*. In one case an organism readily *reproduces* the response to the given problem from past experience. This is called **reproductive thinking**.

The second way requires *something new and different* to achieve the goal, prior learning is of little help here. Such **productive thinking** is (sometimes) argued to involve **insight**. Gestalt psychologists even state that insight problems are a separate category of problems in their own right.

Tasks that might involve insight usually have certain features – they require something new and non-obvious to be done and in most cases they are difficult enough to predict that the initial solution attempt will be unsuccessful. When you solve a problem of this kind you often have a so called **"AHA-experience"** – the solution pops up all of a sudden. At one time you do not have any ideas of the answer to the problem, you do not even feel to make any progress trying out different ideas, but in the next second the problem is solved.

18.12.0.0.5 Fixation

Sometimes, previous experience or familiarity can even make problem solving more difficult. This is the case whenever habitual directions get in the way of finding new directions – an effect called **fixation**.

Functional fixedness

Functional fixedness concerns the solution of *object-use problems*. The basic idea is that when the usual way of using an object is emphasised, it will be far more difficult for a person to use that object in a novel manner.

An example is the **two-string problem**: Knut is left in a room with a chair and a pair of pliers given the task to bind two strings together that are hanging from the ceiling. The problem he faces is that he can never reach both strings at a time because they are just too far away from each other. What can Knut do?



Figure **18.12.1**: Put the two strings together by tying the pliers to one of the strings and then swing it toward the other one.

Mental fixedness

Functional fixedness as involved in the examples above illustrates a **mental set** – a person's tendency to respond to a given task in a manner based on past experience. Because Knut maps an object to a particular function he has difficulties to *vary the way of use*

(pliers as pendulum's weight).

18.12.0.1 Problem-Solving Strategies

When you are presented with a problem—whether it is a complex mathematical problem or a broken printer, how do you solve it? Before finding a solution to the problem, the problem must first be clearly identified. After that, one of many problem solving strategies can be applied, hopefully resulting in a solution. Regardless of strategy, you will likely be guided, consciously or unconsciously, by your knowledge of cause-effect relations among the elements of the problem and the similarity of the problem to previous problems you have solved before. As discussed in earlier sections of this chapter, innate dispositions of the brain to look for and represent causal and similarity relations are key components of general intelligence (Koenigshofer, 2017).

A problem-solving strategy is a plan of action used to find a solution. Different strategies have different action plans associated with them. For example, a well-known strategy is trial and error. The old adage, “If at first you don’t succeed, try, try again” describes trial and error. In terms of your broken printer, you could try checking the ink levels, and if that doesn’t work, you could check to make sure the paper tray isn’t jammed. Or maybe the printer isn’t actually connected to your laptop. When using trial and error, you would continue to try different solutions until you solved your problem. Although trial and error is not typically one of the most time-efficient strategies, it is a commonly used one.

Table 1: Problem-Solving Strategies

Method	Description	Example
Trial and error	Continue trying different solutions until problem is solved	Restarting phone, turning off WiFi, turning off bluetooth in order to determine why your phone is malfunctioning
Algorithm	Step-by-step problem-solving formula	Instruction manual for installing new software on your computer
Heuristic	General problem-solving framework	Working backwards; breaking a task into steps

Another type of strategy is an algorithm. An algorithm is a problem-solving formula that provides you with step-by-step instructions used to achieve a desired outcome (Kahneman, 2011). You can think of an algorithm as a recipe with highly detailed instructions that produce the same result every time they are performed. Algorithms are used frequently in our everyday lives, especially in computer science. When you run a search on the Internet, search engines like Google use algorithms to decide which entries will appear first in your list of results. Facebook also uses algorithms to decide which posts to display on your newsfeed. Can you identify other situations in which algorithms are used?

A heuristic is another type of problem solving strategy. While an algorithm must be followed exactly to produce a correct result, a heuristic is a general problem-solving framework (Tversky & Kahneman, 1974). You can think of these as mental shortcuts that are used to solve problems. A “rule of thumb” is an example of a heuristic. Such a rule saves the person time and energy when making a decision, but despite its time-saving characteristics, it is not always the best method for making a rational decision. Different types of heuristics are used in different types of situations, but the impulse to use a heuristic occurs when one of five conditions is met (Pratkanis, 1989):

- When one is faced with too much information
- When the time to make a decision is limited
- When the decision to be made is unimportant
- When there is access to very little information to use in making the decision
- When an appropriate heuristic happens to come to mind in the same moment

Working backwards is a useful heuristic in which you begin solving the problem by focusing on the end result. Consider this example: You live in Washington, D.C. and have been invited to a wedding at 4 PM on Saturday in Philadelphia. Knowing that Interstate 95 tends to back up any day of the week, you need to plan your route and time your departure accordingly. If you want to be at the wedding service by 3:30 PM, and it takes 2.5 hours to get to Philadelphia without traffic, what time should you leave your

house? You use the working backwards heuristic to plan the events of your day on a regular basis, probably without even thinking about it.

Another useful heuristic is the practice of accomplishing a large goal or task by breaking it into a series of smaller steps. Students often use this common method to complete a large research project or long essay for school. For example, students typically brainstorm, develop a thesis or main topic, research the chosen topic, organize their information into an outline, write a rough draft, revise and edit the rough draft, develop a final draft, organize the references list, and proofread their work before turning in the project. The large task becomes less overwhelming when it is broken down into a series of small steps.

18.12.0.1 Problem Solving as a Search Problem

The idea of regarding problem solving as a search problem originated from Alan Newell and Herbert Simon while trying to design computer programs which could solve certain problems. This led them to develop a program called General Problem Solver which was able to solve any well-defined problem by creating heuristics on the basis of the user's input. This input consisted of objects and operations that could be done on them.

As we already know, every problem is composed of an initial state, intermediate states and a goal state (also: desired or final state), while the initial and goal states characterise the situations before and after solving the problem. The intermediate states describe any possible situation between initial and goal state. The set of operators builds up the transitions between the states. A solution is defined as the sequence of operators which leads from the initial state across intermediate states to the goal state.

The simplest method to solve a problem, defined in these terms, is to search for a solution by just trying one possibility after another (also called trial and error).

As already mentioned above, an organised search, following a specific strategy, might not be helpful for finding a solution to some ill-defined problem, since it is impossible to formalise such problems in a way that a search algorithm can find a solution.

As an example we could just take Knut and his essay: he has to find out about his own opinion and formulate it and he has to make sure he understands the sources texts. But there are no predefined operators he can use, there is no panacea how to get to an opinion and even not how to write it down.

18.12.0.0.1 **Means-End Analysis**

In Means-End Analysis you try to reduce the difference between initial state and goal state by creating sub-goals until a sub-goal can be reached directly (in computer science, what is called recursion works on this basis).

An example of a problem that can be solved by Means-End Analysis is the "Towers of Hanoi"

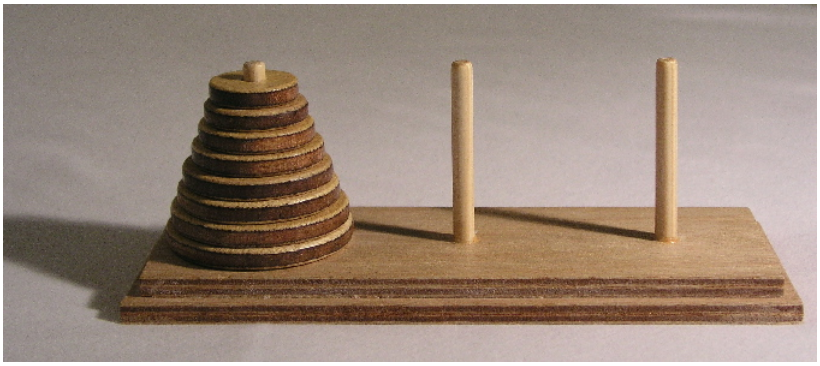


Figure **18.12.2**: Towers of Hanoi with 8 discs – A well defined problem (image from Wikimedia Commons; https://commons.wikimedia.org/wiki/F..._of_Hanoi.jpeg, by User:Evanherk.licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license).

The initial state of this problem is described by the different sized discs being stacked in order of size on the first of three pegs (the “start-peg“). The goal state is described by these discs being stacked on the third pegs (the “end-peg“) in exactly the same order.

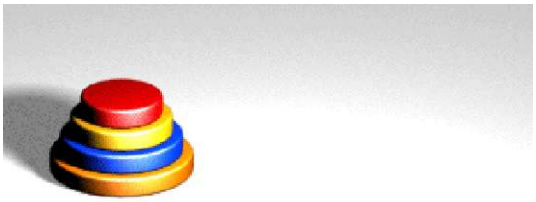
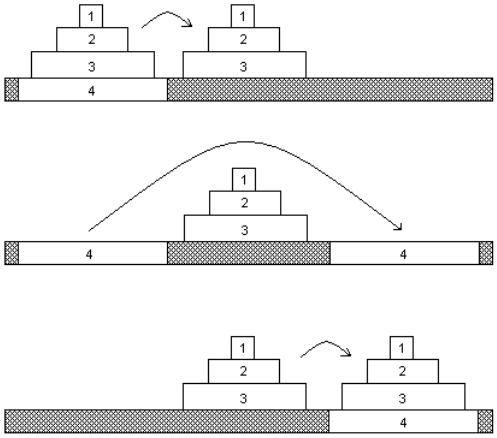


Figure 18.12.3: This animation shows the solution of the game "Tower of Hanoi" with four discs. (image from Wikimedia Commons; https://commons.wikimedia.org/wiki/F...of_Hanoi_4.gif; by André Karwath aka Aka; licensed under the Creative Commons Attribution-Share Alike 2.5 Generic license).

There are three operators:

- You are allowed to move one single disc from one peg to another one
- You are only able to move a disc if it is on top of one stack
- A disc cannot be put onto a smaller one.



In order to use Means-End Analysis we have to create sub-goals. One possible way of doing this is described in the picture:

1. Moving the discs lying on the biggest one onto the second peg.

2. Shifting the biggest disc to the third peg.

3. Moving the other ones onto the third peg, too

You can apply this strategy again and again in order to reduce the problem to the case where you only have to move a single disc – which is then something you are allowed to do.

Strategies of this kind can easily be formulated for a computer; the respective algorithm for the Towers of Hanoi would look like this:

1. move n-1 discs from A to B

2. move disc #n from A to C

3. move n-1 discs from B to C

where n is the total number of discs, A is the first peg, B the second, C the third one. Now the problem is reduced by one with each recursive loop.

Means-end analysis is important to solve everyday-problems – like getting the right train connection: You have to figure out where you catch the first train and where you want to arrive, first of all. Then you have to look for possible changes just in case you do not get a direct connection. Third, you have to figure out what are the best times of departure and arrival, on which platforms you leave and arrive and make it all fit together.

18.12.0.0.2 Analogies

Analogies describe similar structures and interconnect them to clarify and explain certain relations. In a recent study, for example, a song that got stuck in your head is compared to an itching of the brain that can only be scratched by repeating the song over and over again. Useful analogies appears to be based on a psychological mapping of relations between two very disparate types of problems that have abstract relations in common. Applied to STEM problems, Gray and Holyoak (2021) state: "Analogy is a powerful tool for fostering conceptual understanding and transfer in STEM and other fields. Well-constructed analogical comparisons focus attention on the causal-relational structure of STEM concepts, and provide a powerful capability to draw inferences based on a well-understood source domain that can be applied to a novel target domain." Note that similarity between problems of different types in their abstract relations, such as causation, is a key feature of reasoning, problem-solving and inference when forming and using analogies. Recall the discussion of general intelligence in module 14.2. There, similarity relations, causal relations, and predictive relations between events were identified as key components of general intelligence, along with ability to visualize in imagination possible future actions and their probable outcomes prior to commiting to actual behavior in the physical world (Koenigshofer, 2017).

18.12.0.0.3 Restructuring by Using Analogies

One special kind of restructuring, the way already mentioned during the discussion of the Gestalt approach, is analogical problem solving. Here, to find a solution to one problem – the so called target problem, an analogous solution to another problem – the source problem, is presented.

An example for this kind of strategy is the radiation problem posed by K. Duncker in 1945:

As a doctor you have to treat a patient with a malignant, inoperable tumour, buried deep inside the body. There exists a special kind of ray, which is perfectly harmless at a low intensity, but at the sufficient high intensity is able to destroy the tumour – as well as the healthy tissue on his way to it. What can be done to avoid the latter?

When this question was asked to participants in an experiment, most of them couldn't come up with the appropriate answer to the problem. Then they were told a story that went something like this:

A General wanted to capture his enemy's fortress. He gathered a large army to launch a full-scale direct attack, but then learned, that all the roads leading directly towards the fortress were blocked by mines. These roadblocks were designed in such a way, that it was possible for small groups of the fortress-owner's men to pass them safely, but every large group of men would initially set them off. Now the General figured out the following plan: He divided his troops into several smaller groups and made each of them march down a different road, timed in such a way, that the entire army would reunite exactly when reaching the fortress and could hit with full strength.

Here, the story about the General is the source problem, and the radiation problem is the target problem. The fortress is analogous to the tumour and the big army corresponds to the highly intensive ray. Consequently a small group of soldiers represents a ray at low intensity. The solution to the problem is to split the ray up, as the general did with his army, and send the now harmless rays towards the tumour from different angles in such a way that they all meet when reaching it. No healthy tissue is damaged but the tumour itself gets destroyed by the ray at its full intensity.

M. Gick and K. Holyoak presented Duncker's radiation problem to a group of participants in 1980 and 1983. Only 10 percent of them were able to solve the problem right away, 30 percent could solve it when they read the story of the general before. After given an additional hint – to use the story as help – 75 percent of them solved the problem.

With this results, Gick and Holyoak concluded, that analogical problem solving depends on three steps:

1. **Noticing** that an analogical connection exists between the source and the target problem.
2. **Mapping** corresponding parts of the two problems onto each other (fortress → tumour, army → ray, etc.)
3. **Applying** the mapping to generate a parallel solution to the target problem (using little groups of soldiers approaching from different directions → sending several weaker rays from different directions)

18.12.0.0.4 Schemas

The concept that links the target problem with the analogy (the “source problem“) is called problem schema. Gick and Holyoak obtained the activation of a schema on their participants by giving them two stories and asking them to compare and summarize

them. This activation of problem schemata is called “schema induction“.

The two presented texts were picked out of six stories which describe analogical problems and their solution. One of these stories was "The General."

After solving the task the participants were asked to solve the radiation problem. The experiment showed that in order to solve the target problem reading of two stories with analogical problems is more helpful than reading only one story: After reading two stories 52% of the participants were able to solve the radiation problem (only 30% were able to solve it after reading only one story, namely: “The General“).

The process of using a schema or analogy, i.e. applying it to a novel situation, is called *transduction*. One can use a common strategy to solve problems of a new kind.

To create a good schema and finally get to a solution using the schema is a problem-solving skill that requires practice and some background knowledge.

18.12.0.0.5 How Do Experts Solve Problems?

With the term expert we describe someone who devotes large amounts of his or her time and energy to one specific field of interest in which he, subsequently, reaches a certain level of mastery. It should not be of surprise that experts tend to be better in solving problems in their field than novices (people who are beginners or not as well trained in a field as experts) are. They are faster in coming up with solutions and have a higher success rate of right solutions. But what is the difference between the way experts and non-experts solve problems? Research on the nature of expertise has come up with the following conclusions:

When it comes to problems that are situated outside the experts' field, their performance often does not differ from that of novices.

Knowledge: An experiment by Chase and Simon (1973a, b) dealt with the question how well experts and novices are able to reproduce positions of chess pieces on chessboards when these are presented to them only briefly. The results showed that experts were far better in reproducing actual game positions, but that their performance was comparable with that of novices when the chess pieces were arranged randomly on the board. Chase and Simon concluded that the superior performance on actual game positions was due to the ability to recognize familiar patterns: A chess expert has up to 50,000 patterns stored in his memory. In comparison, a good player might know about 1,000 patterns by heart and a novice only few to none at all. This very detailed knowledge is of crucial help when an expert is confronted with a new problem in his field. Still, it is not pure size of knowledge that makes an expert more successful. Experts also organise their knowledge quite differently from novices.

Organization: In 1982 M. Chi and her co-workers took a set of 24 physics problems and presented them to a group of physics professors as well as to a group of students with only one semester of physics. The task was to group the problems based on their similarities. As it turned out the students tended to group the problems based on their surface structure (similarities of objects used in the problem, e.g. on sketches illustrating the problem), whereas the professors used their deep structure (the general physical principles that underlay the problems) as criteria. By recognizing the actual structure of a problem experts are able to connect the given task to the relevant knowledge they already have (e.g. another problem they solved earlier which required the same strategy).

Analysis: Experts often spend more time analyzing a problem before actually trying to solve it. This way of approaching a problem may often result in what appears to be a slow start, but in the long run this strategy is much more effective. A novice, on the other hand, might start working on the problem right away, but often has to realise that he reaches dead ends as he chose a wrong path in the very beginning.

18.12.0.1 Creative Cognition

18.12.0.0.1 **Divergent Thinking**

The term divergent thinking describes a way of thinking that does not lead to one goal, but is open-ended. Problems that are solved this way can have a large number of potential 'solutions' of which none is exactly 'right' or 'wrong', though some might be more suitable than others.

Solving a problem like this involves indirect and productive thinking and is mostly very helpful when somebody faces an ill-defined problem, i.e. when either initial state or goal state cannot be stated clearly and operators are either insufficient or not given at all.

The process of divergent thinking is often associated with creativity, and it undoubtedly leads to many creative ideas. Nevertheless, researches have shown that there is only modest correlation between performance on divergent thinking tasks and other measures of creativity. Additionally it was found that in processes resulting in original and practical inventions things like searching for solutions, being aware of structures and looking for analogies are heavily involved, too.

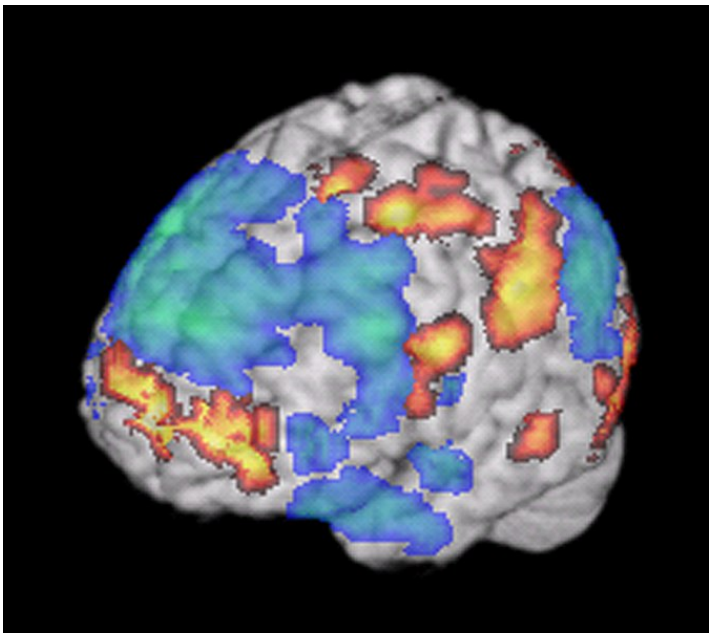


Figure **18.12.4**: functional MRI images of the brains of musicians playing improvised jazz revealed that a large brain region involved in monitoring one's performance shuts down during creative improvisation, while a small region involved in organizing self-initiated thoughts and behaviors is highly activated (Image and modified caption from Wikimedia Commons. File:Creative Improvisation (24130148711).jpg; https://commons.wikimedia.org/wiki/F...130148711).jpg; by NIH Image Gallery; *As a work of the U.S. federal government, the image is in the* ***public domain***.

18.12.0.0.2 **Convergent Thinking**

Convergent thinking patterns are problem solving techniques that unite different ideas or fields to find a solution. The focus of this mindset is speed, logic and accuracy, also identification of facts, reapplying existing techniques, gathering information. The most important factor of this mindset is: there is only one correct answer. You only think of two answers, namely right or wrong. This type of thinking is associated with certain science or standard procedures. People with this type of thinking have logical thinking, are able to memorize patterns, solve problems and work on scientific tests. Most school subjects sharpen this type of thinking ability.

Research shows that the creative process involves both types of thought processes.

18.12.0.1 Brain Mechanisms in Problem Solving

Presenting Neurophysiology in its entirety would be enough to fill several books. Instead, let's focus only on the aspects that are especially relevant to problem solving. Still, this topic is quite complex and problem solving cannot be attributed to one single brain area. Rather there are systems or networks of several brain areas working together to perform a specific problem solving task. This is best shown by an example, playing chess:

Task	Location of Brain activity
• Identifying chess pieces • determining location of pieces • Thinking about making a move • Remembering a pieces move • Planning and executing strategies	• Pathway from Occipital to Temporal Lobe (also called the "what"-pathway of visual processing) • Pathway from Occipital to parietal Lobe (also called the "where"-pathway of visual processing) • Premotor area • Hippocampus (forming new memories) • Prefrontal cortex

Table 2: Brain areas involved in a complex cognitive task.

One of the key tasks, namely **planning and executing strategies**, is performed by the prefrontal cortex (PFC), which also plays an important role for several other tasks correlated with problem solving. This can be made clear from the effects of damage to the PFC on ability to solve problems.

Patients with a lesion in this brain area have difficulty switching from one behavioral pattern to another. A well known example is the wisconsin card-sorting task. A patient with a PFC lesion who is told to separate all blue cards from a deck, would continue sorting out the blue ones, even if the experimenter next told him to sort out all brown cards. Transferred to a more complex problem, this person would most likely fail, because he is not flexible enough to change his strategy after running into a dead end or when the problem changes.

Another example comes from a young homemaker, who had a tumour in the frontal lobe. Even though she was able to cook individual dishes, preparing a whole family meal was an impossible task for her.

Mushiake et al. (2009) note that to achieve a goal in a complex environment, such as problem-solving situations like those above, we must plan multiple steps of action. When planning a series of actions, we have to anticipate future outcomes that will occur as a result of each action, and, in addition, we must mentally organize the temporal sequence of events in order to achieve the goal. These researchers investigated the role of **lateral prefrontal cortex** (PFC) in problem solving in monkeys. They found that "PFC neurons reflected final goals and immediate goals during the preparatory period. [They] also found some PFC neurons reflected each of *all* the forthcoming steps of actions during the preparatory period and they increased their [neural] activity step by step during the execution period. [Furthermore, they] found that the transient increase in synchronous activity of PFC neurons was involved in goal subgoal transformations. [They concluded] that the PFC is involved primarily in the dynamic representation of multiple future events that occur as a consequence of behavioral actions in problem-solving situations" (Mushiake et al., 2009, p. 1). In other words, the **prefrontal cortex** represents in our imagination the sequence of events following each step that we take in solving a particular problem, guiding us step by step to the solution.

As the examples above illustrate, the structure of our brain is of great importance regarding problem solving, i.e. cognitive life. But how was our cognitive apparatus designed? How did perception-action integration as a central species-specific property of humans come about? The answer, as argued extensively in earlier sections of this book, is, of course, natural selection and other forces of genetic evolution. Clearly, animals and humans with genes facilitating brain organization that led to good problem solving skills would be favored by natural selection over genes responsible for brain organization less adept at solving problems. We became equipped with brains organized for effective problem solving because flexible abilities to solve a wide range of problems presented by the environment enhanced ability to survive, to compete for resources, to escape predators, and to reproduce (see chapter on Evolution and Genetics in this text).

In short, good problem solving mechanisms in brains designed for the real world gave a competitive advantage and increased biological fitness. Consequently, humans (and many other animals to a lesser degree) have "innate ability to problem-solve in the real world. Solving real world problems in real time given constraints posed by one's environment is crucial for survival . . . Real world problem solving (RWPS) is different from those that occur in a classroom or in a laboratory during an experiment. They are often dynamic and discontinuous, accompanied by many starts and stops . . . Real world problems are typically ill-defined, and even when they are well-defined, often have open-ended solutions . . . RWPS is quite messy and involves a tight interplay between problem solving, creativity, and insight . . . In psychology and neuroscience, *problem-solving* broadly refers to the inferential steps

taken by an agent [human, animal, or computer] that leads from a given state of affairs to a desired goal state" (Sarathy, 2018, p. 261-2). According to Sarathy (2018), the initial stage of RWPS requires defining the problem and generating a representation of it in working memory. This stage involves activation of parts of the "**prefrontal cortex (PFC)**, **default mode network (DMN)**, and the **dorsal anterior cingulate cortex (dACC)**." The **DMN includes the medial prefrontal cortex, posterior cingulate cortex, and the inferior parietal lobule.** Other **structures sometimes considered part of the network are the lateral temporal cortex, hippocampal formation, and the precuneus.** This network of structures is called "default mode" because these structures show increased activity when one is ***not*** engaged in focused, attentive, goal-directed actions, but rather a "resting state" (a baseline default state) and show decreased neural activity when one is focused and attentive to a particular goal-directed behavior (Raichle, et al., 2001).

18.12.0.1 Moral Reasoning

Jeurissen, et al., (2014) examined a special type of reasoning, moral reasoning, using TMS (Transcranial Magnetic Stimulation). The **dorsolateral prefrontal cortex (DLPFC)** and **temporal-parietal junction (TPJ)** have both been shown to be involved in moral judgments, but this study by Jeurissen, et al., (2014) uses TMS to tease out the different roles these brain areas play in different scenarios involving moral dilemmas.

Moral dilemmas have been categorized by researchers as **moral-impersonal** (e.g., trolley or switch dilemma--**save** the lives of **five** workmen at the expense of the life of **one by switching train** to another track) and **moral-personal** dilemmas (e.g., footbridge dilemma--**push a strange**r in front of a train **to save** the lives of **five** others). In the first scenario, the person just pulls a switch resulting in death of one person to save five, but in the second, the person pushes the victim to their death to save five others.

Dual-process theory proposes that moral decision-making involves two components: an **automatic emotional response** and a **voluntary** application of a **utilitarian decision-rule** (in this case, one death to save five people is worth it). The thought of being responsible for the death of another person elicits an aversive emotional response, but at the same time, cognitive reasoning favors the utilitarian option. **Decision making and social cognition** are often **associated with** the **DLPFC. Neurons** in the **prefrontal cortex** have been found to be involved in cost-benefit analysis and categorize stimuli based on the predicted consequences.

Theory-of-mind (TOM) is a cognitive mechanism which is used when one tries to understand and explain the knowledge, beliefs, and intention of others. **TOM** and **empathy** are often associated with **TPJ functioning**.

In the article by Jeurissen, et al., (2014), brain activity is measured by BOLD. BOLD refers to **Blood-oxygen-level-dependent imaging**, or **BOLD-contrast imaging,** which is a way to measure neural activity in different brain areas in MRI images**.**

Greene et al., 2001 (Links to an external site.), 2004 (Links to an external site.) reported that activity in the **prefrontal cortex** is thought to be important for the **cognitive reasoning process**, which can counteract the automatic emotional response that occurs in moral dilemmas like the one in Jeurissen, et al., (2014). Greene et al. (2001) (Links to an external site.) found that the **medial portions** of the **medial frontal gyrus,** the **posterior cingulate gyrus,** and the **bilateral angular gyrus** showed a **higher BOLD** response in the **moral-personal condition** than the moral-impersonal condition. The **right middle frontal gyrus** and the **bilateral parietal lobes** showed a **lower BOLD** response in the **moral-personal condition** than in the moral impersonal. Furthermore, Greene et al. (2004) (Links to an external site.) showed an **increased BOLD** response for the bilateral **amygdale** in **personal** compared to the impersonal dilemmas.

Given the role of the prefrontal cortex in moral decision-making, Jeurissen, et al., (2014) **hypothesized** that when **magnetically stimulating prefrontal cortex**, they will **selectively influence** the decision process of the **moral personal dilemmas** because the **cognitive reasoning** for which the **DLPFC** is important is **disrupted**, thereby **releasing** the **emotional component** making it more influential in the resolution of the dilemma. Because the **activity** in the **TPJ** is related to **emotional processing and theory of mind** (Saxe and Kanwisher, 2003 (Links to an external site.); Young et al., 2010 (Links to an external site.)), Jeurissen, et al., (2014) **hypothesized** that when **magnetically stimulating** this area, the **TPJ,** during a moral decision, this will **selectively influence** the decision process of **moral-impersonal** dilemmas.

Results of this study by Jeurissen, et al., (2014) showed an **important role** of the **TPJ** in **moral judgment**. Experiments using fMRI (Greene et al., 2004 (Links to an external site.)), have found the **cingulate cortex** to be involved in **moral judgment**. In earlier studies, the cingulate cortex was found to be involved in the **emotional** response. Since the **moral-personal** dilemmas are **more emotional**ly salient, the **higher activity** observed for **TPJ** in the **moral-personal** condition **(more emotional)** is **consistent**

with this view. **Another area** that is hypothesized to be associated with the **emotional** response is the **temporal cortex**. In this study by Jeurissen, et al., (2014)**, magnetic stimulation** of the **right DLPFC** and **right TPJ** shows **roles for right DLPFC (reasoning and utilitarian) and right TPJ (emotion) in moral impersonal and moral personal dilemmas respectively. TMS over the right DLPFC (disrupting neural activity here) leads to** behavior changes consistent with **less cognitive control over emotion**. After right DLPFC stimulation, participants show less feelings of regret than after magnetic stimulation of the right TPJ. This last finding indicates that the **right DLPFC** is involved in **evaluating** the **outcome** of the decision process. In summary, this experiment by Jeurissen, et al., (2014) adds to evidence of a critical role of **right DLPFC and right TPJ** in **moral decision-making** and supports that hypothesis that the **former** is involved in judgments based on **cognitive reasoning** and **anticipation of outcomes,** whereas the **latter** is involved in **emotional processing** related to **moral dilemmas.**

18.12.0.1 Summary

Many different strategies exist for solving problems. Typical strategies include trial and error, applying algorithms, and using heuristics. To solve a large, complicated problem, it often helps to break the problem into smaller steps that can be accomplished individually, leading to an overall solution. The brain mechanisms involved in problem solving vary to some degree depending upon the sensory modalities involved in the problem and its solution, however, the prefrontal cortex is one brain region that appears to be centrally involved in all problem-solving. The prefrontal cortex is required for flexible shifts in attention, for representing the problem in working memory, and for holding steps in problem solving in working memory along with representations of future consequences of those actions permitting planning and execution of plans. Also implicated is the Default Mode Network (DMN) including medial prefrontal cortex, posterior cingulate cortex, and the inferior parietal module, and sometimes the lateral temporal cortex, hippocampus, and the precuneus. Moral reasoning involves a different set of brain areas, primarily the dorsolateral prefrontal cortex (DLPFC) and temporal-parietal junction (TPJ).

Review Questions

1. A specific formula for solving a problem is called _____.
   A. an algorithm
   B. a heuristic
   C. a mental set
   D. trial and error
2. A mental shortcut in the form of a general problem-solving framework is called _____.
   A. an algorithm
   B. a heuristic
   C. a mental set
   D. trial and error

18.12.0.1 References

Gray, M. E., & Holyoak, K. J. (2021). Teaching by analogy: From theory to practice. *Mind, Brain, and Education*, *15* (3), 250-263.

Hunt, L. T., Behrens, T. E., Hosokawa, T., Wallis, J. D., & Kennerley, S. W. (2015). Capturing the temporal evolution of choice across prefrontal cortex. *Elife*, *4*, e11945.

Mushiake, H., Sakamoto, K., Saito, N., Inui, T., Aihara, K., & Tanji, J. (2009). Involvement of the prefrontal cortex in problem solving. *International review of neurobiology*, *85*, 1-11.

Jeurissen, D., Sack, A. T., Roebroeck, A., Russ, B. E., & Pascual-Leone, A. (2014). TMS affects moral judgment, showing the role of DLPFC and TPJ in cognitive and emotional processing. *Frontiers in neuroscience*, *8*, 18.

Kahneman, D. (2011). Thinking, fast and slow. New York: Farrar, Straus, and Giroux.

Koenigshofer, K. A. (2017). General Intelligence: Adaptation to Evolutionarily Familiar Abstract Relational Invariants, Not to Environmental or Evolutionary Novelty. *The Journal of Mind and Behavior*, 119-153.

Pratkanis, A. (1989). The cognitive representation of attitudes. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude structure and function* (pp. 71–98). Hillsdale, NJ: Erlbaum.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, *98*(2), 676-682.

Sawyer, K. (2011). The cognitive neuroscience of creativity: a critical review. *Creat. Res. J.* 23, 137–154. doi: 10.1080/10400419.2011.571191

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

18.12.0.0.1 Brain Mechanisms in Problem Solving

Hunt, L. T., Behrens, T. E., Hosokawa, T., Wallis, J. D., & Kennerley, S. W. (2015). Capturing the temporal evolution of choice across prefrontal cortex. *Elife*, *4*, e11945.

Mushiake, H., Sakamoto, K., Saito, N., Inui, T., Aihara, K., & Tanji, J. (2009). Involvement of the prefrontal cortex in problem solving. *International review of neurobiology*, *85*, 1-11. https://www.sciencedirect.com/scienc...74774209850010

Sawyer, K. (2011). The cognitive neuroscience of creativity: a critical review. *Creat. Res. J.* 23, 137–154. doi: 10.1080/10400419.2011.571191

18.12.0.1 Attributions

"Overview," "Problem Solving Strategies," adapted from Problem Solving by OpenStax Colleg licensed CC BY-NC 4.0 via OER Commons

"Defining Problems," "Problem Solving as a Search Problem," "Creative Cognition," "Brain Mechanisms in Problem-Solving" adapted by Kenneth A. Koenigshofer, Ph.D., from 2.1, 2.2, 2.3, 2.4, 2.5, 2.6 in *Cognitive Psychology and Cognitive Neuroscience* (Wikibooks) https://en.wikibooks.org/wiki/Cognit...e_Neuroscience; unless otherwise noted, LibreTexts content is licensed by CC BY-NC-SA 3.0. Legal; the LibreTexts libraries are Powered by MindTouch

Moral Reasoning was written by Kenneth A. Koenigshofer, Ph.D, Chaffey College.

18.12.1 Categories and Concepts

By Gregory Murphy, New York University

People form mental concepts of categories of objects, which permit them to respond appropriately to new objects they encounter. Most concepts cannot be strictly defined but are organized around the “best” examples or prototypes, which have the properties most common in the category. Objects fall into many different categories, but there is usually a most salient one, called the basic-level category, which is at an intermediate level of specificity (e.g., chairs, rather than furniture or desk chairs). Concepts are closely related to our knowledge of the world, and people can more easily learn concepts that are consistent with their knowledge. Theories of concepts argue either that people learn a summary description of a whole category or else that they learn exemplars of the category. Recent research suggests that there are different ways to learn and represent concepts and that they are accomplished by different neural systems.

18.12.1.1 Learning Objectives

1. Understand the problems with attempting to define categories.
2. Understand typicality and fuzzy category boundaries.
3. Learn about theories of the mental representation of concepts.
4. Learn how knowledge may influence concept learning.

18.12.2 Introduction



Figure **18.12.2**: Although you've (probably) never seen this particular truck before, you know a lot about it because of the knowledge you've accumulated in the past about the features in the category of trucks. [Image: CC0 Public Domain, https://goo.gl/m25gce]

Consider the following set of objects: some dust, papers, a computer monitor, two pens, a cup, and an orange. What do these things have in common? Only that they all happen to be on my desk as I write this. This set of things can be considered a **category**, a set of objects that can be treated as equivalent in some way. But, most of our categories seem much more informative—they share many properties. For example, consider the following categories: trucks, wireless devices, weddings, psychopaths, and trout. Although the objects in a given category are different from one another, they have many commonalities. When you know something is a truck, you know quite a bit about it. The psychology of categories concerns how people learn, remember, and use informative categories such as trucks or psychopaths.

The mental representations we form of categories are called **concepts**. There is a category of trucks in the actual physical world, and I also have a concept of trucks in my head. We assume that people's concepts correspond more or less closely to the actual category, but it can be useful to distinguish the two, as when someone's concept is not really correct.

Concepts are at the core of intelligent behavior. We expect people to be able to know what to do in new situations and when confronting new objects. If you go into a new classroom and see chairs, a blackboard, a projector, and a screen, you know what these things are and how they will be used. You'll sit on one of the chairs and expect the instructor to write on the blackboard or project something onto the screen. You do this *even if you have never seen any of these particular objects before*, because you have concepts of classrooms, chairs, projectors, and so forth, that tell you what they are and what you're supposed to do with them. Furthermore, if someone tells you a new fact about the projector—for example, that it has a halogen bulb—you are likely to extend this fact to other projectors you encounter. In short, *concepts allow* you to extend what you have learned about a limited number of objects to a potentially infinite set of entities (i.e. *generalization*). Notice how *categories and concepts arise from similarity, one of the abstract features of the world that has been genetically internalized into the brain during evolution*, creating *an innate disposition of brains to search for and to represent groupings of similar things, forming one component of general intelligence.* One property of the human brain that distinguishes us from other animals is the high degrees of abstraction in similarity relations that the human brain is capable of encoding compared to the brains of non-human animals (Koenigshofer, 2017).

Simpler organisms, such as *animals and human infants, also have concepts* (Mareschal, Quinn, & Lea, 2010). Squirrels may have a concept of predators, for example, that is specific to their own lives and experiences. However, animals likely have many fewer concepts and cannot understand complex concepts such as mortgages or musical instruments.

You know thousands of categories, most of which you have learned without careful study or instruction. Although this accomplishment may seem simple, we know that it isn't, because it is difficult to program computers to solve such intellectual tasks. If you teach a learning program that a robin, a swallow, and a duck are all birds, it may not recognize a cardinal or peacock as a bird. However, this shortcoming in computers may be at least partially overcome when the type of processing used is parallel distributed processing as employed in artificial neural networks (Koenigshofer, 2017), discussed in this chapter. As we'll shortly see, the problem for computers is that objects in categories are often surprisingly diverse.

18.12.3 Nature of Categories

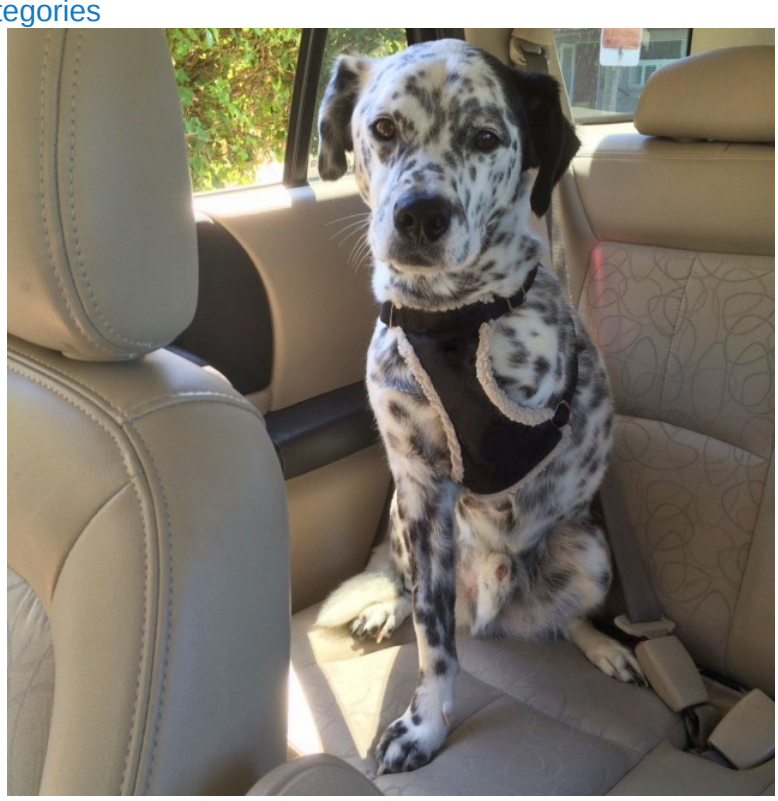


Figure **18.12.3**: Here is a very good dog, but one that does not fit perfectly into a well-defined category where all dogs have four legs. [Image: State Farm, https://goo.gl/KHtu6N, CC BY 2.0, https://goo.gl/BRvSA7]

Traditionally, it has been assumed that categories are *well-defined*. This means that you can give a definition that specifies what is in and out of the category. Such a definition has two parts. First, it provides the *necessary features* for category membership: What must objects have in order to be in it? Second, those features must be *jointly sufficient* for membership: If an object has those features, then it is in the category. For example, if I defined a dog as a four-legged animal that barks, this would mean that every dog is four-legged, an animal, and barks, and also that anything that has all those properties is a dog.

Unfortunately, it has not been possible to find definitions for many familiar categories. Definitions are neat and clear-cut; the world is messy and often unclear. For example, consider our definition of dogs. In reality, not all dogs have four legs; not all dogs bark. I knew a dog that lost her bark with age (this was an improvement); no one doubted that she was still a dog. It is often possible to find some necessary features (e.g., all dogs have blood and breathe), but these features are generally not sufficient to determine category membership (you also have blood and breathe but are not a dog).

Even in domains where one might expect to find clear-cut definitions, such as science and law, there are often problems. For example, many people were upset when Pluto was downgraded from its status as a planet to a dwarf planet in 2006. Upset turned to outrage when they discovered that there was no hard-and-fast definition of planethood: “Aren’t these astronomers scientists? Can’t they make a simple definition?” In fact, they couldn’t. After an astronomical organization tried to make a definition for planets, a number of astronomers complained that it might not include accepted planets such as Neptune and refused to use it. If everything looked like our Earth, our moon, and our sun, it would be easy to give definitions of planets, moons, and stars, but the universe has not conformed to this ideal.

18.12.3.1 Fuzzy Categories

18.12.3.1.1 **Borderline Items**

Experiments also showed that the psychological assumptions of well-defined categories were not correct. Hampton (1979) asked subjects to judge whether a number of items were in different categories. He did not find that items were either clear members or clear nonmembers. Instead, he found many items that were just barely considered category members and others that were just barely not members, with much disagreement among subjects. Sinks were barely considered as members of the kitchen utensil

category, and sponges were barely excluded. People just included seaweed as a vegetable and just barely excluded tomatoes and gourds. Hampton found that members and nonmembers formed a continuum, with no obvious break in people's membership judgments. If categories were well defined, such examples should be very rare. Many studies since then have found such *borderline members* that are not clearly in or clearly out of the category.

Furniture	Fruit
chair	orange
table	banana
desk	pear
bookcase	plum
lamp	strawberry
cushion	pineapple
rug	lemon
stove	honeydew
picture	date
vase	tomato

Table 1. Examples of two categories, with members ordered by typicality (from Rosch & Mervis, 1975)

McCloskey and Glucksberg (1978) found further evidence for borderline membership by asking people to judge category membership twice, separated by two weeks. They found that when people made repeated category judgments such as "Is an olive a fruit?" or "Is a sponge a kitchen utensil?" they changed their minds about borderline items—up to 22 percent of the time. So, not only do people disagree with one another about borderline items, they disagree with themselves! As a result, researchers often say that categories are *fuzzy*, that is, they have unclear boundaries that can shift over time.

18.12.3.1.2 Typicality

A related finding that turns out to be most important is that even among items that clearly are in a category, some seem to be "better" members than others (Rosch, 1973). Among birds, for example, robins and sparrows are very **typical**. In contrast, ostriches and penguins are very *atypical* (meaning not typical). If someone says, "There's a bird in my yard," the image you have will be of a smallish passerine bird such as a robin, not an eagle or hummingbird or turkey.

You can find out which category members are typical merely by asking people. Table 1 shows a list of category members in order of their rated typicality. Typicality is perhaps the most important variable in predicting how people interact with categories. The following text box is a partial list of what typicality influences.

We can understand the two phenomena of borderline members and typicality as two sides of the same coin. Think of the most typical category member: This is often called the category *prototype*. Items that are less and less similar to the prototype become less and less typical. At some point, these less typical items become so atypical that you start to doubt whether they are in the

category at all. Is a rug really an example of furniture? It's in the home like chairs and tables, but it's also different from most furniture in its structure and use. From day to day, you might change your mind as to whether this atypical example is in or out of the category. So, changes in typicality ultimately lead to borderline members.

Influences of Typicality on Cognition
• **Typical items are judged category members more often** (Hampton, 1979).
• **Speed of categorization is faster for typical items** (Rips, Shoben, & Smith, 1973).
• **Typical members are learned before atypical ones** (Rosch & Mervis, 1975).
• **Learning a category is easier if typical examples are provided** (Mervis & Pani, 1980).
• **In language comprehension, references to typical members are understood more easily** (Garrod & Sanford, 1977).
• **In language production, people tend to say typical items before atypical ones (e.g., "apples and lemons" rather than "lemons and apples")** (Onishi, Murphy, & Bock, 2008).

Table 2:
Text Box 1

18.12.4 Source of Typicality

Intuitively, it is not surprising that robins are better examples of birds than penguins are, or that a table is a more typical kind of furniture than is a rug. But given that robins and penguins are known to be birds, why should one be more typical than the other? One possible answer is the frequency with which we encounter the object: We see a lot more robins than penguins, so they must be more typical. Frequency does have some effect, but it is actually not the most important variable (Rosch, Simpson, & Miller, 1976). For example, I see both rugs and tables every single day, but one of them is much more typical as furniture than the other.

The best account of what makes something typical comes from Rosch and Mervis's (1975) *family resemblance theory*. They proposed that items are likely to be typical if they (a) have the features that are frequent in the category and (b) do not have features frequent in other categories. Let's compare two extremes, robins and penguins. Robins are small flying birds that sing, live in nests in trees, migrate in winter, hop around on your lawn, and so on. Most of these properties are found in many other birds. In contrast, penguins do not fly, do not sing, do not live in nests or in trees, do not hop around on your lawn. Furthermore, they have properties that are common in other categories, such as swimming expertly and having wings that look and act like fins. These properties are more often found in fish than in birds.



Figure 18.12.4: When you think of “bird,” how closely does the robin resemble your general figure? [Image: CC0 Public Domain, https://goo.gl/m25gce]

According to Rosch and Mervis, then, it is not because a robin is a very common bird that makes it typical. Rather, it is because the robin has the shape, size, body parts, and behaviors that are very common (i.e. most similar) among birds—and not common among fish, mammals, bugs, and so forth.

In a classic experiment, Rosch and Mervis (1975) made up two new categories, with arbitrary features. Subjects viewed example after example and had to learn which example was in which category. Rosch and Mervis constructed some items that had features that were common in the category and other items that had features less common in the category. The subjects learned the first type of item before they learned the second type. Furthermore, they then rated the items with common features as more typical. In another experiment, Rosch and Mervis constructed items that differed in how many features were shared with a *different* category. The more features were shared, the longer it took subjects to learn which category the item was in. These experiments, and many later studies, support both parts of the family resemblance theory.

18.12.5 Category Hierarchies

Many important categories fall into *hierarchies*, in which more concrete categories are nested inside larger, abstract categories. For example, consider the categories: brown bear, bear, mammal, vertebrate, animal, entity. Clearly, all brown bears are bears; all bears are mammals; all mammals are vertebrates; and so on. Any given object typically does not fall into just one category—it could be in a dozen different categories, some of which are structured in this hierarchical manner. Examples of biological categories come to mind most easily, but within the realm of human artifacts, hierarchical structures can readily be found: desk chair, chair, furniture, artifact, object.

Brown (1958), a child language researcher, was perhaps the first to note that there seems to be a preference for which category we use to label things. If your office desk chair is in the way, you’ll probably say, “Move that chair,” rather than “Move that desk chair” or “piece of furniture.” Brown thought that the use of a single, consistent name probably helped children to learn the name for things. And, indeed, children’s first labels for categories tend to be exactly those names that adults prefer to use (Anglin, 1977).

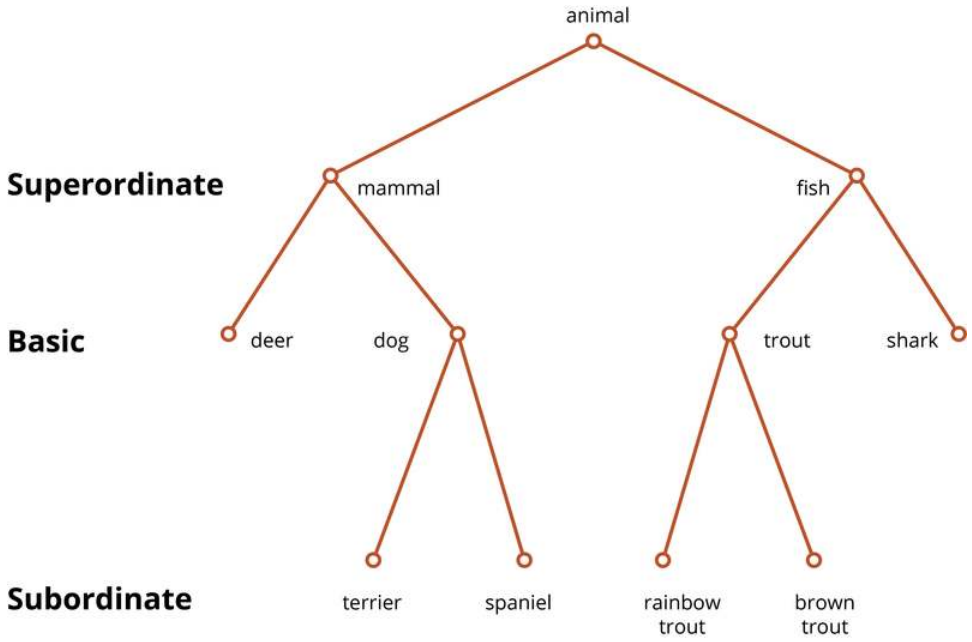


Figure **18.12.5**: This is a highly simplified illustration of hierarchically organized categories, with the superordinate, basic, and subordinate levels labeled. Keep in mind that there may be even more specific subordinates (e.g., wire-haired terriers) and more general superordinates (e.g., living thing)

This preference is referred to as a preference for the **basic level of categorization**, and it was first studied in detail by Eleanor Rosch and her students (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). The basic level represents a kind of Goldilocks effect, in which the category used for something is not too small (northern brown bear) and not too big (animal), but is just right (bear). The simplest way to identify an object's basic-level category is to discover how it would be labeled in a neutral situation. Rosch et al. (1976) showed subjects pictures and asked them to provide the first name that came to mind. They found that 1,595 names were at the basic level, with 14 more specific names (*subordinates*) used. Only once did anyone use a more general name (*superordinate*). Furthermore, in printed text, basic-level labels are much more frequent than most subordinate or superordinate labels (e.g., Wisniewski & Murphy, 1989).

The preference for the basic level is not merely a matter of labeling. Basic-level categories are usually easier to learn. As Brown noted, children use these categories first in language learning, and superordinates are especially difficult for children to fully acquire.[1] People are faster at identifying objects as members of basic-level categories (Rosch et al., 1976).

Rosch et al. (1976) initially proposed that basic-level categories cut the world at its joints, that is, merely reflect the big differences between categories like chairs and tables or between cats and mice that exist in the world. However, it turns out that which level is basic is not universal. North Americans are likely to use names like *tree*, *fish*, and *bird* to label natural objects. But people in less industrialized societies seldom use these labels and instead use more specific words, equivalent to *elm*, *trout*, and *finch* (Berlin, 1992). Because Americans and many other people living in industrialized societies know so much less than our ancestors did about the natural world, our basic level has "moved up" to what would have been the superordinate level a century ago. Furthermore, experts in a domain often have a preferred level that is more specific than that of non-experts. Birdwatchers see sparrows rather than just birds, and carpenters see roofing hammers rather than just hammers (Tanaka & Taylor, 1991). This all suggests that the preferred level is not (only) based on how different categories are in the world, but that people's knowledge and interest in the categories has an important effect.

One explanation of the basic-level preference is that basic-level categories are more *differentiated:* The category members are similar to one another, but they are different from members of other categories (Murphy & Brownell, 1985; Rosch et al., 1976). (The alert reader will note a similarity to the explanation of typicality I gave above. However, here we're talking about the entire category and not individual members.) Chairs are pretty similar to one another, sharing a lot of features (legs, a seat, a back, similar size and shape); they also don't share that many features with other furniture. Superordinate categories are not as useful because their members are not very similar to one another. What features are common to most furniture? There are very few. Subordinate categories are not as useful, because they're very similar to other categories: Desk chairs are quite similar to dining room chairs and

easy chairs. As a result, it can be difficult to decide which subordinate category an object is in (Murphy & Brownell, 1985). Experts can differ from novices in which categories are the most differentiated, because they know different things about the categories, therefore changing how similar the categories are.

> *[1] This is a controversial claim, as some say that infants learn superordinates before anything else (Mandler, 2004). However, if true, then it is very puzzling that older children have great difficulty learning the correct meaning of words for superordinates, as well as in learning artificial superordinate categories (Horton & Markman, 1980; Mervis, 1987). However, it seems fair to say that the answer to this question is not yet fully known.*

18.12.6 Theories of Concept Representation

Now that we know these facts about the psychology of concepts, the question arises of how concepts are mentally represented. There have been two main answers. The first, somewhat confusingly called the *prototype theory* suggests that people have a *summary representation* of the category, a mental description that is meant to apply to the category as a whole. (The significance of *summary* will become apparent when the next theory is described.) This description can be represented as a set of *weighted features* (Smith & Medin, 1981). The features are weighted by their frequency in the category. For the category of birds, having wings and feathers would have a very high weight; eating worms would have a lower weight; living in Antarctica would have a lower weight still, but not zero, as some birds do live there.

Donker	Blegdav
has thick windows	has steel windows
is red	is purple
divers live there	farmers live there
is under water	is in the desert
get there by submarine	get there by submarine
has fish as pets	has polar bears as pets

Figure **18.12.6**: If you were asked, “What kind of animal is this?” according to prototype theory, you would consult your summary representations of different categories and then select the one that is most similar to this image—probably a lizard! [Image: Adhi Rachdian, https://goo.gl/dQyUwf, CC BY 2.0, https://goo.gl/BRvSA7]

The idea behind prototype theory is that when you learn a category, you learn a general description that applies to the category as a whole: Birds have wings and usually fly; some eat worms; some swim underwater to catch fish. People can state these generalizations, and sometimes we learn about categories by reading or hearing such statements (“The kimodo dragon can grow to be 10 feet long”).

When you try to classify an item, you see how well it matches that weighted list of features. For example, if you saw something with wings and feathers fly onto your front lawn and eat a worm, you could (unconsciously) consult your concepts and see which ones contained the features you observed. This example possesses many of the highly weighted bird features, and so it should be easy to identify as a bird.

This theory readily explains the phenomena we discussed earlier. Typical category members have more, higher-weighted features. Therefore, it is easier to match them to your conceptual representation. Less typical items have fewer or lower-weighted features (and they may have features of other concepts). Therefore, they don't match your representation as well (less similarity). This makes people less certain in classifying such items. Borderline items may have features in common with multiple categories or not be very close to any of them. For example, edible seaweed does not have many of the common features of vegetables but also is not close to any other food concept (meat, fish, fruit, etc.), making it hard to know what kind of food it is.

A very different account of concept representation is the *exemplar theory* (**exemplar** being a fancy name for an example; Medin & Schaffer, 1978). This theory denies that there is a summary representation. Instead, the theory claims that your concept of vegetables is remembered examples of vegetables you have seen. This could of course be hundreds or thousands of exemplars over the course of your life, though we don't know for sure how many exemplars you actually remember.

How does this theory explain classification? When you see an object, you (unconsciously) compare it to the exemplars in your memory, and you judge how **similar** it is to exemplars in different categories. For example, if you see some object on your plate and want to identify it, it will probably activate memories of vegetables, meats, fruit, and so on. In order to categorize this object, you calculate how **similar** it is to each exemplar in your memory. These **similarity scores** are added up for each category. Perhaps the object is very similar to a large number of vegetable exemplars, moderately similar to a few fruit, and only minimally similar to some exemplars of meat you remember. These **similarity scores are compared, and the category with the highest score is chosen**.[2]

Why would someone propose such a theory of concepts? One answer is that in many experiments studying concepts, people learn concepts by seeing exemplars over and over again until they learn to classify them correctly. Under such conditions, it seems likely that people eventually memorize the exemplars (Smith & Minda, 1998). There is also evidence that ***close similarity*** to well-remembered objects has a **large effect on classification**. Allen and Brooks (1991) taught people to classify items by following a rule. However, they also had their subjects study the items, which were richly detailed. In a later test, the experimenters gave people new items that were very similar to one of the old items but were in a different category. That is, they changed one property so that the item no longer followed the rule. They discovered that people were often fooled by such items. **Rather than following the category rule they had been taught, they** seemed to **recognize the new item as being very similar** to an old one and so put it, incorrectly, into the same category.

Many experiments have been done to compare the prototype and exemplar theories. Overall, the **exemplar theory seems to have won most of these comparisons**. However, the experiments are somewhat limited in that they usually involve a small number of exemplars that people view over and over again. It is not so clear that exemplar theory can explain real-world classification in which people do not spend much time learning individual items (how much time do you spend studying squirrels? or chairs?). Also, given that some part of our knowledge of categories is learned through general statements we read or hear, it seems that there must be room for a summary description separate from exemplar memory.

Many researchers would now acknowledge that concepts are represented through multiple cognitive systems. For example, your knowledge of dogs may be in part through general descriptions such as "dogs have four legs." But you probably also have strong memories of some exemplars (your family dog, Lassie) that influence your categorization. Furthermore, some categories also involve rules (e.g., a strike in baseball). How these systems work together is the subject of current study.

> *[2] Actually, the decision of which category is chosen is more complex than this, but the details are beyond this discussion.*

18.12.7 Knowledge

The final topic has to do with how concepts fit with our broader knowledge of the world. We have been talking very generally about people learning the features of concepts. For example, they see a number of birds and then learn that birds generally have wings, or perhaps they remember bird exemplars. From this perspective, it makes no difference what those exemplars or features are—people just learn them. But consider two possible concepts of buildings and their features in Table 2.

Donker	Blegdav
has thick windows	has steel windows
is red	is purple
divers live there	farmers live there
is under water	is in the desert
get there by submarine	get there by submarine
has fish as pets	has polar bears as pets

Table 3. Examples of two fictional concepts

Imagine you had to learn these two concepts by seeing exemplars of them, each exemplar having some of the features listed for the concept (as well as some idiosyncratic features). Learning the donker concept would be pretty easy. It seems to be a kind of underwater building, perhaps for deep-sea explorers. Its features seem to go together. In contrast, the blegdav doesn't really make sense. If it's in the desert, how can you get there by submarine, and why do they have polar bears as pets? Why would farmers live in the desert or use submarines? What good would steel windows do in such a building? This concept seems peculiar. In fact, if people are asked to learn new concepts that make sense, such as donkers, they learn them quite a bit faster than concepts such as blegdavs that don't make sense (Murphy & Allopenna, 1994). Furthermore, the features that seem connected to one another (such as being underwater and getting there by submarine) are learned better than features that don't seem related to the others (such as being red).

Such effects demonstrate that when we learn new concepts, we try to connect them to the knowledge we already have about the world. If you were to learn about a new animal that doesn't seem to eat or reproduce, you would be very puzzled and think that you must have gotten something wrong. By themselves, the prototype and exemplar theories don't predict this. They simply say that you learn descriptions or exemplars, and they don't put any constraints on what those descriptions or exemplars are. However, the *knowledge approach* to concepts emphasizes that concepts are meant to tell us about real things in the world, and so our knowledge of the world is used in learning and thinking about concepts.

We can see this effect of knowledge when we learn about new pieces of technology. For example, most people could easily learn about tablet computers (such as iPads) when they were first introduced by drawing on their knowledge of laptops, cell phones, and related technology. Of course, this reliance on past knowledge can also lead to errors, as when people don't learn about features of their new tablet that weren't present in their cell phone or expect the tablet to be able to do something it can't.

One important aspect of people's knowledge about categories is called **psychological essentialism** (Gelman, 2003; Medin & Ortony, 1989). People tend to believe that some categories—most notably natural kinds such as animals, plants, or minerals—have an underlying property that is found only in that category and that causes its other features. Most categories don't actually have essences, but this is sometimes a firmly held belief. For example, many people will state that there is something about dogs, perhaps some specific gene or set of genes, that all dogs have and that makes them bark, have fur, and look the way they do. Therefore, decisions about whether something is a dog do not depend only on features that you can easily see but also on the assumed presence of this cause.

Donker	Blegdav
has thick windows	has steel windows
is red	is purple
divers live there	farmers live there
is under water	is in the desert
get there by submarine	get there by submarine
has fish as pets	has polar bears as pets

Figure **18.12.7**: Although it may seem natural that different species have an unchangeable "essence," consider evolution and everything's development from common ancestors. [Image: Marc Dragiewicz, https://goo.gl/E9v4eR, CC BY-NC-SA 2.0, https://goo.gl/Toc0ZF]

Belief in an essence can be revealed through experiments describing fictional objects. Keil (1989) described to adults and children a fiendish operation in which someone took a raccoon, dyed its hair black with a white stripe down the middle, and implanted a "sac of super-smelly yucky stuff" under its tail. The subjects were shown a picture of a skunk and told that this is now what the animal looks like. What is it? Adults and children over the age of 4 all agreed that the animal is still a raccoon. It may look and even act like a skunk, but a raccoon cannot change its stripes (or whatever!)—it will always be a raccoon.

Importantly, the same effect was not found when Keil described a coffeepot that was operated on to look like and function as a bird feeder. Subjects agreed that it was now a bird feeder. Artifacts don't have an essence.

Signs of essentialism include (a) objects are believed to be either in or out of the category, with no in-between; (b) resistance to change of category membership or of properties connected to the essence; and (c) for living things, the essence is passed on to progeny.

Essentialism is probably helpful in dealing with much of the natural world, but it may be less helpful when it is applied to humans. Considerable evidence suggests that people think of gender, racial, and ethnic groups as having essences, which serves to emphasize the difference between groups and even justify discrimination (Hirschfeld, 1996). Historically, group differences were described by inheriting the blood of one's family or group. "Bad blood" was not just an expression but a belief that negative properties were inherited and could not be changed. After all, if it is in the nature of "those people" to be dishonest (or clannish or athletic ...), then that could hardly be changed, any more than a raccoon can change into a skunk.

Research on categories of people is an exciting ongoing enterprise, and we still do not know as much as we would like to about how concepts of different kinds of people are learned in childhood and how they may (or may not) change in adulthood. Essentialism doesn't apply only to person categories, but it is one important factor in how we think of groups.

18.12.7.1 Summary

Concepts are central to our everyday thought. When we are planning for the future or thinking about our past, we think about specific events and objects in terms of their categories. If you're visiting a friend with a new baby, you have some expectations about what the baby will do, what gifts would be appropriate, how you should behave toward it, and so on. Knowing about the category of babies helps you to effectively plan and behave when you encounter this child you've never seen before. Such inferences from knowledge about a category are highly adaptive and an important component of thinking and intelligence.

Learning about those categories is a complex process that involves seeing exemplars (babies), hearing or reading general descriptions ("Babies like black-and-white pictures"), general knowledge (babies have kidneys), and learning the occasional rule (all babies have a rooting reflex). Current research is focusing on how these different processes take place in the brain. It seems likely that these different aspects of concepts are accomplished by different neural structures (Maddox & Ashby, 2004). However, it is clear that the brain is genetically predisposed to seek out similarities in the environment and to represent groupings of things forming categories that can be used to make inferences about new instances of the category which have never been encountered before. In this way knowledge is organized and expectations from this knowledge allow improved adaptation to newly encountered environmental objects and situations by virtue of their similarity to a known category previously formed (Koenigshofer, 2017).

Another interesting topic is how concepts differ across cultures. As different cultures have different interests and different kinds of interactions with the world, it seems clear that their concepts will somehow reflect those differences. On the other hand, the structure of the physical world also imposes a strong constraint on what kinds of categories are actually useful. The interplay of culture, the environment, and basic cognitive processes in establishing concepts has yet to be fully investigated.

18.12.7.2 Discussion Questions

1. Pick a couple of familiar categories and try to come up with definitions for them. When you evaluate each proposal (a) is it in fact accurate as a definition, and (b) is it a definition that people might actually use in identifying category members?
2. For the same categories, can you identify members that seem to be “better” and “worse” members? What about these items makes them typical and atypical?
3. Going around the room, point to some common objects (including things people are wearing or brought with them) and identify what the basic-level category is for that item. What are superordinate and subordinate categories for the same items?
4. List some features of a common category such as tables. The knowledge view suggests that you know reasons for why these particular features occur together. Can you articulate some of those reasons? Do the same thing for an animal category.
5. Choose three common categories: a natural kind, a human artifact, and a social event. Discuss with class members from other countries or cultures whether the corresponding categories in their cultures differ. Can you make a hypothesis about when such categories are likely to differ and when they are not?

18.12.7.3 Vocabulary

Basic-level category
The neutral, preferred category for a given object, at an intermediate level of specificity.

Category
A set of entities that are equivalent in some way. Usually the items are similar to one another.

Concept
The mental representation of a category.

Exemplar
An example in memory that is labeled as being in a particular category.

Psychological essentialism
The belief that members of a category have an unseen property that causes them to be in the category and to have the properties associated with it.

Typicality
The difference in “goodness” of category members, ranging from the most typical (the prototype) to borderline members.

18.12.7.4 References

- Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule.*Journal of Experimental Psychology: General*, *120*, 3–19.
- Anglin, J. M. (1977). *Word, object, and conceptual developmen*t. New York, NY: W. W. Norton.
- Berlin, B. (1992). *Ethnobiological classification: Principles of categorization of plants and animals in traditional societies*. Princeton, NJ: Princeton University Press.
- Brown, R. (1958). How shall a thing be called? *Psychological Review, 65*, 14–21.
- Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford, UK: Oxford University Press.
- Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior, 18*, 441–461.
- Hirschfeld, L. A. (1996). *Race in the making: Cognition, culture, and the child's construction of human kinds*. Cambridge, MA: MIT Press.
- Horton, M. S., & Markman, E. M. (1980). Developmental differences in the acquisition of basic and superordinate categories. *Child Development*, 51, 708–719.
- Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

- Koenigshofer, K. A. (2017). General Intelligence: Adaptation to Evolutionarily Familiar Abstract Relational Invariants, Not to Environmental or Evolutionary Novelty. *The Journal of Mind and Behavior*, 38(2):119-153.
- Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-based systems of perceptual category learning. *Behavioural Processes*, *66*, 309–332.
- Mandler, J. M. (2004). *The foundations of mind: Origins of conceptual thought*. Oxford, UK: Oxford University Press.
- Mareschal, D., Quinn, P. C., & Lea, S. E. G. (Eds.) (2010). *The making of human concepts*. Oxford, UK: Oxford University Press.
- McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition*, *6*, 462–472.
- Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). Cambridge, UK: Cambridge University Press.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207–238.
- Mervis, C. B. (1987). Child-basic object categories and early lexical development. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 201–233). Cambridge, UK: Cambridge University Press.
- Murphy, G. L., & Allopenna, P. D. (1994). The locus of knowledge effects in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 904–919.
- Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 70–84.
- Norenzayan, A., Smith, E. E., Kim, B. J., & Nisbett, R. E. (2002). Cultural preferences for formal versus intuitive reasoning. *Cognitive Science*, *26*, 653–684.
- Rosch, E., & Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605.
- Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). *Basic objects in natural categories. Cognitive Psychology, 8*, 382–439.
- Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 491–502.
- Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111–144). New York, NY: Academic Press.
- Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.
- Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1411–1436.
- Tanaka, J. W., & Taylor, M. E. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *15*, 121–149.
- Wisniewski, E. J., & Murphy, G. L. (1989). Superordinate and basic category names in discourse: A textual analysis. *Discourse Processes*, *12*, 245–261.

18.12.7.5 Attributions

18.12.7.6 Authors



- Gregory Murphy is Professor of Psychology at New York University. He previously taught at the University of Illinois and Brown University. His research focuses on concepts and reasoning, and he is the author of The Big Book of Concepts (MIT Press, 2002).

18.12.7.7 Creative Commons License



18.12.7.7.1 **Adapted by Kenneth Koenigshofer, PhD, from Categories and Concepts by Gregory Murphy, licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Permissions beyond the scope of this license may be available in our Licensing Agreement.**

18.12.7.8 How to cite this Noba module using APA Style

Murphy, G. (2021). Categories and concepts. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology.* Champaign, IL: DEF publishers. Retrieved from http://noba.to/6vu4cpkt

---

This page titled 18.12: Chapter 14- Problem Solving, Categories and Concepts is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

18.13: Chapter 14- The Brain and Evolution of Cognition

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Briefly describe two general approaches to human cognition and its evolution: continuity vs. qualitative gap
2. List and briefly describe the theories of the evolution of human intelligence discussed in this module
3. Briefly explain the possible role of cooking in human brain evolution
4. List the brain areas associated most closely with cognition and intelligence in humans and in several other animal groups

18.13.1 Overview

In this module, we examine ideas about the evolutionary origins of human cognition and intelligence. A number of approaches to this issue are discussed. Attempts to understand the relationship between human intelligence and intelligence in other species can be divided into two primary categories. One is based on the assumption of continuity between intelligence in human and non-human animals. The second assumes that there is a qualitative gap between human and non-human animal intelligence. Factors hypothesized to be important in the evolution of human intelligence include bipedalism, encephalization and increased information processing capacity, greater density of cortical neurons per unit of cortical mass, absolute number of neurons in the brain, meat-eating, cooking, complex social organization, "mental time travel," anticipation of future drive states, greater capacity for abstraction and visualization of future, ability to envision future actions and their outcomes, anticipatory cognition, tool use, connectivity between cortical areas and cortical networks, niche construction, and language.

18.13.2 Evolution of Cognition

by Kenneth A. Koenigshofer, Ph.D., Chaffey College

Animals use a range of mechanisms to respond to environmental change. These vary in complexity and sophistication from simple reflexes and simple forms of learning such as habituation to complex learning, reasoning and cognition. Cognition in this comparative context refers to the ability to acquire, process, and retain information to guide behavior and decision-making toward successful adaptation (van Horik & Emery, 2011).

Two different approaches to human cognition have dominated research and theory about evolution of cognition and intelligence. One assumes **continuity** between human and animal cognition and that **differences are primarily a matter of degree**. The **second approach** hypothesizes that there are **major qualitative gaps** or jumps in cognition and intelligence between humans and animals.

Gibson (2002) argues that the differences in cognition and intelligence between humans and other animals are a matter of degree and thus assumes continuity. According to Gibson, human cognition and intelligence exceeds that of other animals because of "increased information processing capacities" of the human brain due to enlargement of the human neocortex, cerebellum, and basal ganglia. Gibson proposes that increased information processing capabilities of the human brain led to development of human language, social cognition, and tool-making.

Herculano-Houzel (2012) offers anatomical data which supports continuity between human and non-human primate brains, thereby supporting Gibson's views. She argues that the human brain is a "scaled-up primate brain." Regarding the cerebral cortex specifically, she notes that "primate cortices contain many more neurons than nonprimate cortices of a similar mass" (Herculano-Houzel, et al., 2015, p. 159). This greater density of cortical neurons in primate cerebral cortex compared to other mammal species means that "primates are . . . subject to a different scaling rule, with more neurons for a given body mass compared to other mammalian clades. . . while larger bodies have neurons in the the ROB [rest of the brain] that are on average larger in proportion to the linear dimension of the body, the number of brain neurons is not dictated simply by body mass" (Herculano-Houzel, et al.,

2015, p. 161). In other words, bigger bodies do not necessarily mean more brain neurons, but perhaps just bigger neurons. More specifically, humans have the largest brain and the largest number of neurons among the primates, perhaps as many as three times the number of neurons found in gorillas and orangutans (which have the next largest living primate brains after humans), yet gorillas can grow to be three times the body size of humans. This larger body size has not resulted in greater numbers of brain neurons in gorillas.

Why did human brains become so much larger, with so many more neurons, than expected for their body size (i.e. greater encephalization) when compared to other primates, and mammals in general? To answer this question, we need to consider the large amount of metabolic energy that a large brain with a large number of neurons requires. The energy required to run a brain increases as a function of the number of neurons that make up the brain. Brains with a lot of neurons require a lot of metabolic energy. Without sufficient energy supplies from the food available to a species, the number of brain neurons can only increase to a limit set by the available metabolic energy gained by feeding. Eating meat is one way to obtain more calories from the environment, and eating cooked meat is even better. Given that "the energetic cost of the brain is a linear function of its numbers of neurons," Fonseca-Azevedo and Herculano-Houzel (2012, p. 18571) argue that "metabolic limitations that result from the number of hours available for feeding and the low caloric yield of raw foods impose a tradeoff between body size and number of brain neurons [in non-human primates] . . . This limitation was probably overcome in Homo erectus with the shift to a cooked diet. Absent the requirement to spend most available hours of the day feeding, the combination of newly freed time and a large number of brain neurons affordable on a cooked diet may thus have been a major positive driving force to the rapid increase in brain size in human evolution." This argument is based on the observation that cooked foods are easier to digest and thus require less metabolic energy to process, leaving more usable energy after digestion compared to uncooked food. The use of fire and the invention of cooking, especially of meat, a concentrated source of metabolic energy (see below), by our ancient ancestors overcame the metabolic limitations that constrained brain size in other primates. This cultural innovation set the stage for larger relative brain size and greater number of brain neurons in human evolution beginning with Homo erectus, the oldest member of genus Homo for which evidence of the use of fire for cooking has been found. In short, eating meat, especially cooked meat, provided the metabolic energy required to support bigger brains with larger numbers of neurons, freeing hominid brain evolution beginning with Homo erectus from the metabolic limitations that prevented increased encephalization in other primates and in mammals in general.

Lakatos and Janka (2008) propose that upright posture and bipedalism in human ancestors freed the hands for evolution of improved manual dexterity leading to tool-making. Weapons improved the hunting of animals. Eating meat, as just discussed, provided more calories and fat in the diet making possible the evolution of larger brains which require large amounts of caloric energy (modern human brains comprise 2% of body weight but consume 20% of the body's metabolic energy). The human brain tripled in size over 3.5 million years from an average of 450 cm^3 in Australopithecinae to 1350 cm^3 in Homo sapiens. The appearance and stabilization of the FOXP2 gene (see last module in this chapter) set the stage for human language, which, according to Lakatos and Janka (p. 220), was the "basis for abrupt evolution of human intelligence," and coincided with the first appearance and the rapid spread of Homo sapiens over the Earth. They point out that the "level of intelligence is related anatomically to the number of cortical neurons and physiologically to the speed of conductivity of neural pathways, the latter being dependent on the degree of myelinization" (p.220). In other words, both the number of neurons and the amount of myelin on axons in a brain (the more myelinated axons, the faster the conduction speeds between neurons and the faster and more efficient information processing becomes, presumably increasing intelligence. In addition, they claim that language provided more effective communication among cooperating hunters and food-gatherers, requiring less energy to achieve effective cooperation, thereby providing competitive advantage in obtaining food. They also suggest that better mental skills, favored in sexual competition for mates, helped stabilize "cleverness" genes. They support the view that major qualitative changes occurred during evolution of human cognition. On their view, evolution of language was the basic condition required for conscious thinking, creating a qualitative change in human cognition which made humans and human cognition different from all other species.

Another hypothesis is the "social brain" theory of the evolution of human cognition. This stems from the observation that primate brains are roughly twice the size expected based on body size, compared to other mammals, and that the size of cerebral cortex is correlated with the size of the primate social group characteristic of a species (Barrett & Henzi, 2005). The basic idea is that living in groups created local competition among members of the social group for scarce resources which would lead to the evolution of cognitive skills to out-wit one's competitors in the social group. Increased cognitive skills would select for even greater cognitive skills because of the increased competition. However, this view of competition among individuals as the driving force in the evolution of primate cognition doesn't take into account mounting evidence that much of primate social behavior, especially in

humans, involves altruistic behavior including cooperation, sharing, trade, group defense, and other prosocial behaviors which can evolve when members of a group are inter-dependent for survival and reproduction of viable offspring. This may occur through kin selection (see previous section on kin selection in Chapter 3) or **group selection** wherein natural selection acts on the group as a whole (Barrett & Henzi, 2005). Nevertheless, social life does involve cognitive skills not required in non-social animals, although monkeys may use less demanding cognitive skills than do apes in their social interactions. A major problem with the social brain hypothesis as an explanation for the evolution of cognition is that the argument "can sometimes appear circular: primates have large brains because their social lives are cognitively demanding, and their lives are cognitively demanding because they have large brains that allow them to produce more complex forms of social behavior" (Barrett & Henzi, 2005).

Reader, et al. (2011) note that there are conflicting hypotheses regarding the nature and structure of primate cognition. One view is that primate cognition is divided into specialized, independently evolving modules ("massive modularity"), each of which processes information using "rules" specific to solution of a specific class of adaptive problem--such as mate selection, or predator avoidance, or securing and maintaining status in one's social group, and so on. The opposite of this modular view of cognition is that primate intelligence is a single general process, attributable to a single 'general intelligence' factor, g, the evolutionary origin of which remains controversial (see Koenigshofer, 2017). Reader, et al. (2011) collected ecologically relevant measures of cognition including reported incidences of behavioral innovation, social learning, tool use, extractive foraging and tactical deception, in 62 primate species. All measures exhibited strong positive associations, after statistically controlling for multiple potential confounds, creating a highly correlated composite of cognitive traits which they took as a species-level measure of general intelligence, which they called, "primate g(S)." They argue that *primate g(S)* suggests that "social, technical and ecological abilities have coevolved in primates, indicative of an across-species general intelligence that includes elements of cultural intelligence," such as social learning ability and tool use. They also reported that primate g(S) correlated "with both brain volume and captive learning performance measures." These researchers conclude that their "findings question the independence of cognitive traits and do not support 'massive modularity' in primate cognition, nor an exclusively social model of primate intelligence" given that no relationship between social group size (complexity) and cognitive performance in lab tests was found. This suggests that social complexity as measured by group size may not have been as important a driver of the evolution of primate general intelligence as is widely believed. "High general intelligence has independently evolved at least four times, with convergent evolution in capuchins, baboons, macaques and great apes" (Reader, Hager, & Laland, 2011, p. 1017). "The strong correlation between distinct measures of primate cognitive performance is strikingly evocative of the correlations in performance on different IQ tests observed in humans. A possible explanation for this correspondence is that the *g* factor reported in humans reflects underlying general processes that evolved in common ancestors and are thus shared in our extant primate relatives" (Reader, Hager, & Laland, 2011, p. 1024). This view is consistent with arguments made in module 14.2 that general intelligence is found in many non-human animals as well as in humans (Koenigshofer, 2017).

Another intriguing hypothesis about the evolution of cognition and the gap or discontinuity between human cognition and cognition in other animals is the concept of mental time travel. "Mental time travel comprises the mental reconstruction of personal events from the past (episodic memory) and the mental construction of possible events in the future" (Suddendorf, Addis, & Corballis, 2011, p. 344). This projection of the self into future-time using visual imagery (i.e. imagination) was discussed as an important component of general intelligence (Koenigshofer, 2017) in module 14.2. According to Suddendorf and colleagues, the ability to "time travel" to the past doesn't occur in human children until about three and a half years old, and is associated with episodic memory, a type of autobiographical memory dependent upon functioning of the prefrontal cortex involving understanding that the memory is of an experience of a past self. This awareness of "pastness" may be beyond the capability of most animals, particularly if they do not have a conscious sense of self. Furthermore, mental time travel requires ability to dissociate from one's current state of mind to generate a personal memory or a projection into possible future states of the world (imagination again). This involves so-called "executive functions" of the frontal lobe. "'Executive function' is an umbrella term for the mental operations which enable an individual to disengage from the immediate context in order to guide behavior by reference to mental models and future goals" (Hughes, Russel, & Robbins, 1994, p. 477, as referenced in Suddendorf, et al., 2011). As was discussed in earlier modules, this ability requires the **prefrontal cortex**. Persons with autism and those with prefrontal damage have difficulty with executive functions. This cognitive impairment is revealed by inability to switch behavior previously employed when changes in the task require, and also by apparent difficulty in using mental models as evidenced by deficits in planning future behavior. Although animals in general certainly anticipate future in various ways, including implicit learning such as conditioning, whether they construct mental models of future is for most species an open question; however, for chimpanzees at least, evidence from

Kohler's classic studies showed sudden insight by chimpanzees in problem solving implying "constructive thought with an eye to the future solution of a problem" suggesting mental time travel to the future in these animals. Studies by Jane Goodall of chimpanzee tool-making "also suggest flexible forethought. For example, the chimpanzees at Gombe [in Tanzania] manufactured pointed tools from sticks at one place to use them later for termite fishing at another place that was out of sight (Goodall, 1986)." Furthermore, chimpanzees have been observed to transport stones for cracking open nuts from one region where stones are found to another feeding location when stones are not found. This shows ability to imagine future and to engage in current behavior that anticipates and plans for an expected future.

However, one hypothesis by Bischof (1978, 1985) is that non-human animals cannot anticipate future needs or drive states and so are bound by their present motivational state. For example, a full-bellied lion is no threat to passing zebra, but a full-bellied human may be (Suddendorf, Addis, & Corballis, 2011, p. 349). This difference may reflect a trend in evolution. "Bischof (1985) suggests that, in the course of evolution, there was a progressively increasing gap between drive and action. Great apes display quite extensive gaps; they can postpone the immediate enactment of their current drive, and make plans to receive gratification at a later point in time." The role of human anticipatory cognition is not only central to human behavior but shapes human society as well (Suddendorf, Addis, & Corballis, 2011, p. 351). For example, we grow, harvest, store, and ship food for future use using all sorts of complex behavioral practices oriented to the future use of the food. We have complex legal systems that are based on the idea that current punishment for a past crime will prevent future crime. We have invented nuclear arsenals ready for launch based on the assumption that such weapons will prevent future attacks on us. We have the institution of marriage which we anticipate will help satisfy our needs for emotional and financial security (including needs for shelter and food), for reproduction of offspring, and for future sexual gratification.

Osvath and Gärdenfors (2005, p. 1), following Bischof's hypothesis, state "Anticipatory cognition, that is, the ability to mentally represent future needs, is a uniquely human trait that has arisen along the hominid line . . . [resulting in] stone tool manufacture, . . . transports over long ranges of tools as well as food and of the use of accumulation spots. . . .[T]his niche promoted the selection for anticipatory cognition, in particular planning for future goals. Once established, anticipatory cognition opened up for further cultural developments, such as long ranging migration, division of labor, and advanced co-operation and communication, all of which one finds evidence for in Homo ergaster/erectus. . . . A distinctive feature of human thinking that contrasts with the cognition of other primates is our capacity to form mental representations of the distant future. . . . planning in primates and other animals suggest that they can only plan for present needs (this is dubbed the Bischof-Kohler hypothesis) . . . "

The evolution of anticipatory cognition resulted in "niche construction," a term from evolutionary biology. "Niche construction can be defined as the systematic changes that organisms bring about in their environments (Day, Laland & Odling-Smee, 2003). A wide variety of organisms construct parts of their environment: spiders making webs, birds building nests, beavers constructing dams and plants altering their chemical surroundings" (Osvath and Gärdenfors (2005, p. 2). Of particular interest is the fact that niche construction "can actually change the evolutionary dynamics" by altering selection pressures leading to "unusual evolutionary dynamics, . . . [including] ecological inheritance. This is characterized by offspring inheriting the altered ecology from its ancestors – an ecology with its own selective pressures. Implementing niche construction in evolutionary theory involves two kinds of modifying processes, natural selection and niche construction, as well as two kinds of descent, genetic and ecological inheritance (Day, Laland & Odling-Smee, 2003)" (Osvath and Gärdenfors (2005, p. 2). According to Osvath and Gärdenfors (2005, p.1), "the cultural niche that was created by the use of Oldowan [stone] tools, including transport of tools and carcasses (Plummer, 2004), has led to a selection for anticipatory cognition, and in particular anticipatory planning." These events produced a qualitative gap in cognitive evolution between humans and apes--our capacity for representation of the distant future, including planning for future needs, capacities which appear to be absent in other animals including our nearest primate relatives, the great apes.

"The ability to envision various actions and their consequences is a necessary requirement for an animal to be capable of planning. . . . An organism is planning its actions if it has . . . representations of (1) goal and start situations, (2) sequences of actions, and (3) the outcomes of actions. . . . The hominids have presumably been the most cultural dependent species ever and thus the most powerful and flexible niche constructors (to various degrees, of course, and with Homo sapiens at the present extreme). A consequence of this is that the hominids should exhibit great evolutionary resistance to changes in the natural environment, but also be capable of far-reaching evolutionary change caused by some ground-breaking cultural innovation (Laland, Odling-Smee & Feldman, 2000). The key point is that some of our major cognitive adaptations might not only be a result of new demands from a changing natural environment, but to a larger degree emerge from a highly constructed environment, including both social and

artefactual elements. It is not sufficient to look at habitat, climate and other natural ecological factors, when trying to understand how humans evolved" (Osvath and Gärdenfors (2005, p. 2). Niche construction must also be considered, and Osvath and Gärdenfors argue that niche construction played a large role in human cognitive evolution.

They explain further: "Culture is the most forceful medium in niche construction, because culturally transmitted traits spread much more rapidly than genetically transmitted traits do. On one hand, culture can create a very strong selection and thereby increase the rate of evolutionary changes. On the other hand, culture can buffer out particular natural selection pressures (Laland, Odling-Smee & Feldman, 2000). A cultural innovation, constituting a form of niche construction, can create a behavioural drive that accelerates evolution. The same innovation can create a niche that blocks natural selective pressures. For example, humans live in cold climates without having any major biological adaptations to cold, because they rely on the cultural niche created by clothes and external heat sources. The consequence is that the naked human body does not exhibit cold adapted traits to the extent that would be expected from a mammal living in a cold climate. . . The general thesis is that the hominids themselves created a great deal of the selective pressures that eventually made us human."

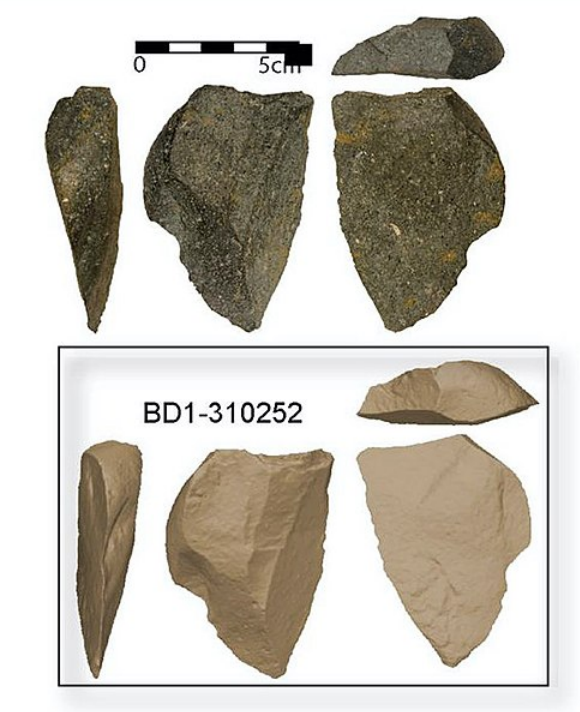




Figure 18.13.1: Human niche construction: from stone tools to skyscrapers. (Images from Wikimedia Commons. (Left) File:Bokol Dora Stone Tool (cropped).jpg; https://commons.wikimedia.org/wiki/F..._(cropped).jpg; by David R. Braun; licensed under the Creative Commons Attribution-Share Alike 2.0 Generic license. (Right) File:Nakanoshima Skyscrapers in 201705 002.jpg; https://commons.wikimedia.org/wiki/F...201705_002.jpg; by Mc681; licensed under the Creative Commons Attribution-Share Alike 4.0 International license.

Another hypothesis about the difference between human and animal cognition emphasizes capacities for abstraction. Ability for high levels of abstraction in causal and similarity relations (Penn et al., 2008) and in covariation detection (Wasserman, 1993b) may be central to the phylogenetically unprecedented achievements of the human intellect (Koenigshofer, 2017), including niche construction. In a review of the comparative literature, Penn et al. (2008) find “no compelling evidence” that any nonhuman animal can reason about the “relation between relations,” something that humans do readily. Neither do they find convincing evidence for analogical reasoning (dependent upon judgments of abstract similarities across domains) in nonhuman animals. Children readily understand the similarity between a dog and its dog house and a bird and its nest. Consistent with Penn et al. (2008), Walker and Gopnik (2014) report that although human toddlers readily detect abstract relational causality quickly with only a few training trials, non-human primates have great difficulty with similar tasks even after hundreds of trials. Penn et al. (2008, p. 123) believe that what distinguishes human cognition from that in non-human animals is “the ability to reason about higher-order relations,” and that this capacity for high levels of abstraction “subserves a wide variety of distinctively human capabilities.”

As discussed in an earlier module, research in molecular genetics suggests one possible explanation for how human ability for high levels of abstraction may have come about during the course of evolution. Pollard (2009) comparing human and chimpanzee genomes found “massive mutations” in humans in the “DNA switches” controlling size and complexity of cerebral cortex, extending the period of prenatal cell division in human cerebral cortex by several days compared to our closest primate relatives. Research using artificial neural networks suggests that increasing cortical complexity leads to sudden leaps in ability for abstraction and rule-like understanding of general principles (Clark, 1993), a hallmark of high intelligence. Findings by Penn et al. (2008) strongly suggest that superior ability for abstraction of relational information may be the key component explaining differences in general intelligence between humans and nonhuman animals. These abilities may involve **anterior dorsolateral prefrontal cortex** in humans (Kroger et al., 2002; Reeber et al., 1998; see Koenigshofer, 2017).

Frontal cortex is found in other mammals as well, so the question remains as to why the human brain with its frontal lobe has capabilities which exceed those of all other animals. In this regard, Roth and Dicke (2012) note "Primates are, on average, more intelligent than other mammals, with great apes and finally humans on top. They generally have larger brains and cortices, and because of higher relative cortex volume and neuron packing density (NPD), they have much more cortical neurons than other mammalian taxa with the same brain size. Likewise, information processing capacity is generally higher in primates due to short interneuronal distance and high axonal conduction velocity [see discussion above of findings by Herculano-Houzel, et al., 2015, and Lakatos and Janka, 2008]. Across primate taxa, differences in intelligence correlate best with differences in number of cortical neurons and synapses plus information processing speed. The human brain stands out by having a large cortical volume with relatively high NPD, high conduction velocity, and high cortical parcellation. All aspects of human intelligence are present at least in rudimentary form in nonhuman primates or some mammals or vertebrates except syntactical language. The latter can be regarded as a very potent 'intelligence amplifier.'" This last point is related to human ability to share knowledge and ideas facilitating cultural transmission of learned information and cultural evolution.

Another important finding involves the connectivity between cortical areas. For example, prefrontal white matter volume is disproportionately larger in humans than in other primates (Schoenemann, et al., 2005). However, this finding might be related to more gyri in the frontal cortex of humans compared to other primates. Nevertheless, these observations may reflect "connections between cortical areas capable of relaying and processing information at greater speeds, facilitating complex cognitive function" (Schenker, et al., 2005, p.564).

Williams (2020) emphasizes the role of cognitive interactions within human groups to account for the cognitive achievements of humans compared to other animals. According to Williams, General Collective Intelligence (GCI) can be defined as a system that organizes groups into a single collective cognition with the potential for vastly greater general problem solving ability than that of any individual in the group. According to Williams, GCI represents a phase transition in human intelligence. When humans collectively put their cognitive capacities and activity together to solve a particular problem a collective intelligence emerges which is far more powerful than the cognitive abilities of any single individual. Communication and social cooperation between

individual humans "putting their heads together" facilitated by cultural innovations for sharing information and ideas, within and across generations (such as writing, the internet, scientific organizations, research universities and other group cognitive efforts), create a kind of "superintelligence" not found in any other species. The powerful adaptive effects of collaborative problem solving combined with cultural transmission of learned information and ideas from generation to generation, built upon by each succeeding generation, is essential to the cognitive achievements which make us so different from other species. While other species such as dogs, apes, killer whales, dolphins, elephants and so on are quite intelligent, none has created an international space station or flown robots to Mars.

Nevertheless, cognition and intelligence have evolved in a wide variety of species. "Within the animal kingdom, complex brains and high intelligence have evolved several to many times independently, e.g. among ecdysozoans in some groups of insects (e.g. blattoid, dipteran, hymenopteran taxa, e.g. bees, wasps), among lophotrochozoans in octopodid molluscs (e.g. octopus), among vertebrates in teleosts (e.g. cichlid fish), corvid and psittacid birds (e.g. ravens, crows, parrots), and cetaceans e.g. whales, dolphins, porpoise), elephants and primates. High levels of intelligence are invariantly bound to multimodal centres such as the **mushroom bodies in insects**, the **vertical lobe in octopodids**, the **pallium in birds** and the **cerebral cortex in primates**, all of which contain highly ordered associative neuronal networks. The driving forces for high intelligence may vary among the mentioned taxa, e.g. needs for spatial learning and foraging strategies in insects and cephalopods, for social learning in cichlids, instrumental learning and spatial orientation in birds and social as well as instrumental learning in primates" (Roth, 2015, p. 0049). According to Roth, high intelligence, which evolved in a large variety of species, is dependent not only upon factors identified by others such as number of brain neurons, their packing density, myelination and speed of processing, but also upon the presence of "multimodal centers" where processing events from multiple sensory modalities can be integrated into whole object and whole event representations. Keep in mind, that as was discussed in Module 14.2, cognition and intelligence in any species are part of the guidance systems that direct behavior toward successful adaptation in response to the challenges and opportunities presented by the environment. In the human case, the control of fire for cooking food, especially meat, permitting larger numbers of brain neurons in Homo erectus and in the members of the Homo genus which followed appears, to have been critical for setting hominids on the path to brain expansion and increasingly sophisticated cognition. But the fossil record shows us that a large brain was a prominent feature not only of ancient Homo sapiens in Africa but also of the Neanderthals in Europe, east Asia, and the Middle East. Although, Neanderthals had a larger visual cortex than modern humans, accounting for some of their enlarged cranial capacity, their frontal lobes, associated with human intelligence, were approximately the same size as the frontal lobes of modern humans, suggesting the possibility of high intelligence in the Neanderthals.

18.13.3 Differentiation of the Frontal Cortex

The role of the frontal cortex in human intelligence cannot assume a uniform set of regions comprising this cortex. Instead research shows that the frontal cortex is differentiated into regions specialized for different functional roles. For example, Schenker et al. (2005, p. 547-548) note that "this large cortical territory comprises anatomical subdivisions with rather distinct functional attributes. The dorsal sector of the frontal lobe is involved in perception, response selection, working memory, and problem solving (Owen, 1997; Bechara et al., 1998; Petrides, 2000; Pochon et al., 2001). It also includes territories activated during language production in humans (Foundas, 2001) and most of the primary motor and premotor cortices (Zilles et al., 1995). The mesial frontal cortex, including the anterior cingulate, is important for processing emotional stimuli and for production of affective responses (Cummings, 1993; Rezai et al., 1993), planning and initiation of voluntary motor sequences (Tanji and Mushiake, 1996), theory of mind (Fletcher et al., 1995; Gallagher et al., 2000; Stuss et al.,2001), and attention management (Carter et al., 1999; Dagher et al., 1999). The orbital frontal cortex is known to be involved in the evaluation of actions based on emotional reinforcers (Damasio, 1994; Stone et al., 1998; Rolls, 2000)." It is not unreasonable to speculate that these functions of the orbital frontal cortex may play a role in human ability to plan behavior for future needs, which according to the Bischof's hypothesis, discussed above, is a key feature of human cognition setting it apart from cognition in other primates.

18.13.4 Evolution of Increased Brain Size in Neanderthals and Early Humans

As discussed above, theorists have hypothesized many different factors to account for the expansion of the human brain and increases in intelligence. Fonseca-Azevedo and Herculano-Houzel (2012) hypothesized that a cooked diet was key to greater encephalization (larger brain and more neurons)--humans have the largest brain and the largest number of neurons among the living primates, perhaps as many as three times the number of neurons found in gorillas and orangutans. Lakatos and Janka (2008)

proposed that upright posture and bipedalism in human ancestors freed the hands for improved tool-making and production of weapons which improved hunting, providing a richer diet of meat capable of sustaining evolution of larger brains. Also recall the "social brain" hypothesis that the evolution of intelligence and cognition in humans was stimulated by increased information processing required by complex social interactions.

18.13.4.1 Evolution of Brain Size and Intelligence in Neanderthals

Now consider the Neanderthals whose Mousterian manufacturing indicates higher levels of conceptualization. Neanderthal fossil skulls show brain size equal to or slightly larger than that of modern humans (see figure below). The size of frontal cortex in humans and Neanderthals is essentially the same. Considering their large brain size and frontal lobes comparable to those of modern humans, some anthropologists suggest that Neanderthal cognitive and social capacities were not significantly different from that in humans 30,000 years ago (Hayden, 2012). Given these facts, we may wonder to what extent the above hypotheses about the evolution of human intelligence might also apply to Neanderthals. Spear points at Neanderthal sites certainly suggest hunting and meat-eating by Neanderthals. Neanderthals also cooked their food (Hayden, 2012; Henry, 2017). There is evidence that Neanderthals lived in groups (although perhaps smaller than humans of 40,000 years ago) and may have had complex social interactions (Duveau et al., 2019; Hayden, 2012). Could any or all of these factors, hypothesized as catalysts for the evolution of complex cognition in humans, have also been catalysts for the evolution of greater brain size and higher intelligence in Neanderthals?

18.13.4.2 Cognitive Competition between Early Humans and Neanderthals in Eurasia

Another intriguing question involves possible competition between early humans and Neanderthals. Some researchers suggest that competition from humans migrating to Europe and Asia from Africa eventually caused the extinction of the Neanderthals. However, another possible effect of such competition may have been beneficial for both species. Could competition between early humans and Neanderthals have been a factor stimulating evolution of higher intelligence in both species, with humans eventually winning out? This hypothesis is weakened by the fact that no fossils of Neanderthals have ever been found in Africa, but only in Europe, the Middle East, and parts of east Asia. Without such fossil finds of Neanderthals in Africa we must assume that no interactions in Africa between early Homo species and Neanderthals were possible, eliminating competition between the two species as a potential cause of increased brain size and intelligence in early humans in Africa. But what about potential competition between the two species in Eurasia? Was contact between the two species frequent and sustained enough to create significant cognitive competition between them which might have influenced their brain evolution?

It is now generally accepted by scientists that Neanderthals and modern humans coexisted for tens of thousands of years, Neanderthals having become extinct only some 40,000 years ago, while the origin of modern humans is thought to have taken place some 200,000 to 300,000 years ago and human migration out of Africa to Europe and Asia is thought to have occurred about 70,000 to 100,000 years ago. Furthermore, about 1-4% of human DNA is Neanderthal in origin indicating interbreeding, suggesting significant contact between Neanderthals and humans in several places outside of Africa and during several periods between 40,000 and 100,000 years ago (Alex, 2018). Anthropological evidence suggests that Neanderthals were cognitively advanced. For example, they made stone tools for specialized tasks. These tools, found in a cave in France, were made using a manufacturing technique (dubbed Mousterian by experts) characterized by preparing the core of raw stone material from which flakes were struck and then worked. Sharper tools with a finer edge can be produced using this technique. Neanderthals shaped these flakes into tools like scrapers, blades, and projectile points--specifically spear points. In fact, at Neanderthal caves sites in the Middle East, there are a higher percentage of spear points found than at neighboring Homo sapiens sites. Mousterian tools are a technological advance, taking a high degree of conceptualization and knowledge of the properties of the stone (see section on Material Culture in this textbook).

Were Neanderthals smart enough to provide cognitive competition for early humans? Human and Neanderthal DNA is 99.8% the same including genes important for brain expansion and language (Alex, 2018). As already mentioned, measurements of endocranial volume of Neanderthal skulls indicate that Neanderthals possessed a brain similar in size or even slightly larger than the brains of recent humans, with a larger visual cortex in Neanderthals, but a larger cerebellum in humans (Alex, 2018; Kochiyama, et al., 2018). But was any potential competition between Neanderthals and ancient humans sustained long enough to impact brain evolution? As discussed, we do know that the interactions between the two species were sufficient for some interbreeding between them, at least in Eurasia. Whether interactions between the two species led to cognitive competition between

the two species that was sufficient to affect brain evolution of one or both species is unknown, but remains an intriguing question. However, one thing we do know is that most of the encephalization of the two species was already complete before early humans left Africa and therefore before any interactions could have taken place between them.

Sometime between 520,000 and 630,000 years ago the human and Neanderthal lines diverged from a common ancestor (Homo heidelbergensis; see figure below) and took separate evolutionary paths. Those who migrated to Europe eventually evolved into Neanderthals, while those in Africa gave rise to modern humans. These early humans later migrated out of Africa to Europe and other regions of the Earth. By 150,000 years ago both species had similar average brain size of about 1400 cubic centimeters, three times larger than the brains of modern day chimpanzees, and they also had very similar within-species variation in cranial capacity. The average brain volume of Neanderthals of 1410 cubic centimeters actually exceeds the 1349 cubic centimeter average of modern humans (Alex, 2018). These events appear to have occurred before any interactions between the two species, suggesting that the bulk of brain evolution in both took place before any encounters between them, further weakening the mutual competition hypothesis discussed above as a significant cause of the encephalization that occurred in the evolution of early humans and Neanderthals after they diverged from Homo heidelbergensis.

What accounts for the similar large brain size in both Neanderthals and humans? Were the genetics for large brain size present in the common ancestor of both before the two lines diverged? Or, alternatively, did the similar encephalization in the two species result from convergent evolution? Common genetics appears to be at least one important factor. *Homo heidelbergensis* is thought to be the common ancestor of *Homo neanderthalensis* in Europe and *Homo sapiens* in Africa (see section on Human Evolution in Chapter 3). The trend toward larger brain size was already evident in Homo habilis and Homo erectus, predecessors of Homo heidelbergensis, suggesting that the increase in encephalization in Neanderthals and Homo sapiens was a continuation of the trend toward encephalization already present in the genus Homo. The figure below from section 3.7 shows these evolutionary relationships.

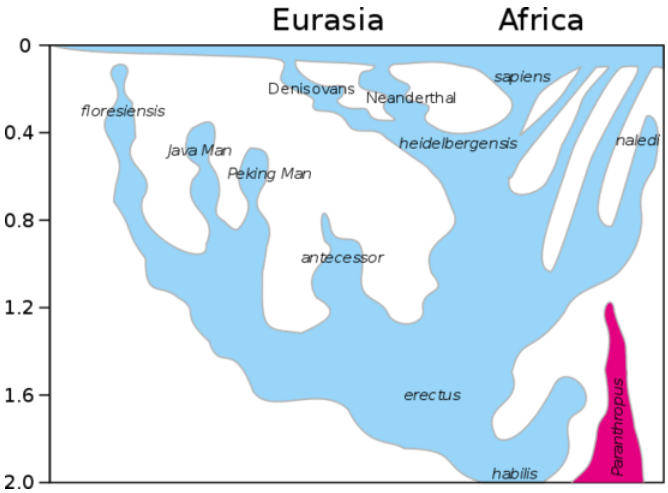


Figure **18.13.2**: Evolutionary relationships among hominins. Note the divergence of Neanderthal and sapiens (the human line) from heidelbergensis, which evolved from habilis and erectus. (Image from Wikimedia Commons; File:Homo lineage 2017update.svg; https://commons.wikimedia.org/wiki/F...2017update.svg; by User:Conquistador, User:Dbachmann; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

One hypothesis about the disappearance of the Neanderthals was that competition from early humans led to their extinction some 30,000 to 40,000 years ago (Gilpin, et al., 2016). However, as suggested above, is it possible that competition between Neanderthals and humans in Europe, south Asia, and the Middle East (where fossils of early humans and Neanderthals have been found) might have played some role in stimulating the evolution of larger, smarter brains in both species? In the figure below, note the larger skull of the Neanderthal compared to the skull of a modern human. Many of the species of the Homo genus coexisted with one another and could have interacted with one another. Perhaps competition between various Homo species played some role in the trend toward greater encephalization seen in the genus Homo beginning with species which diverged from Homo erectus. However, this hypothesis requires evidence of significant and sustained competition between these species, and that evidence has remained elusive. Perhaps future fossil finds will shed light on this intriguing possibility. On the other hand, small populations of

these species may have precluded significant overlap of their ranges and reduced the chances of significant interaction and sustained competition among them.

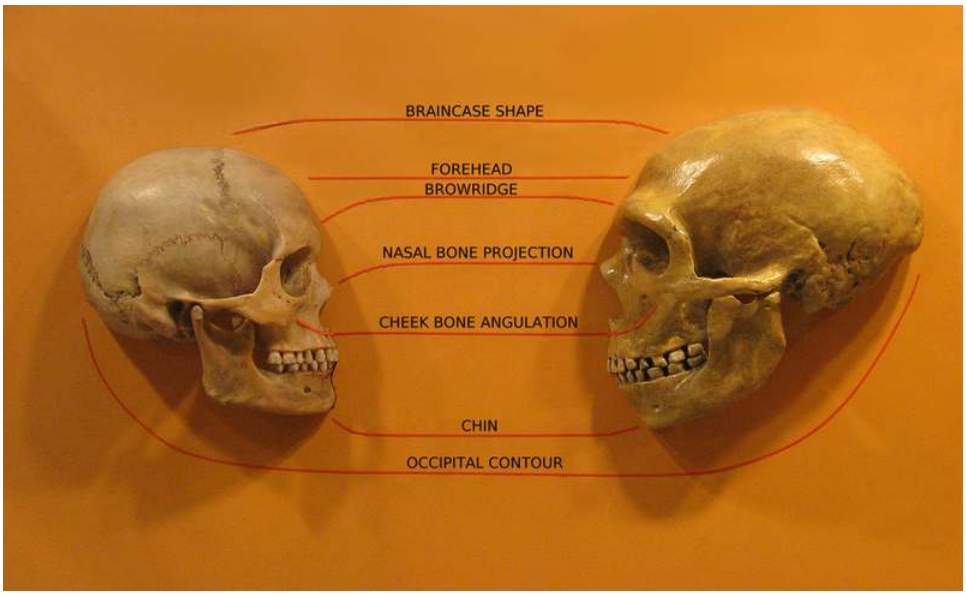


Figure 18.13.3: Comparison of Modern Human and Neanderthal skulls from the Cleveland Museum of Natural History. The cranial space in Neanderthal reveals a brain size slightly larger than the brain size of modern humans. This and other anthropological evidence suggests the possibility of significant competition between the two species, perhaps catalyzing the evolution of cognition in both. (Image and first sentence in caption from Wikimedia Commons, remainder of caption by Kenneth Koenigshofer, PhD, Chaffey College; File:Sapiens neanderthal comparison en.png; https://commons.wikimedia.org/wiki/F...parison_en.png; by hairymuseummatt (original photo), KaterBegemot (derivative work); licensed under the Creative Commons Attribution-Share Alike 2.0 Generic license).

18.13.5 An Integrative Model of the Evolution and Structure of Intelligence

A significant controversy among theorists is whether the human mind is **modular or general purpose**. The older model of the mind, originating with the British Empiricist philosophers of the 17th and 18th centuries and adopted by early behaviorists, sees the brain as a **general purpose processor** handling all kinds of adaptive problems without need for any specialized processing brain modules. An alternative model favored by many evolutionary psychologists such as Cosmides and Tooby at the University of California, Santa Barbara, is that the human mind is **highly modular** in nature, consisting of large numbers of specialized, domain-specific (specific to specific problem type) information processing modules which operate more or less independently of one another. Such modules or mini-computers are hypothesized to be specialized by evolution to operate on different types of information to achieve solutions to different types of adaptive problems. Although many evolutionary psychologists favor this modular view of the mind and brain, composed of special purpose brain modules, many psychologists disagree with this view, in part because it leaves out general intelligence and the flexibility in problem solving that seems to require general problem solving capabilities.

One solution that integrates both views is a model of the human brain that incorporates both types of processing mechanisms--some brain circuits specialized to solve specific adaptive problems such as mate selection, predator avoidance, and so on, and, in addition, brain mechanisms of general intelligence. Many evolutionary psychologists do not dispute the idea that some form of general intelligence exists; the problem is that they have been unable to understand how general intelligence, capable of solving noncurrent "novel problems," could have even evolved (Cosmides & Tooby, 2002; Kanasawa, 2004, 2010). This is because problems that are truly novel have not happened before and therefore have not been recurrent over many generations. This creates a theoretical problem because natural selection can only work on selection criteria consistently present generation after generation.

However, this problem has been solved. An analysis of the structure of events, including adaptive problems, reveals that every event contains two components: 1) abstract relational properties, such as cause-effect, similarity, and predictive relations, which

have been consistently present in the environment generation after generation, and 2) specific event details which are idiosyncratic, nonrecurrent, perhaps even truly novel, and therefore transient and confined to a single generation (Koenigshofer, 2017). Any adaptive problem is a mixture of these two elements. Although event details which are idiosyncratic to specific situations encountered by different individuals may be unique and novel, and thus highly variable across generations, nevertheless the relational properties such as cause-effect exist across situations and adaptive problems generation after generation.

Because of the repetition of such relational properties (causal, similarity, and predictive relations) over countless generations in countless events, adaptive problems, and adaptive opportunities, natural selection has utilized these relational regularities of the world as selection criteria for evolution of important components of brain organization and function. As a consequence, information about these relational properties of the world has been genetically incorporated into the innate organization of the brains of many species, including our own (Koenigshofer, 2017). This innate knowledge about these relational properties of the world forms a set of cognitive instincts that collectively make up what psychologists call general intelligence (Koenigshofer, 2017), the g-factor first discovered by Spearman in the early 1900s.

On this view, novel problems are only novel in their details, but constant or invariant across generations in their abstract relational structures. It is the constancy of these relational structures over generations that provided the regularities across generations which acted as selection criteria for the evolution of innate knowledge about these abstract, relational invariants--the key components of general intelligence. Application of this innate knowledge of general intelligence permits the solution to a near infinite variety of problems and adaptive opportunities, regular in their relational structure, but novel in their details (Koenigshofer, 2017). According to this integrative model of intelligence, many specialized modules (the modular view of intelligence) act along with general intelligence--the innate knowledge of causal, similarity, and predictive relations--to flexibly solve problems and seize adaptive opportunities, accounting for the unprecedented achievements of the human intellect. These cultural innovations by humans and their preservation over generations by cultural transmission have permitted humans each generation to draw upon the innovations and inventions of prior generations, building upon them, generation after generation, ratcheting up human adaptive success, far beyond that found in any other species.

18.13.6 References

Alex, B. (2018). Discover Magazine. Neanderthal Brains: Bigger, Not Necessarily Better; https://www.discovermagazine.com/pla...ssarily-better; retrieved 3/22/2022.

Barrett, L., & Henzi, P. (2005). The social nature of primate cognition. *Proceedings of the Royal Society B: Biological Sciences*, *272* (1575), 1865-1875.

Duveau, J., Berillon, G., Verna, C., Laisné, G., & Cliquet, D. (2019). The composition of a Neandertal social group revealed by the hominin footprints at Le Rozel (Normandy, France). *Proceedings of the National Academy of Sciences*, *116*(39), 19409-19414.

Fonseca-Azevedo, K., & Herculano-Houzel, S. (2012). Metabolic constraint imposes tradeoff between body size and number of brain neurons in human evolution. *Proceedings of the National Academy of Sciences*, *109* (45), 18571-18576.

Gibson, K. R. (2002). Evolution of human intelligence: The roles of brain size and mental construction. *Brain, behavior and evolution*, *59* (1-2), 10-20.

Gilpin, W., Feldman, M. W., & Aoki, K. (2016). An ecocultural model predicts Neanderthal extinction through competition with modern humans. *Proceedings of the National Academy of Sciences*, *113* (8), 2134-2139.

Hayden, B. (2012). Neandertal social structure?. *Oxford Journal of Archaeology*, *31* (1), 1-26.

Henry, A. G. (2017). Neanderthal cooking and the costs of fire. *Current Anthropology*, *58* (S16), S329-S336.

Herculano-Houzel, S. (2012). The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proceedings of the National Academy of Sciences*, *109* (Supplement 1), 10661-10668.

Herculano-Houzel, S., Catania, K., Manger, P. R., & Kaas, J. H. (2015). Mammalian brains are made of these: a dataset of the numbers and densities of neuronal and nonneuronal cells in the brain of glires, primates, scandentia, eulipotyphlans, afrotherians and artiodactyls, and their relationship with body mass. *Brain, Behavior and Evolution*, *86*(3-4), 145-163.

Kochiyama, T., Ogihara, N., Tanabe, H.C. *et al.* (2018). Reconstructing the Neanderthal brain using computational anatomy. *Sci Rep* **8,** 6296. https://doi.org/10.1038/s41598-018-24331-0

Lakatos, L., & Janka, Z. (2008). Evolution of human brain and intelligence. *Ideggyogyaszati szemle*, *61* (7-8), 220-229.

Osvath, M., & Gärdenfors, P. (2005). Oldowan culture and the evolution of anticipatory cognition. *Lund University Cognitive Studies*, *122*, 1-16.

Reader, S. M., Hager, Y., & Laland, K. N. (2011). The evolution of primate general and cultural intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1017-1027.

Roth, G. (2015). Convergent evolution of complex brains and high intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370* (1684), 20150049.

Roth, G., & Dicke, U. (2012). Evolution of the brain and intelligence in primates. *Progress in brain research*, *195*, 413-430.Suddendorf, T., Addis, D. R., & Corballis, M. C. (2011). Mental time travel and shaping of the human mind. *M. Bar*, 344-354.

Schenker, N. M., Desgouttes, A. M. & Semendeferi, K. (2005) Neural connectivity and cortical substrates of cognition in hominoids. Journal of Human Evolution 49(5):547–69

Schoenemann, P. T., Sheehan, M. J. & Glotzer, L. D. (2005) Prefrontal white matter volume is disproportionately larger in humans than in other primates. Nature Neuroscience 8(2):242–52.CrossRefGoogle ScholarPubMed

van Horik, J., & Emery, N. J. (2011). Evolution of cognition. Wiley Interdisciplinary Reviews: Cognitive Science, *2(6), 621-633.*

Williams, A. E. (2020). Human intelligence and general collective intelligence as phase changes in animal intelligence.

18.13.7 Attributions

"Evolution of Cognition" by Kenneth A. Koenigshofer, Ph.D., Chaffey College, licensed under CC BY 4.0.

This page titled 18.13: Chapter 14- The Brain and Evolution of Cognition is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

18.14: Chapter 14- Mirror Neurons, Theory of Mind, Social Cognition, and Neuroscience of Its Disorders

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Describe mirror and canonical neurons, their locations in the brain, and their possible functions
2. Describe theory of mind (ToM) and its role in social cognition
3. Define the social brain

18.14.0.1 Overview

In the early 1990s, researchers working with monkeys identified a group of neurons in their frontal cortex that were activated when the monkeys made a particular motion with their hand or mouth. There was nothing surprising in that, since these neurons were located in what is considered a “motor” portion of the brain. What was surprising, though, was that this same group of neurons was also activated in monkeys who were not performing the motion themselves, but rather watching other monkeys perform it. Hence these neurons were given the name “mirror neurons”. Mirror neurons are one possible basis in the brain for what may be called social intelligence or social cognition, information processing necessary for successful interactions with conspecifics. The new field of social neuroscience studies the neurological structures and processes that carry out the information processing required for adaptive interactions with others, basis for all human social behavior. One of these capacities, labeled “theory of mind,” is ability to perceive and understand mental states of others, essential in human social life. Autism is in the category of *pervasive developmental disorders* and may result in part from impaired theory of mind and abnormalities in a number of brain structures collectively referred to by some neuroscientists as the "social brain."

18.14.0.1 Are Mirror Neurons the Basis for Communication?

Mirror neurons were discovered in area F5 of the ventral premotor cortex of macaque monkeys by researchers at the University of Parma, Italy, in 1992. The researchers found that these neurons had some very distinctive characteristics : they fired not only when a monkey performed a voluntary gesture (for example, turned a handle to open a door), but also when a monkey watched another monkey perform this same action. The part of the brain where these mirror neurons are found in monkeys corresponds to the part of the human brain known as Broca’s area, which since the 19th century has been known to play an important role in language. In addition to their being located in a brain area associated with language in humans, two other things about mirror neurons have led many researchers to suggest that they may play a role in the evolution and learning of language: these neurons tell us about the intentions of the people around us, and they help us to imitate the movements of other people’s lips and tongues.

Another category of neurons present in area F5, **canonical neurons**, may, like mirror neurons, be involved in human language faculties. The special characteristic of canonical neurons is that they fire when an individual simply sees a graspable object. For example, if a monkey looks at a ball, the canonical neurons that fire are the same ones that will fire if the monkey decides to actually grasp the ball. In contrast, the monkey’s mirror neurons will not be activated by the mere sight of a ball, but only if the monkey either grasps the ball or sees another monkey do so.

In essence, mirror neurons react to visual stimuli that represent an interaction between a biological means of action (such as the hand or mouth) and an object. These neurons thus act as agents for recognizing purposive actions as opposed to simple movements.

Even in the case of canonical neurons, which can be activated by the sight of a graspable object in the absence of movement, the internal representation is that of a purposive action, and not just a simple movement of the hand or arm.

This is what has led some researchers to think that mirror neurons might help to explain the cognitive foundations of language, by providing the neural substrate for the human ability to understand the meaning of other people’s actions, which is the basis for all social relations. This system of correspondences between perceptions and actions would help us to infer other people’s mental

states and to interpret their actions as intentional behaviors arising from these states. We can then easily imagine how this mechanism for interpreting gestural communication might have been applied to verbal communication as well.

The hypothesis advanced is that the motor system, through its mirror neurons, is involved in perceiving speech, and that through evolution, the “motor resonance” generated by the mirror neurons has been diverted (or exapted) from its original function to serve the needs of language. One has to be impressed by the economy of such a cognitive system, in which one individual understands what other individuals are doing (or saying) on the basis of the internal neural representation of his or her own motor capabilities.

The point is that intentional communication between two individuals differs from the simple cries of alarm by which animals signal danger to all members of their group indiscriminately. Intentional communication, in contrast, requires one individual who is transmitting information and a second who is paying attention to receive it. Among all the possible origins of language , the first form of intentional communication among humans may have arisen from the imitation of gestures and facial expressions. Thus mirror neurons may have played a role in sharing these common representations and, eventually, a common language.

Mirror neurons have been shown to have some other interesting characteristics as well. First of all, as noted above, they will be activated when you see someone else’s hand grasp an object, but not when you see a tool grasp the same object. The explanation is that unlike tools and other human artifacts, human body parts are represented in the motor and premotor areas of the brain’s frontal lobes. Second—and here is where the implications for language become really interesting—mirror neurons do not react to just any movements of the hand or mouth, but only to movements that are involved in goal-directed actions.

> In other words, it is only when an action has a meaning that it activates the mirror neurons. Their response is thus associated with the expression of intentionality—that is, with the purpose of the observed gesture. For example, certain mirror neurons that are activated when a monkey manipulates an object will remain quiet when the monkey uses the same muscles, connected to the same neurons, to perform a similar action for a different purpose, such as to scratch itself or to pick an insect out of its fur.

It is becoming more and more apparent that the brain’s motor system does not only control movements but can also in a certain sense read the actions performed by other individuals. Mirror neurons may thus play a fundamental role in all human social behavior, including language.

18.14.0.1 Summary

Mirror neurons discovered in area F5 of the frontal cortex in the macaque monkey, an area that corresponds to Broca's area of the left frontal lobe of humans, fire action potentials when a monkey performs a voluntary movement (for example, turning a handle to open a door), but also when a monkey watches another monkey perform this same action. if a monkey looks at a ball, the canonical neurons that fire are the same ones that will fire if the monkey decides to actually grasp the ball. However, by contrast, the monkey’s mirror neurons will not be activated by the mere sight of a ball, but only if the monkey either grasps the ball or sees another monkey do so. Mirror neurons do not react to just any movements of the hand or mouth, but only to movements that are involved in goal-directed actions. These neurons thus act as agents for recognizing purposive actions as opposed to simple movements. The hypothesis is that the motor system, through its mirror neurons, is involved in perceiving speech, and the intentions of others, and may thus play a fundamental role in all human social behavior.

18.14.0.1 Attributions

Are Mirror Neurons the Basis for Communication? by Bruno Dubuc, The Brain from Top to Bottom under a Copyleft license.

18.14.0.1 The Social Brain and Social Neuroscience

The brain mechanisms of social cognition and social intelligence help us understand other people and accomplish adaptive interactions with them. Deficits in these abilities due to damage in the "social brain" can lead to disorders such as autism.

18.14.1 **Social Neuroscience**

By Tiffany A. Ito and Jennifer T. Kubota

University of Colorado Boulder, University of Delaware

This module provides an overview of the new field of social neuroscience, which combines the use of neuroscience methods and theories to understand how other people influence our thoughts, feelings, and behavior. The module reviews research measuring

neural and hormonal responses to understand how we make judgments about other people and react to stress. Through these examples, it illustrates how social neuroscience addresses three different questions: (1) how our understanding of social behavior can be expanded when we consider neural and physiological responses, (2) what the actual biological systems are that implement social behavior (e.g., what specific brain areas are associated with specific social tasks), and (3) how biological systems are impacted by social processes.

- Electroencephalography (EEG)
- fMRI
- mentalizing
- neuroendocrinology
- Prejudice
- Social neuroscience
- social psychology
- stereotyping
- stress

18.14.1.1 Learning Objectives

- Define social neuroscience and describe its three major goals.
- Describe how measures of brain activity such as EEG and fMRI are used to make inferences about social processes.
- Discuss how social categorization occurs.
- Describe how simulation may be used to make inferences about others.
- Discuss the ways in which other people can cause stress and also protect us against stress.

Psychology has a long tradition of trying to better understand how we think and act. For example, in 1939 Heinrich Kluver and Paul Bucy removed (i.e. lesioned) the temporal lobes in some rhesus monkeys and observed the effect on behavior. Included in these **lesions** was a subcortical area of the brain called the **amygdala**. After surgery, the monkeys experienced profound behavioral changes, including loss of fear. These results provided initial evidence that the amygdala plays a role in emotional responses, a finding that has since been confirmed by subsequent studies (Phelps & LeDoux, 2005; Whalen & Phelps, 2009).

18.14.2 What Is Social Neuroscience?

Social neuroscience seeks to understand how we think about and act toward other people. More specifically, we can think of social neuroscience as an interdisciplinary field that uses a range of neuroscience measures to understand social behavior. As such, social neuroscience studies the same topics as social psychology, but does so from a multilevel perspective that includes the study of the brain and body. Figure 1 shows the scope of social neuroscience with respect to the older fields of social psychology and neuroscience. Although the field is relatively new – the term first appeared in 1992 (Cacioppo & Berntson, 1992) – it has grown rapidly, thanks to technological advances in brain science, and to the recognition that neural and physiological information are critical to understanding how we interact with other people.

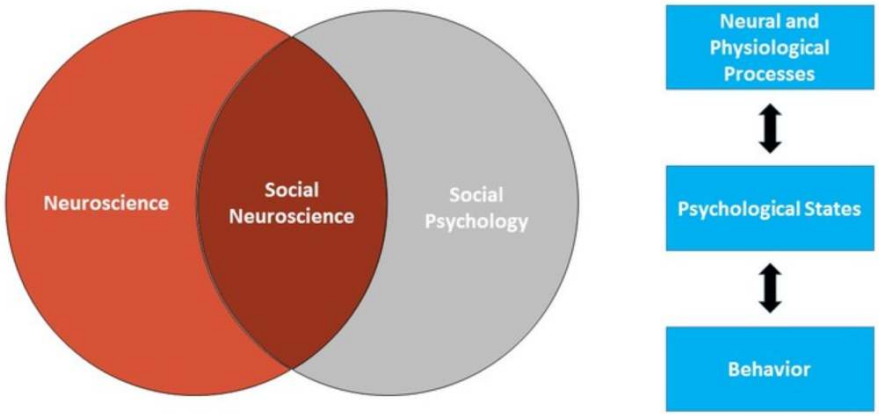


Figure 18.14.1: Social neuroscience is the intersection of social psychology and neuroscience. Under this multilevel approach, neural/physiological processes and behavior are two things we can measure or observe. Psychological states cannot be directly observed, but understanding them is the goal. Social neuroscientists use the observable neural/physiological processes and behavioral responses to make inferences about unobservable psychological states. The bidirectional arrows show that all levels of analysis are assumed to influence each other (e.g., psychological states can influence neural responses, and neural responses can influence psychological states).

Social neuroscience can be thought of as both a methodological approach (using measures of the brain and body to study social processes) and a theoretical orientation (seeing the benefits of integrating neuroscience into the study of social psychology). The overall approach in social neuroscience is to understand the psychological processes that underlie our social behavior. Because those psychological processes are intrapsychic phenomena that cannot be directly observed, social neuroscientists rely on a combination of measurable or observable neural and physiological responses as well as actual overt behavior to make inferences about psychological states (see Figure 1). Using this approach, social neuroscientists have been able to pursue three different types of questions: (1) What more can we learn about social behavior when we consider neural and physiological responses? (2) What are the actual biological systems that implement social behavior (e.g., what specific brain areas are associated with specific social tasks)? and (3) How are biological systems impacted by social processes?

How Automatically Do We Judge Other People?

Social categorization is the act of mentally classifying someone as belonging in a group. Why do we do this? It is an effective mental shortcut. Rather than effortfully thinking about every detail of every person we encounter, social categorization allows us to rely on information we already know about the person's group. For example, by classifying your restaurant server as a man, you can quickly activate all the information you have stored about men and use it to guide your behavior. But this shortcut comes with potentially high costs. The stored group beliefs might not be very accurate, and even when they do accurately describe some group members, they are unlikely to be true for every member you encounter. In addition, many beliefs we associate with groups – called **stereotypes** – are negative. This means that relying on social categorization can often lead people to make negative assumptions about others.

The potential costs of social categorization make it important to understand how social categorization occurs. Is it rare or does it occur often? Is it something we can easily stop, or is it hard to override? One difficulty answering these questions is that people are not always consciously aware of what they are doing. In this case, we might not always realize when we are categorizing someone. Another concern is that even when people are aware of their behavior, they can be reluctant to accurately report it to an experimenter. In the case of social categorization, subjects might worry they will look bad if they accurately report classifying someone into a group associated with negative stereotypes. For instance, many racial groups are associated with some negative stereotypes, and subjects may worry that admitting to classifying someone into one of those groups means they believe and use those negative stereotypes.



Figure 18.14.2: This man is wearing an elastic electrode cap into which individual electrodes (inside the white circles) are sewn into standardized locations. [Image: Hans, CC0 Public Domain, https://goo.gl/m25gce]

Social neuroscience has been useful for studying how social categorization occurs without having to rely on self-report measures, instead measuring brain activity differences that occur when people encounter members of different social groups. Much of this work has been recorded using the **electroencephalogram**, or **EEG**. EEG is a measure of electrical activity generated by the brain's neurons. Comparing this electrical activity at a given point in time against what a person is thinking and doing at that same time allows us to make inferences about brain activity associated with specific psychological states. One particularly nice feature of EEG is that it provides very precise timing information about when brain activity occurs. EEG is measured non-invasively with small electrodes that rest on the surface of the scalp. This is often done with a stretchy elastic cap, like the one shown in Figure 2, into which the small electrodes are sewn. Researchers simply pull the cap onto the subject's head to get the electrodes into place; wearing it is similar to wearing a swim cap. The subject can then be asked to think about different topics or engage in different tasks as brain activity is measured.

To study social categorization, subjects have been shown pictures of people who belong to different social groups. Brain activity recorded from many individual trials (e.g., looking at lots of different Black individuals) is then averaged together to get an overall idea of how the brain responds when viewing individuals who belong to a particular social group. These studies suggest that social categorization is an **automatic process** – something that happens with little conscious awareness or control – especially for dimensions like gender, race, and age (Ito & Urland, 2003; Mouchetant-Rostaing & Giard, 2003). The studies specifically show that brain activity differs when subjects view members of different social groups (e.g., men versus women, Blacks versus Whites), suggesting that the group differences are being encoded and processed by the perceiver. One interesting finding is that these brain changes occur both when subjects are purposely asked to categorize the people into social groups (e.g., to judge whether the person is Black or White), and also when they are asked to do something that draws attention away from group classifications (e.g., making a personality judgment about the person) (Ito & Urland, 2005). This tells us that we do not have to intend to make group classifications in order for them to happen. It is also very interesting to consider how quickly the changes in brain responses occur. Brain activity is altered by viewing members of different groups within 200 milliseconds of seeing a person's face. That is just two-

tenths of a second. Such a fast response lends further support to the idea that social categorization occurs automatically and may not depend on conscious intention.

Overall, this research suggests that we engage in social categorization very frequently. In fact, it appears to happen automatically (i.e., without us consciously intending for it to happen) in most situations for dimensions like gender, age, and race. Since classifying someone into a group is the first step to activating a group stereotype, this research provides important information about how easily stereotypes can be activated. And because it is hard for people to accurately report on things that happen so quickly, this issue has been difficult to study using more traditional self-report measures. Using EEGs has, therefore, been helpful in providing interesting new insights into social behavior.

18.14.3 Do We Use Our Own Behavior to Help Us Understand Others?

Classifying someone into a social group then activating the associated stereotype is one way to make inferences about others. However, it is not the only method. Another strategy is to imagine what our own thoughts, feelings, and behaviors would be in a similar situation. Then we can use our simulated reaction as a best guess about how someone else will respond (Goldman, 2005). After all, we are experts in our own feelings, thoughts, and tendencies. It might be hard to know what other people are feeling and thinking, but we can always ask ourselves how we would feel and act if we were in their shoes.

There has been some debate about whether **simulation** is used to get into the minds of others (Carruthers & Smith, 1996; Gallese & Goldman, 1998). Social neuroscience research has addressed this question by looking at the brain areas used when people think about themselves and others. If the same brain areas are active for the two types of judgments, it lends support to the idea that the self may be used to make inferences about others via simulation.

We know that an area in the prefrontal cortex called the **medial prefrontal cortex** (mPFC) – located in the middle of the frontal lobe – is active when people think about themselves (Kelley, Macrae, Wyland, Caglar, Inati, & Heatherton, 2002). This conclusion comes from studies using **functional magnetic resonance imaging**, or fMRI. While EEG measures the brain's electrical activity, fMRI measures changes in the oxygenation of blood flowing in the brain. When neurons become more active, blood flow to the area increases to bring more oxygen and glucose to the active cells. fMRI allows us to image these changes in oxygenation by placing people in an fMRI machine or scanner (Figure 3), which consists of large magnets that create strong magnetic fields. The magnets affect the alignment of the oxygen molecules within the blood (i.e., how they are tilted). As the oxygen molecules move in and out of alignment with the magnetic fields, their nuclei produce energy that can be detected with special sensors placed close to the head. Recording fMRI involves having the subject lay on a small bed that is then rolled into the scanner. While fMRI does require subjects to lie still within the small scanner and the large magnets involved are noisy, the scanning itself is safe and painless. Like EEG, the subject can then be asked to think about different topics or engage in different tasks as brain activity is measured. If we know what a person is thinking or doing when fMRI detects a blood flow increase to a particular brain area, we can infer that part of the brain is involved with the thought or action. fMRI is particularly useful for identifying which particular brain areas are active at a given point in time.



Figure **18.14.3**: Functional magnetic resonance imaging (fMRI) scanner used to image the brain while people perform tasks. The scanner allows researchers to view the changes in blood oxygenation in specific locations in the brain during a task. Images are collected using powerful magnets and radio waves that shift the position of atoms in oxygenated blood that rushes to areas involved in performing the task. MRIs are non-invasive and there are no known risks from exposure to the magnetic fields or radio waves. Subjects lay on the while bed with their head inside the head coil, then the bed is moved into the scanner. [Image: Janne Moren, https://goo.gl/MKb2jn, CC BY-NC-SA 2.0, https://goo.gl/Toc0ZF]

The conclusion that the mPFC is associated with the self comes from studies measuring fMRI while subjects think about themselves (e.g., saying whether traits are descriptive of themselves). Using this knowledge, other researchers have looked at whether the same brain area is active when people make inferences about others. Mitchell, Neil Macrae, and Banaji (2005) showed subjects pictures of strangers and had them judge either how pleased the person was to have his or her picture taken or how symmetrical the face appeared. Judging whether someone is pleased about being photographed requires making an inference about someone's internal feelings – we call this **mentalizing**. By contrast, facial symmetry judgments are based solely on physical appearances and do not involve mentalizing. A comparison of brain activity during the two types of judgments shows more activity in the mPFC when making the mental versus physical judgments, suggesting this brain area is involved when inferring the internal beliefs of others.

There are two other notable aspects of this study. First, mentalizing about others also increased activity in a variety of regions important for many aspects of social processing, including a region important in representing biological motion (**superior temporal sulcus** or STS), an area critical for emotional processing (amygdala), and a region also involved in thinking about the beliefs of others (**temporal parietal junction**, TPJ) (Gobbini & Haxby, 2007; Schultz, Imamizu, Kawato, & Frith, 2004) (Figure 4). This finding shows that a distributed and interacting set of brain areas is likely to be involved in social processing. Second, activity in the most ventral part of the mPFC (the part closer to the belly rather than toward the top of the head), which has been most consistently associated with thinking about the self, was particularly active when subjects mentalized about people they rated as similar to themselves. Simulation is thought to be most likely for similar others, so this finding lends support to the conclusion that we use simulation to mentalize about others. After all, if you encounter someone who has the same musical taste as you, you will probably assume you have other things in common with him. By contrast, if you learn that someone loves music that you hate, you might expect him to differ from you in other ways (Srivastava, Guglielmo, & Beer, 2010). Using a simulation of our own feelings and thoughts will be most accurate if we have reason to think the person's internal experiences are like our own. Thus, we may be most likely to use simulation to make inferences about others if we think they are similar to us.

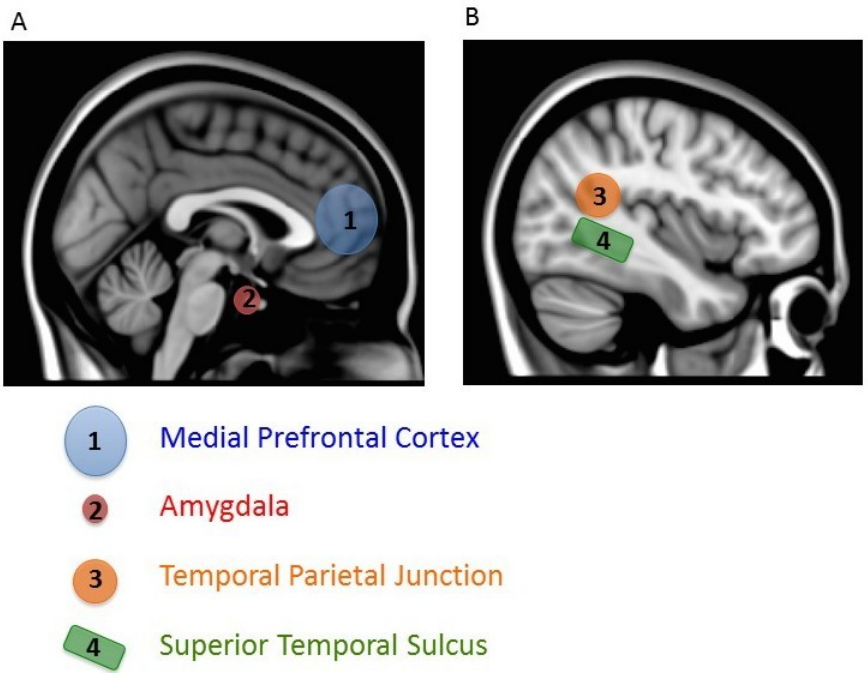


Figure **18.14.4**: The areas of the brain most commonly associated with processing of self and others. Panel A is a sagittal view, looking at the inside of the brain as if it were sliced in half. Panel B is a lateral view, showing the brain from the outside. The medial prefrontal cortex (mPFC) (Structure 1) is commonly activated when thinking about one's self and when thinking about similar others. Additionally, the amygdala (Structure 2) is important for learning about and detecting important things in our environments and plays an important role in fear learning and expression. The temporal parietal junction (TPJ) (Structure 3), located at the intersection of the parietal and temporal lobes, is activated when people think about the beliefs of others. Activity in the superior temporal sulcus (STS) (Structure 4) is commonly observed when people view biological motion. The common activation of this network of regions when people think about the feelings, thoughts, and intentions of others indicates that the processing of others involves a number of complex psychological processes.

This research is a good example of how social neuroscience is revealing the **functional neuroanatomy** of social behavior. That is, it tells us which brain areas are involved with social behavior. The mPFC (as well as other areas such as the STS, amygdala, and TPJ) is involved in making judgments about the self and others. This research also provides new information about how inferences are made about others. Whereas some have doubted the widespread use of simulation as a means for making inferences about others, the activation of the mPFC when mentalizing about others, and the sensitivity of this activation to similarity between self and other, provides evidence that simulation occurs.

18.14.4 What Is the Cost of Social Stress?

Stress is an unfortunately frequent experience for many of us. **Stress** – which can be broadly defined as a threat or challenge to our well-being – can result from everyday events like a course exam or more extreme events such as experiencing a natural disaster. When faced with a stressor, **sympathetic nervous system** activity increases in order to prepare our body to respond to the challenge. This produces what Selye (1950) called a **fight or flight response**. The release of **hormones**, which act as messengers from one part of an organism (e.g., a cell or gland) to another part of the organism, is part of the stress response.

A small amount of stress can actually help us stay alert and active. In comparison, sustained stressors, or chronic stress, detrimentally affect our health and impair performance (Al'Absi, Hugdahl, & Lovallo, 2002; Black, 2002; Lazarus, 1974). This happens in part through the chronic secretion of stress-related hormones (e.g., Davidson, Pizzagalli, Nitschke, & Putnam, 2002; Dickerson, Gable, Irwin, Aziz, & Kemeny, 2009). In particular, stress activates the **hypothalamic-pituitary-adrenal (HPA) axis** to release **cortisol** (see Figure 5 for a discussion). Chronic stress, by way of increases in cortisol, impairs attention, memory, and self-control (Arnsten, 2009). Cortisol levels can be measured non-invasively in bodily fluids, including blood and saliva. Researchers often collect a cortisol sample before and after a potentially stressful task. In one common collection method, subjects place polymer swabs under their tongue for 1 to 2 minutes to soak up saliva. The saliva samples are then stored and analyzed later to determine the level of cortisol present at each time point.

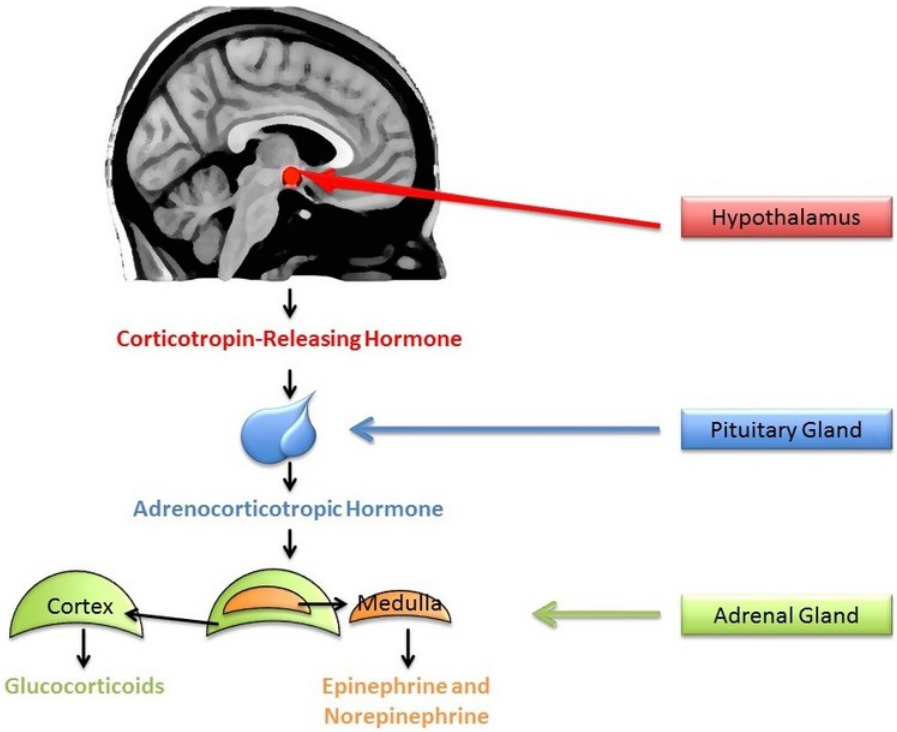


Figure **18.14.5**: The Hypothalamic-pituitary-adrenal (HPA) axis. Black arrows represent the stress response pathway starting in the brain at the hypothalamus (an area within the brain). Stress triggers neurons in the hypothalamus to release corticotrophin-releasing hormone (CRH). The CRH is transported to the pituitary gland, another area in the brain, that activates the secretion of andrenocorticotropic hormone (ACTH). In turn, ACTH stimulates the adrenal glands that sit on top of the kidneys. The adrenal glands are composed of the outer adrenal cortex and inner adrenal medulla. The adrenal cortex secretes glucorcorticoids (including cortisol) and the medulla secretes epinephrine and norepinephrine. Stress, both psychological and physical, activates the HPA axis and results in the systemic release of cortisol, epinephrine, and norepinephrine.

Whereas early stress researchers studied the effects of physical stressors like loud noises, social neuroscientists have been instrumental in studying how our interactions with other people can cause stress. This question has been addressed through **neuroendocrinology**, or the study of how the brain and hormones act in concert to coordinate the physiology of the body. One contribution of this work has been in understanding the conditions under which other people can cause stress. In one study, Dickerson, Mycek, and Zaldivar (2008) asked undergraduates to deliver a speech either alone or to two other people. When the students gave the speech in front of others, there was a marked increase in cortisol compared with when they were asked to give a speech alone. This suggests that like chronic physical stress, everyday social stressors, like having your performance judged by others, induces a stress response. Interestingly, simply giving a speech in the same room with someone who is doing something else did not induce a stress response. This suggests that the mere presence of others is not stressful, but rather it is the potential for them to judge us that induces stress.

Worrying about what other people think of us is not the only source of social stress in our lives. Other research has shown that interacting with people who belong to different social groups than us – what social psychologists call **outgroup** members – can increase physiological stress responses. For example, cardiovascular responses associated with stress like contractility of the heart ventricles and the amount of blood pumped by the heart (what is called cardiac output) are increased when interacting with outgroup as compared with **ingroup** members (i.e., people who belong to the same social group we do) (Mendes, Blascovich, Likel, & Hunter, 2002). This stress may derive from the expectation that interactions with dissimilar others will be uncomfortable (Stephan & Stephan, 1985) or concern about being judged as unfriendly and prejudiced if the interaction goes poorly (Plant & Devine, 2003).

The research just reviewed shows that events in our social lives can be stressful, but are social interactions always bad for us? No. In fact, while others can be the source of much stress, they are also a major buffer against stress. Research on **social support** shows that relying on a network of individuals in tough times gives us tools for dealing with stress and can ward off loneliness (Cacioppo & Patrick, 2008). For instance, people who report greater social support show a smaller increase in cortisol when performing a speech in front of two evaluators (Eisenberger, Taylor, Gable, Hilmert, & Lieberman, 2007).

What determines whether others will increase or decrease stress? What matters is the context of the social interaction. When it has potential to reflect badly on the self, social interaction can be stressful, but when it provides support and comfort, social interaction can protect us from the negative effects of stress. Using neuroendocrinology by measuring hormonal changes in the body has helped researchers better understand how social factors impact our body and ultimately our health.

18.14.5 Conclusions

Human beings are intensely social creatures – our lives are intertwined with other people and our health and well-being depend on others. Social neuroscience helps us to understand the critical function of how we make sense of and interact with other people. This module provides an introduction to what social neuroscience is and what we have already learned from it, but there is much still to understand. As we move forward, one exciting future direction will be to better understand how different parts of the brain and body interact to produce the numerous and complex patterns of social behavior that humans display. We hinted at some of this complexity when we reviewed research showing that while the mPFC is involved in mentalizing, other areas such as the STS, amygdala, and TPJ are as well. There are likely additional brain areas involved as well, interacting in ways we do not yet fully understand. These brain areas in turn control other aspects of the body to coordinate our responses during social interactions. Social neuroscience will continue to investigate these questions, revealing new information about how social processes occur, while also increasing our understanding of basic neural and physiological processes.

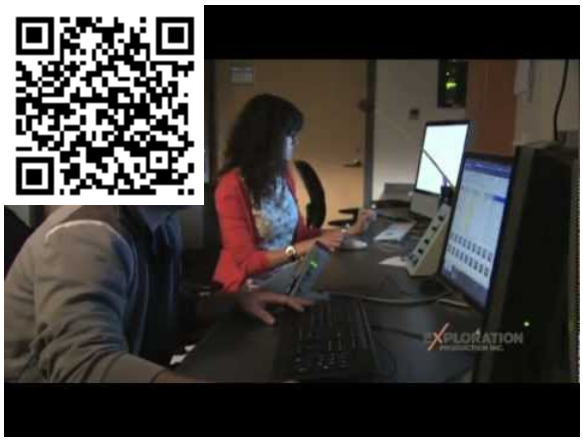
18.14.5.1 Outside Resources

Society for Social Neuroscience

http://www.s4sn.org

Video: See a demonstration of fMRI data being collected.





Video: See an example of EEG data being collected.





Video: View two tasks frequently used in the lab to create stress – giving a speech in front of strangers, and doing math computations out loud in front of others. Notice how some subjects show obvious signs of stress, but in some situations, cortisol changes suggest that even people who appear calm are experiencing a physiological response associated with stress.





Video: Watch a video used by Fritz Heider and Marianne Simmel in a landmark study on social perception published in 1944. Their goal was to investigate how we perceive other people, and they studied it by seeing how readily we apply people-like interpretations to non-social stimuli.





18.14.5.2 Discussion Questions

1. Categorizing someone as a member of a social group can activate group stereotypes. EEG research suggests that social categorization occurs quickly and often automatically. What does this tell us about the likelihood of stereotyping occurring? How can we use this information to develop ways to stop stereotyping from happening?
2. Watch this video, similar to what was used by Fritz Heider and Marianne Simmel in a landmark study on social perception published in 1944, and imagine telling a friend what happened in the video. http://intentionperception.org/wp-co...ider_Flash.swf. After watching the video, think about the following: Did you describe the motion of the objects solely in geometric terms (e.g., a large triangle moved from the left to the right), or did you describe the movements as actions of animate beings, maybe even of people (e.g., the circle goes into the house and shuts the door)? In the original research, 33 of 34 subjects described the action of the shapes using human terms. What does this tell us about our tendency to mentalize?
3. Consider the types of things you find stressful. How many of them are social in nature (e.g., are related to your interactions with other people)? Why do you think our social relations have such potential for stress? In what ways can social relations be beneficial and serve as a buffer for stress?

18.14.5.3 Vocabulary

Amygdala

A region located deep within the brain in the medial area (toward the center) of the temporal lobes (parallel to the ears). If you could draw a line through your eye sloping toward the back of your head and another line between your two ears, the amygdala would be located at the intersection of these lines. The amygdala is involved in detecting relevant stimuli in our environment and has been implicated in emotional responses.

Automatic process

When a thought, feeling, or behavior occurs with little or no mental effort. Typically, automatic processes are described as involuntary or spontaneous, often resulting from a great deal of practice or repetition.

Cortisol

A hormone made by the adrenal glands, within the cortex. Cortisol helps the body maintain blood pressure and immune function. Cortisol increases when the body is under stress.

Electroencephalogram

A measure of electrical activity generated by the brain's neurons.

Fight or flight response

The physiological response that occurs in response to a perceived threat, preparing the body for actions needed to deal with the threat.

Functional magnetic resonance imaging

A measure of changes in the oxygenation of blood flow as areas in the brain become active.

Functional neuroanatomy

Classifying how regions within the nervous system relate to psychology and behavior.

Hormones

Chemicals released by cells in the brain or body that affect cells in other parts of the brain or body.

Hypothalamic-pituitary-adrenal (HPA) axis

A system that involves the hypothalamus (within the brain), the pituitary gland (within the brain), and the adrenal glands (at the top of the kidneys). This system helps maintain homeostasis (keeping the body's systems within normal ranges) by regulating digestion, immune function, mood, temperature, and energy use. Through this, the HPA regulates the body's response to stress and injury.

Ingroup

A social group to which an individual identifies or belongs.

Lesions

Damage or tissue abnormality due, for example, to an injury, surgery, or a vascular problem.

Medial prefrontal cortex

An area of the brain located in the middle of the frontal lobes (at the front of the head), active when people mentalize about the self and others.

Mentalizing

The act of representing the mental states of oneself and others. Mentalizing allows humans to interpret the intentions, beliefs, and emotional states of others.

Neuroendocrinology

The study of how the brain and hormones act in concert to coordinate the physiology of the body.

Outgroup

A social group to which an individual does not identify or belong.

Simulation

Imaginary or real imitation of other people's behavior or feelings.

Social categorization

The act of mentally classifying someone into a social group (e.g., as female, elderly, a librarian).

Social support

A subjective feeling of psychological or physical comfort provided by family, friends, and others.

Stereotypes

The beliefs or attributes we associate with a specific social group. Stereotyping refers to the act of assuming that because someone is a member of a particular group, he or she possesses the group's attributes. For example, stereotyping occurs when we assume someone is unemotional just because he is man, or particularly athletic just because she is African American.

Stress

A threat or challenge to our well-being. Stress can have both a psychological component, which consists of our subjective thoughts and feelings about being threatened or challenged, as well as a physiological component, which consists of our body's response to the threat or challenge (see "fight or flight response").

Superior temporal sulcus

The sulcus (a fissure in the surface of the brain) that separates the superior temporal gyrus from the middle temporal gyrus. Located in the temporal lobes (parallel to the ears), it is involved in perception of biological motion or the movement of animate

objects.

Sympathetic nervous system

A branch of the autonomic nervous system that controls many of the body's internal organs. Activity of the SNS generally mobilizes the body's fight or flight response.

Temporal parietal junction

The area where the temporal lobes (parallel to the ears) and parieta lobes (at the top of the head toward the back) meet. This area is important in mentalizing and distinguishing between the self and others.

18.14.5.4 References

Al'Absi, M., Hugdahl, K., & Lovallo, W. (2002). Adrenocortical stress responses and altered working memory performance. *Psychophysiology, 39*(1), 95–99.

Arnsten, A. F. T. (2009). Stress signaling pathways that impair prefrontal cortex structure and function. *Nature Neuroscience Reviews, 10*(6), 410–422.

Black, P. (2002). Stress and the inflammatory response: A review of neurogenic inflammation. *Brain, Behavior, & Immunity, 16*, 622–653.

Cacioppo, J. T., & Berntson, G. G. (1992). Social psychological contributions to the decade of the brain: Doctrine of multilevel analysis. *American Psychologist, 47*, 1019–1028.

Cacioppo, J. T., & Patrick, B. (2008). *Loneliness: Human nature and the need for social connection*. New York, NY: W. W. Norton & Company.

Carruthers, P. and Smith, P. (1996). *Theories of Theories of Mind*. New York, NY: Cambridge University Press.

Davidson, R. J., Pizzagalli, D., Nitschke, J. B., & Putnam, K. (2002). Depression: Perspectives from affective neuroscience. *Annual Review of Psychology, 53*, 545–574.

Dickerson, S. S., Gable, S. L., Irwin, M. R., Aziz, N., & Kemeny, M. E. (2009). Social-evaluative threat and proinflammatory cytokine regulation an experimental laboratory investigation. *Psychological Science, 20*, 1237–1244.

Dickerson, S. S., Mycek, P. J., & Zaldivar, F. (2008). Negative social evaluation, but not mere social presence, elicits cortisol responses to a laboratory stressor task. *Health Psychology, 27*(1), 116–121.

Eisenberger, N. I., Taylor, S. E., Gable, S. L., Hilmert, C. J., & Lieberman, M. D. (2007). Neural pathways link social support to attenuated neuroendocrine stress responses. *Neuroimage, 35*(4), 1601–1612.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2*, 493–501.

Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia, 45*(1), 32–41.

Goldman, A. I. (2005). Imitation, mind reading, and simulation. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation: From neuroscience to social science (Vol. 2: Imitation, human development, and culture*, pp. 79–93). Cambridge, MA: MIT Press.

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology, 85*, 616–626.

Ito, T.A., & Urland, G.R. (2005). The influence of processing objectives on the perception of faces: An ERP study of race and gender perception. *Cognitive, Affective, and Behavioral Neuroscience, 5*, 21–36.

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience, 14*, 785–794.

Lazarus, R. S., (1974). Psychological stress and coping in adaptation and illness. *International Journal of Psychiatry in Medicine, 5*, 321–333.

Mendes, W. B., Blascovich, J., Lickel, B., & Hunter, S. (2002). Challenge and threat during social interactions with White and Black men. *Personality and Social Psychology Bulletin, 28*, 939–952.

Mitchell, J. P., Neil Macrae, C., & Banaji, M. R. (2005). Forming impressions of people versus inanimate objects: social-cognitive processing in the medial prefrontal cortex. *Neuroimage*, *26*(1), 251–257.

Mouchetant-Rostaing, Y., & Giard, M. H. (2003). Electrophysiological correlates of age and gender perception on human faces. *Journal of Cognitive Neuroscience*, *15*, 900–910.

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron*, *48*, 175.

Plant, E. A., & Devine, P. G. (2003). The antecedents and implications of interracial anxiety. *Personality and Social Psychology Bulletin, 29*, 790–801.

Schultz, J., Imamizu, H., Kawato, M., & Frith, C. D. (2004). Activation of the human superior temporal gyrus during observation of goal attribution by intentional objects. *Journal of Cognitive Neuroscience*, *16*, 1695–1705.

Selye, H. (1950). *The physiology and pathology of exposure to stress*. Montreal: Acta Inc.

Srivastava, S., Guglielmo, S., & Beer, J. S. (2010). Perceiving others' personalities: Examining the dimensionality, assumed similarity to the self, and stability of perceiver effects. *Journal of Personality and Social Psychology*, *98*, 520.

Stephan, W. G., & Stephan, C. W. (1985). Intergroup anxiety. *Journal of Social Issues*, *41*(3), 157–175.

Whalen, P. J., & Phelps, E. A. (2009). *The human amygdala*. New York, NY: The Guilford Press.

18.14.5.5 Attributions

Adapted by Kenneth A. Koenigshofer, PhD, from Ito, T. A. & Kubota, J. T. (2021). Social neuroscience. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology.* Champaign, IL: DEF publishers. Retrieved from http://noba.to/qyekc5gf

Authors



- Tiffany A. Ito is a Professor of Psychology and Neuroscience at the University of Colorado Boulder. Her research integrates neuroscience methods and theories to better understand social processes, with a particular focus on aspects of stereotyping and prejudice.



- Jennifer Kubota is an assistant professor at University of Delaware. She received her Ph.D. in social neuroscience from The University of Colorado Boulder. Her work focuses on the psychological and neural substrates of impression formation and their relation to decision-making.

18.14.5.6 Creative Commons License

Social Neuroscience by Tiffany A. Ito and Jennifer T. Kubota is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Permissions beyond the scope of this license may be available in our Licensing Agreement.

Minor editing of this content by Kenneth A. Koenigshofer, PhD. Chaffey College.

18.14.5.7 Theory of Mind and Social Cognition

18.14.5.8 Overview

We now continue exploration of social cognition by examining an important cognitive ability in humans essential for successful social interaction--theory of mind, a capacity that some scholars prefer to label “mentalizing” or “mindreading,” the ability to perceive and interpret other people’s behavior in terms of their mental states. The capacity may have evolved sometime in the last few million years. Theory of mind is thought to be a prerequisite for natural language acquisition, strategic social interaction, reflexive thought, and moral judgment. Humans need to understand minds in order to engage in the kinds of complex interactions that social communities (small and large) require--interactions that have given rise to the complex products of human cultural evolution. The capacity is severely limited in autism.

18.14.5.9 **Theory of Mind**

By Bertram Malle

Brown University

One of the most remarkable human capacities is to perceive and understand mental states. This capacity, often labeled “theory of mind,” consists of an array of psychological processes that play essential roles in human social life. We review some of these roles, examine what happens when the capacity is deficient, and explore the many processes that make up the capacity to understand minds.

Share this URL

http://noba.to/a8wpytg3

- Autism Spectrum Disorder
- Empathy
- Explanations
- Intentionality
- Perspective taking
- Social cognition

18.14.5.10 Learning Objectives

- Explain what theory of mind is.
- Enumerate the many domains of social life in which theory of mind is critical.
- Describe some characteristics of how autistic individuals differ in their processing of others’ minds.
- Describe and explain some of the many concepts and processes that comprise the human understanding of minds.
- Have a basic understanding of how ordinary people explain unintentional and intentional behavior.

18.14.6 Introduction

One of the most fascinating human capacities is the ability to perceive and interpret other people’s behavior in terms of their mental states. Having an appreciation for the workings of another person’s mind is considered a prerequisite for natural language acquisition (Baldwin & Tomasello, 1998), strategic social interaction (Zhang, Hedden, & Chia, 2012), reflexive thought (Bogdan, 2000), and moral judgment (Guglielmo, Monroe, & Malle, 2009). This capacity develops from early beginnings in the first year of life to the adult’s fast and often effortless understanding of others’ thoughts, feelings, and intentions. And though we must speculate about its evolutionary origin, we do have indications that the capacity evolved sometime in the last few million years.

In this module we will focus on two questions: What is the role of understanding others’ minds in human social life? And what is known about the mental processes that underlie such understanding? For simplicity, we will label this understanding “**theory of mind**,” even though it is not literally a “theory” that people have about the mind; rather, it is a capacity that some scholars prefer to

label “mentalizing” or “mindreading.” But we will go behind all these labels by breaking down the capacity into distinct components: the specific concepts and mental processes that underlie the human understanding of minds.

First, let’s get clear about the roles that this understanding plays in social life.

18.14.7 The Role of Theory of Mind in Social Life



Figure **18.14.6**: We rely on the theory of mind in social situations to infer what others are thinking and feeling. Among other things, this capability helps us work successfully in teams. [Image: Office of Public Affairs, https://goo.gl/O8zvFj, CC BY-SA 2.0, https://goo.gl/rxiUsF]

Put yourself in this scene: You observe two people’s movements, one behind a large wooden object, the other reaching behind him and then holding a thin object in front of the other. Without a theory of mind you would neither understand what this movement stream meant nor be able to predict either person’s likely responses. With the capacity to interpret certain physical movements in terms of mental states, perceivers can parse this complex scene into intentional actions of reaching and giving (Baird & Baldwin, 2001); they can interpret the actions as instances of offering and trading; and with an appropriate cultural script, they know that all that was going on was a customer pulling out her credit card with the intention to pay the cashier behind the register. People’s theory of mind thus frames and interprets perceptions of human behavior in a particular way—as perceptions of agents who can act intentionally and who have desires, beliefs, and other mental states that guide their actions (Perner, 1991; Wellman, 1990).

Not only would social perceivers without a theory of mind be utterly lost in a simple payment interaction; without a theory of mind, there would probably be no such things as cashiers, credit cards, and payment (Tomasello, 2003). Plain and simple, humans need to understand minds in order to engage in the kinds of complex interactions that social communities (small and large) require. And it is these complex social interactions that have given rise, in human cultural evolution, to houses, cities, and nations; to books, money, and computers; to education, law, and science.

The list of social interactions that rely deeply on theory of mind is long; here are a few highlights.

- Teaching another person new actions or rules by taking into account what the learner knows or doesn’t know and how one might best make him understand.
- Learning the words of a language by monitoring what other people attend to and are trying to do when they use certain words.
- Figuring out our social standing by trying to guess what others think and feel about us.
- Sharing experiences by telling a friend how much we liked a movie or by showing her something beautiful.

- Collaborating on a task by signaling to one another that we share a goal and understand and trust the other's intention to pursue this joint goal.

18.14.8 Autism and Theory of Mind



Figure **18.14.7**: Individuals with autism can have a harder time using the theory of mind because it involves processing facial expressions and inferring people's intentions. A look that might convey a lot of meaning to most people conveys little or nothing to someone with autism. [Image: WarzauWynn, https://goo.gl/gUO8HE, CC BY-NC 2.0, https://goo.gl/tgFydH]

Another way of appreciating the enormous impact that theory of mind has on social interactions is to study what happens when the capacity is severely limited, as in the case of autism (Tager-Flusberg, 2007). In a fascinating discussion in which (high-functioning) autistic individuals talk about their difficulties with other people's minds (Blackburn, Gottschewski, George, & L—, 2000), one person reports: "I know people's faces down to the acne scars on the left corners of their chins . . . and how the hairs of their eyebrows curl. . . . The best I can do is start picking up bits of data during my encounter with them because there's not much else I can do. . . . I'm not sure what kind of information about them I'm attempting to process." What seems to be missing, as another person with autism remarks, is an "automatic processing of 'people information.'" Some autistic people report that they perceive others "in a more analytical way." This analytical mode of processing, however, is very tiresome and slow: "Given time I may be able to analyze someone in various ways, and seem to get good results, but may not pick up on certain aspects of an interaction until I am obsessing over it hours or days later" (Blackburn et al., 2000).

So what is this magical potion that allows most people to gain quick and automatic access to other people's minds and to recognize the meaning underlying human behavior? Scientific research has accumulated a good deal of knowledge in the past few decades, and here is a synopsis of what we know.

18.14.9 The Mental Processes Underlying Theory of Mind

The first thing to note is that "theory of mind" is not a single thing. What underlies people's capacity to recognize and understand mental states is a whole host of components—a toolbox, as it were, for many different but related tasks in the social world (Malle, 2008). Figure 1 shows some of the most important tools, organized in a way that reflects the complexity of involved processes: from simple and automatic on the bottom to complex and deliberate on the top. This organization also reflects development—from tools that infants master within the first 6–12 months to tools they need to acquire over the next 3–5 years. Strikingly, the organization also reflects evolution: monkeys have available the tools on the bottom; chimpanzees have available the tools at the second level; but only humans master the remaining tools above. Let's look at a few of them in more detail.

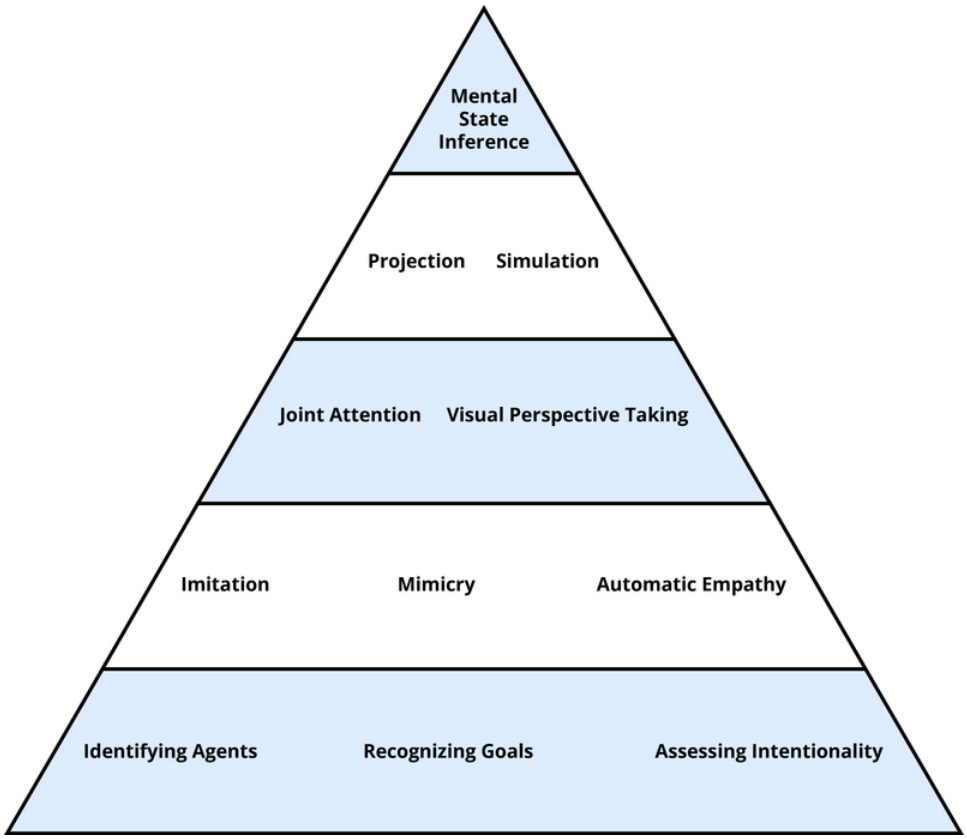


Figure **18.14.8**: Some of the major tools of theory of mind, with the bottom showing simple, automatic, early developing, and evolutionarily old processes, and the top showing complex, more deliberate, late developing, and evolutionarily recent processes.

18.14.10 Agents, Goals, and Intentionality

The *agent* category allows humans to identify those moving objects in the world that can act on their own. Features that even very young children take to be indicators of being an agent include being self-propelled, having eyes, and reacting systematically to the interaction partner's behavior, such as following gaze or imitating (Johnson, 2000; Premack, 1990).

The process of *recognizing goals* builds on this agent category, because agents are characteristically directed toward goal objects, which means they seek out, track, and often physically contact said objects. Even before the end of their first year, infants recognize that humans reach toward an object they strive for even if that object changes location or if the path to the object contains obstacles (Gergely, Nádasdy, Csibra, & Bíró, 1995; Woodward, 1998). What it means to recognize goals, therefore, is to see the systematic and predictable relationship between a particular agent pursuing a particular object across various circumstances.

Through learning to recognize the many ways by which agents pursue goals, humans learn to pick out behaviors that are **intentional**. The concept of **intentionality** is more sophisticated than the *goal* concept. For one thing, human perceivers recognize that some behaviors can be unintentional even if they were goal-directed—such as when you unintentionally make a fool of yourself even though you had the earnest goal of impressing your date. To act intentionally you need, aside from a goal, the right kinds of beliefs about how to achieve the goal. Moreover, the adult concept of intentionality requires that an agent have the *skill* to perform the intentional action in question: If I am flipping a coin, trying to make it land on heads, and if I get it to land on heads on my first try, you would not judge my action of making it land on heads as intentional—you would say it was luck (Malle & Knobe, 1997).

18.14.11 Imitation, Synchrony, and Empathy



Figure **18.14.9**: It's natural when having a conversation to unconsciously synchronize with our partners. [Image: Jacopo Aneghini Photos , https://goo.gl/QDpPln, CC BY-NC 2.0, https://goo.gl/VnKlK8]

Imitation and empathy are two other basic capacities that aid the understanding of mind from childhood on (Meltzoff & Decety, 2003). *Imitation* is the human tendency to carefully observe others' behaviors and do as they do—even if it is the first time the perceiver has seen this behavior. A subtle, automatic form of imitation is called **mimicry**, and when people mutually mimic one another they can reach a state of **synchrony**. Have you ever noticed when two people in conversation take on similar gestures, body positions, even tone of voice? They "synchronize" their behaviors by way of (largely) unconscious imitation. Such synchrony can happen even at very low levels, such as negative physiological arousal (Levenson & Ruef, 1992), though the famous claim of synchrony in women's menstrual cycles is a myth (Yang & Schank, 2006). Interestingly, people who enjoy an interaction synchronize their behaviors more, and increased synchrony (even manipulated in an experiment) makes people enjoy their interaction more (Chartrand & Bargh, 1999). Some research findings suggest that synchronizing is made possible by brain mechanisms that tightly link perceptual information with motor information (when I see you move your arm, my arm-moving program is activated). In monkeys, highly specialized so-called **mirror neurons** fire both when the monkey sees a certain action and when it performs that same action (Rizzolatti, Fogassi, & Gallese, 2001). In humans, however, things are a bit more complex. In many everyday settings, people perceive uncountable behaviors and fortunately don't copy all of them (just consider walking in a crowd—hundreds of your mirror neurons would fire in a blaze of confusion). Human imitation and mirroring is selective, triggering primarily actions that are relevant to the perceiver's current state or aim.

Automatic empathy builds on imitation and synchrony in a clever way. If Bill is sad and expresses this emotion in his face and body, and if Elena watches or interacts with Bill, then she will subtly imitate his dejected behavior and, through well-practiced associations of certain behaviors and emotions, she will feel a little sad as well (Sonnby-Borgström, Jönsson, & Svensson, 2003). Thus, she empathizes with him—whether she wants to or not. Try it yourself. Type "sad human faces" into your Internet search engine and select images from your results. Look at 20 photos and pay careful attention to what happens to your face and to your mood. Do you feel almost a "pull" of some of your facial muscles? Do you feel a tinge of melancholy?

18.14.12 Joint Attention, Visual Perspective Taking

Going beyond the automatic, humans are capable of actively engaging with other people's mental states, such as when they enter into situations of **joint attention**—like Marissa and Noah, who are each looking at an object and are both aware that each of them is looking at the object. This sounds more complicated than it really is. Just point to an object when a 3-year old is around and notice how both the child and you check in with each other, ensuring that you are really *jointly* engaging with the object. Such shared engagement is critical for children to learn the meaning of objects—both their value (is it safe and rewarding to approach?) and the words that refer to them (what do you call this?). When I hold up my keyboard and show it to you, we are jointly attending to it, and if I then say it's called "Tastatur" in German, you know that I am referring to the keyboard and not to the table on which it had been resting.

Another important capacity of engagement is **visual perspective taking**: You are sitting at a dinner table and advise another person on where the salt is—do you consider that it is to *her* left even though it is to *your* right? When we overcome our egocentric perspective this way, we imaginatively adopt the other person's spatial viewpoint and determine how the world looks from their perspective. In fact, there is evidence that we mentally "rotate" toward the other's spatial location, because the farther away the person sits (e.g., 60, 90, or 120 degrees away from you) the longer it takes to adopt the person's perspective (Michelon & Zacks, 2006).

18.14.13 Projection, Simulation (and the Specter of Egocentrism)

When imagining what it might be like to be in another person's *psychological* position, humans have to go beyond mental rotation. One tool to understand the other's thoughts or feelings is **simulation**—using one's own mental states as a model for others' mental states: "What would it feel like sitting across from the stern interrogator? *I* would feel scared . . ." An even simpler form of such modeling is the assumption that the other thinks, feels, wants what we do—which has been called the "like-me" assumption (Meltzoff, 2007) or the inclination toward *social* **projection** (Krueger, 2007). In a sense, this is an absence of perspective taking, because we assume that the other's perspective equals our own. This can be an effective strategy if we share with the other person the same environment, background, knowledge, and goals, but it gets us into trouble when this presumed common ground is in reality lacking. Let's say you know that Brianna doesn't like Fred's new curtains, but you hear her exclaim to Fred, "These are beautiful!" Now you have to predict whether Fred can figure out that Brianna was being sarcastic. It turns out that you will have a hard time suppressing your own knowledge in this case and you may overestimate how easy it is for Fred to spot the sarcasm (Keysar, 1994). Similarly, you will overestimate how visible that pimple is on your chin—even though it feels big and ugly to you, in reality very few people will ever notice it (Gilovich & Savitsky, 1999). So the next time when you spot a magnificent bird high up in the tree and you get impatient with your friend who just can't see what is clearly obvious, remember: it's obvious to *you*.

What all these examples show is that people use their own current state—of knowledge, concern, or perception—to grasp other people's mental states. And though they often do so correctly, they also get things wrong at times. This is why couples counselors, political advisors, and Buddhists agree on at least one thing: we all need to try harder to recognize our egocentrism and actively take other people's perspective—that is, grasp their actual mental states, even if (or especially when) they are different from our own.

18.14.14 Explicit Mental State Inference

The ability to truly take another person's perspective requires that we separate what we want, feel, and know from what the other person is likely to want, feel, and know. To do so humans make use of a variety of information. For one thing, they rely on stored knowledge—both general knowledge ("Everybody would be nervous when threatened by a man with a gun") and agent-specific knowledge ("Joe was fearless because he was trained in martial arts"). For another, they critically rely on perceived facts of the concrete situation—such as what is happening to the agent, the agent's facial expressions and behaviors, and what the person saw or didn't see.

This capacity of integrating multiple lines of information into a mental-state inference develops steadily within the first few years of life, and this process has led to a substantial body of research (Wellman, Cross, & Watson, 2001). The research began with a clever experiment by Wimmer and Perner (1983), who tested whether children can pass a **false-belief test** (see Figure 2). The child is shown a picture story of Sally, who puts her ball in a basket and leaves the room. While Sally is out of the room, Anne comes

along and takes the ball from the basket and puts it inside a box. The child is then asked *where* Sally thinks the ball is located when she comes back to the room. Is she going to look first *in the box* or *in the basket*?

The right answer is that she will look in the basket, because that's where she put it and thinks it is; but we have to infer this *false belief* against our own better knowledge that the ball is in the box. This is very difficult for children before the age of 4, and it usually takes some cognitive effort in adults (Epley, Morewedge, & Keysar, 2004).

The challenge is clear: People are good at automatically relating to other people, using their own minds as a fitting model for others' minds. But people need to recognize when to step out of their own perspective and truly represent the other person's perspective—which may harbor very different thoughts, feelings, and intentions.

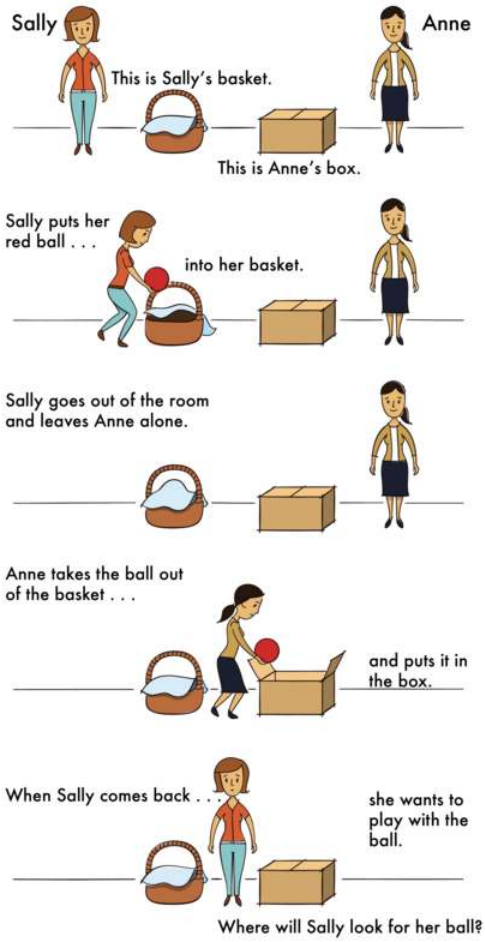


Figure **18.14.10**: Sally–Anne task to test children's ability to infer false beliefs.

18.14.15 Tools in Summary

We have seen that the human understanding of other minds relies on many tools. People process such information as motion, faces, and gestures and categorize it into such concepts as agent, intentional action, or fear. They rely on relatively automatic psychological processes, such as imitation, joint attention, and projection. And they rely on more effortful processes, such as simulation and mental-state inference. These processes all link behavior that humans observe to mental states that humans infer. If we call this stunning capacity a "theory," it is a theory of mind *and* behavior.

18.14.16 Folk Explanations of Behavior

Nowhere is this mind–behavior link clearer than in **people's explanations of behavior**—when they try to understand why somebody acted or felt a certain way. People have a strong need to answer such "why" questions, from the trivial to the significant: why the neighbor's teenage daughter is wearing a short skirt in the middle of winter; why the policeman is suddenly so friendly; why the murderer killed three people. The need to explain this last behavior seems puzzling, because typical benefits of explanation

are absent: We do not need to predict or control the criminal's behavior since we will never have anything to do with him. Nonetheless, we have an insatiable desire to *understand*, to find *meaning* in this person's behavior—and in people's behavior generally.

This makes evolutionary sense for at least two reasons: 1) because we are highly social creatures by virtue of our genetic evolution as a species, and success in the social group, including being valued by others in your group, was especially important to our survival in our evolutionary past (if you "don't get" others and consequently aren't valued and liked by the group, they are less likely to risk themselves to save you from a predator or to share food and other resources with you), strong motivation to understand other people would be favored by natural selection; 2) strong motivation to understand behavior of others increases chances of being able to predict behavior of others, and to predict their intentions, thereby allowing prediction of their future behavior allowing strategic preparation for it increasing your success in your social interactions, essential to human adaptation.



Figure 18.14.11: When people behave in ways that don't make sense to us we often feel compelled to come up with reasonable explanations. What are their motives? What circumstances would lead to this? How could this happen? [Image: lwpkommunikacio, https://goo.gl/5x1SAN, CC BY 2.0, https://goo.gl/axKpvj]

Older theories of how people explain and understand behavior suggested that people merely identify causes of the behavior (e.g., Kelley, 1967). That is true for most unintentional behaviors—tripping, having a headache, calling someone by the wrong name. But to explain *intentional* behaviors, people use a more sophisticated framework of interpretation, which follows directly from their concept of intentionality and the associated mental states they infer (Malle, 2004). We have already mentioned the complexity of people's concept of intentionality; here it is in full (Malle & Knobe, 1997): For an agent to perform a behavior intentionally, she must have a desire for an outcome (what we had called a goal), beliefs about how a particular action leads to the outcome, and an intention to perform that action; if the agent then actually performs the action with awareness and skill, people take it to be an intentional action. To explain *why* the agent performed the action, humans try to make the inverse inference of *what desire* and *what beliefs* the agent had that led her to so act, and these inferred desires and beliefs are the *reasons* for which she acted. What was her reason for wearing a short skirt in the winter? "She wanted to annoy her mother." What was the policeman's reason for suddenly being so nice? "He thought he was speaking with an influential politician." What was his reason for killing three people? In fact, with such extreme actions, people are often at a loss for an answer. If they do offer an answer, they frequently retreat to "causal history explanations" (Malle, 1999), which step outside the agent's own reasoning and refer instead to more general

background facts—for example, that he was mentally ill or a member of an extremist group. But people clearly prefer to explain others' actions by referring to their beliefs and desires, the specific reasons for which they acted.

By relying on a theory of mind, explanations of behavior make meaningful what would otherwise be inexplicable motions—just like in our initial example of two persons passing some object between them. We recognize that the customer wanted to pay and that's why she passed her credit card to the cashier, who in turn knew that he was given a credit card and swiped it. It all seems perfectly clear, almost trivial to us. But that is only because humans have a theory of mind and use it to retrieve the relevant knowledge, simulate the other people's perspective, infer beliefs and desires, and explain what a given action means. Humans do this effortlessly and often accurately. Moreover, they do it within seconds or less. What's so special about that? Well, it takes years for a child to develop this capacity, and it took our species a few million years to evolve it. That's pretty special.

18.14.16.1 Outside Resources

Blog: On the debate about menstrual synchrony

http://blogs.scientificamerican.com/context-and-variation/2011/11/16/menstrual-synchrony/

Blog: On the debates over mirror neurons

http://blogs.scientificamerican.com/guest-blog/2012/11/06/whats-so-special-about-mirror-neurons/

Book: First and last chapters of Zunshine, L. (2006). *Why we read fiction: Theory of mind and the novel*. Columbus, OH: Ohio State University Press.

https://ohiostatepress.org/Books/Book PDFs/Zunshine Why.pdf

Movie: A movie that portrays the social difficulties of a person with autism: *Adam* (Fox Searchlight Pictures, 2009)

http://www.imdb.com/title/tt1185836/?ref_=fn_tt_tt_1

ToM and Autism TEDx Talks

https://www.ted.com/playlists/153/the_autism_spectrum

Video: TED talk on autism

http://www.ted.com/talks/temple_grandin_the_world_needs_all_kinds_of_minds.html

Video: TED talk on empathy

http://blog.ted.com/2011/04/18/a-radical-experiment-in-empathy-sam-richards-at-ted-com/

Video: TED talk on theory of mind and moral judgment

http://www.ted.com/talks/rebecca_saxe_how_brains_make_moral_judgments.html

Video: Test used by Baron Cohen (prior to the core study) to investigate whether autistic children had a theory of mind by using a false belief task.





Video: Theory of mind development





18.14.16.2 Discussion Questions

1. Recall a situation in which you tried to infer what a person was thinking or feeling but you just couldn't figure it out, and recall another situation in which you tried the same but succeeded. Which tools were you able to use in the successful case that you didn't or couldn't use in the failed case?
2. Mindfulness training improves keen awareness of one's own mental states. Look up a few such training programs (easily found online) and develop a similar training program to improve awareness of other people's minds.
3. In the near future we will have robots that closely interact with people. Which theory of mind tools should a robot definitely have? Which ones are less important? Why?
4. Humans assume that everybody has the capacity to make choices and perform intentional actions. But in a sense, a choice is just a series of brain states, caused by previous brain states and states of the world, all governed by the physical laws of the universe. Is the concept of choice an illusion?
5. The capacity to understand others' minds is intimately related to another unique human capacity: language. How might these two capacities have evolved? Together? One before the other? Which one?

18.14.16.3 Vocabulary

Automatic empathy

A social perceiver unwittingly taking on the internal state of another person, usually because of mimicking the person's expressive behavior and thereby feeling the expressed emotion.

False-belief test

An experimental procedure that assesses whether a perceiver recognizes that another person has a false belief—a belief that contradicts reality.

Folk explanations of behavior

People's natural explanations for why somebody did something, felt something, etc. (differing substantially for unintentional and intentional behaviors).

Intention

An agent's mental state of committing to perform an action that the agent believes will bring about a desired outcome.

Intentionality

The quality of an agent's performing a behavior intentionally—that is, with skill and awareness and executing an intention (which is in turn based on a desire and relevant beliefs).

Joint attention

Two people attending to the same object and being aware that they both are attending to it.

Mimicry

Copying others' behavior, usually without awareness.

Mirror neurons

Neurons identified in monkey brains that fire both when the monkey performs a certain action and when it perceives another agent performing that action.

Projection

A social perceiver's assumption that the other person wants, knows, or feels the same as the perceiver wants, know, or feels.

Simulation

The process of representing the other person's mental state.

Synchrony

Two people displaying the same behaviors or having the same internal states (typically because of mutual mimicry).

Theory of mind

The human capacity to understand minds, a capacity that is made up of a collection of concepts (e.g., agent, intentionality) and processes (e.g., goal detection, imitation, empathy, perspective taking).

Visual perspective taking

Can refer to visual perspective taking (perceiving something from another person's spatial vantage point) or more generally to effortful mental state inference (trying to infer the other person's thoughts, desires, emotions).

18.14.16.4 References

Baird, J. A., & Baldwin, D. A. (2001). Making sense of human behavior: Action parsing and intentional inference. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 193–206). Cambridge, MA:MIT Press.

Baldwin, D. A., & Tomasello, M. (1998). Word learning: A window on early pragmatic understanding. In E. V. Clark (Ed.), *The proceedings of the twenty-ninth annual child language research forum* (pp. 3–23). Chicago, IL: Center for the Study of Language and Information.

Blackburn, J., Gottschewski, K., George, E., & L—, N. (2000, May). A discussion about theory of mind: From an autistic perspective. *Proceedings of Autism Europe's 6th International Congress*. Glasgow. Retrieved from http://archive.autistics.org/library/AE2000-ToM.html.

Bogdan, R. (2000). *Minding minds: Evolving a reflexing mind by interpreting others*. Cambridge, MA: MIT Press.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology, 76*, 893–910.

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology, 40*, 760–768.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56*, 165–193.

Gilovich, T., & Savitsky, K. (1999). The spotlight effect and the illusion of transparency: Egocentric assessments of how we are seen by others. *Current Directions in Psychological Science, 8*, 165–168.

Guglielmo, S., Monroe, A. E., & Malle, B. F. (2009). At the heart of morality lies folk psychology. *Inquiry: An Interdisciplinary Journal of Philosophy, 52*, 449–466.

Johnson, S. C. (2000). The recognition of mentalistic agents in infancy. *Trends in Cognitive Sciences, 4*, 22–28.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15, pp. 192–240). Lincoln: University of Nebraska Press.

Keysar, B. (1994). The illusory transparency of intention: Linguistic perspective taking in text. *Cognitive Psychology, 26*, 165–208.

Krueger, J. I. (2007). From social projection to social behaviour. *European Review of Social Psychology, 18*, 1–35.

Levenson, R. W., & Ruef, A. M. (1992). Empathy: A physiological substrate. *Journal of Personality and Social Psychology, 63*, 234–246.

Malle, B. F. (2008). The fundamental tools, and possibly universals, of social cognition. In R. M. Sorrentino & S. Yamaguchi (Eds.), *Handbook of motivation and cognition across cultures* (pp. 267–296). New York, NY: Elsevier/Academic Press.

Malle, B. F. (2004). *How the mind explains behavior: Folk explanations, meaning, and social interaction*. Cambridge, MA: MIT Press.

Malle, B. F. (1999). How people explain behavior: A new theoretical framework. *Personality and Social Psychology Review, 3*, 23–48.

Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology, 33*, 101–121.

Meltzoff, A. N. (2007). “Like me”: A foundation for social cognition. *Developmental Science, 10*, 126–134.

Meltzoff, A. N., & Decety, J. (2003). What imitation tells us about social cognition: A rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*, 491–500.

Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception & Psychophysics, 68*, 327–337.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Premack, D. (1990). The infant’s theory of self-propelled objects. *Cognition, 36*, 1–16.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience, 2*, 661–670.

Sonnby-Borgström, M., Jönsson, P., & Svensson, O. (2003). Emotional empathy as related to mimicry reactions at different levels of information processing. *Journal of Nonverbal Behavior, 27*, 3–23.

Tager-Flusberg, H. (2007). Evaluating the theory-of-mind hypothesis of autism. *Current Directions in Psychological Science, 16*, 311–315.

Tomasello, M. (2003). The key is social cognition. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 47–57). Cambridge, MA: MIT Press.

Wellman, H. M. (1990). *The child’s theory of mind*. Cambridge, MA: MIT Press.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*, 655–684.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*, 1–34.

Yang, Z., & Schank, J. C. (2006). Women do not synchronize their menstrual cycles. *Human Nature*, *17*, 433–447.

Zhang, J., Hedden, T., & Chia, A. (2012). Perspective-taking and depth of theory-of-mind reasoning in sequential-move games. *Cognitive Science*, *36*, 560–573.

18.14.16.5 Attributions

Adapted by Kenneth A. Koenigshofer, Ph.D. from Malle, B. (2021). Theory of mind. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology.* Champaign, IL: DEF publishers. Retrieved from http://noba.to/a8wpytg3

18.14.16.6 **Authors**



- Bertram Malle, Professor at Brown University, was trained in psychology, philosophy, and linguistics in Austria and at Stanford University. He received an SESP Dissertation award, an NSF CAREER award, and was president of the Society of Philosophy and Psychology. His research focuses on social cognition, moral judgment, and more recently human-robot interaction.

18.14.16.7 Creative Commons License

Theory of Mind by Bertram Malle is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Permissions beyond the scope of this license may be available in our Licensing Agreement.

Paragraph on evolutionary rationale for motivation to understand others in "Folk Explanations of Behavior" added by Kenneth A. Koenigshofer, PhD. Chaffey College.

18.14.16.8 **Disorders of the Social Brain**

18.14.16.9 Overview

We now continue exploration of social cognition by examining a spectrum of disorders of social processing and the brain mechanisms involved. Here we encounter social neuroscience once again. Autism is in the category of *pervasive developmental disorders*, which includes Asperger's disorder, childhood disintegrative disorder, autistic disorder, and pervasive developmental disorder - not otherwise specified. These disorders, together, are labeled *autism spectrum disorder* (ASD). ASD is defined by the presence of profound difficulties in social interactions and communication combined with the presence of repetitive or restricted interests, cognition and behaviors. The social brain is a set of interconnected neuroanatomical structures that process social information, enabling the recognition of other individuals and the evaluation their mental states. The social brain is hypothesized to consist of the amygdala, the orbital frontal cortex (OFC), fusiform gyrus (FG), and the posterior superior temporal sulcus (STS) region, among other structures. Because autism is a developmental disorder, it is particularly important to diagnose and treat ASD early in life.

18.14.17 **Autism: Insights from the Study of the Social Brain**

By Kevin A. Pelphrey

Yale University

People with autism spectrum disorder (ASD) suffer from a profound social disability. Social neuroscience is the study of the parts of the brain that support social interactions or the "social brain." This module provides an overview of ASD and focuses on understanding how social brain dysfunction leads to ASD. Our increasing understanding of the social brain and its dysfunction in

ASD will allow us to better identify the genes that cause ASD and will help us to create and pick out treatments to better match individuals. Because social brain systems emerge in infancy, social neuroscience can help us to figure out how to diagnose ASD even before the symptoms of ASD are clearly present. This is a hopeful time because social brain systems remain malleable well into adulthood and thus open to creative new interventions that are informed by state-of-the-art science.

Share this URL

http://noba.to/yqdepwgt

- Autism Spectrum Disorder
- Social brain
- Social neuroscience

18.14.17.10 Learning Objectives

- Know the basic symptoms of ASD.
- Distinguish components of the social brain and understand their dysfunction in ASD.
- Appreciate how social neuroscience may facilitate the diagnosis and treatment of ASD.

18.14.18 Defining Autism Spectrum Disorder

Autism Spectrum Disorder (ASD) is a developmental disorder that usually emerges in the first three years and persists throughout the individual's life. Though the key symptoms of ASD fall into three general categories (see below), each person with ASD exhibits symptoms in these domains in different ways and to varying degrees. This *phenotypic heterogeneity* reflects the high degree of variability in the genes underlying ASD (Geschwind & Levitt, 2007). Though we have identified genetic differences associated with individual cases of ASD, each accounts for only a small number of the actual cases, suggesting that no single genetic cause will apply in the majority of people with ASD. There is currently no biological test for ASD.

Autism is in the category of *pervasive developmental disorders*, which includes Asperger's disorder, childhood disintegrative disorder, autistic disorder, and pervasive developmental disorder - not otherwise specified. These disorders, together, are labeled *autism spectrum disorder* (ASD). ASD is defined by the presence of profound difficulties in social interactions and communication combined with the presence of repetitive or restricted interests, cognitions and behaviors. The diagnostic process involves a combination of parental report and clinical observation. Children with significant impairments across the social/communication domain who also exhibit repetitive behaviors can qualify for the ASD diagnosis. There is wide variability in the precise symptom profile an individual may exhibit.

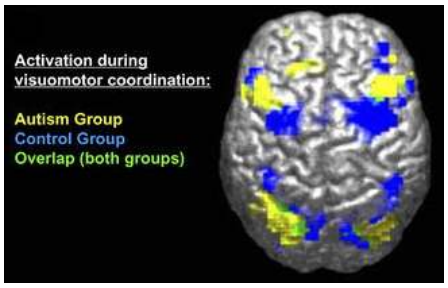


Figure **18.14.12**: The red lines indicate the scan paths (collection of eye movements) used by people with (right column) and without (left column) autism to explore faces. Modified from Pelphrey et al., (2002).

Since Kanner first described ASD in 1943, important commonalities in symptom presentation have been used to compile criteria for the diagnosis of ASD. These diagnostic criteria have evolved during the past 70 years and continue to evolve (e.g., see the recent changes to the diagnostic criteria on the American Psychiatric Association's website, http://www.dsm5.org/), yet impaired social functioning remains a required symptom for an ASD diagnosis. Deficits in social functioning are present in varying degrees for simple behaviors such as eye contact, and complex behaviors like navigating the give and take of a group conversation for individuals of all functioning levels (i.e. high or low IQ). Moreover, difficulties with social information processing occur in both visual (e.g., Pelphrey et al., 2002) and auditory (e.g., Dawson, Meltzoff, Osterling, Rinaldi, & Brown, 1998) sensory modalities.

Consider the results of an eye tracking study in which Pelphrey and colleagues (2002) observed that individuals with autism did not make use of the eyes when judging facial expressions of emotion (see right panels of Figure 1). While repetitive behaviors or language deficits are seen in other disorders (e.g., obsessive-compulsive disorder and specific language impairment, respectively), basic social deficits of this nature are unique to ASD. Onset of the social deficits appears to precede difficulties in other domains (Osterling, Dawson, & Munson, 2002) and may emerge as early as 6 months of age (Maestro et al., 2002).

18.14.19 Defining the Social Brain

Within the past few decades, research has elucidated specific brain circuits that support perception of humans and other species. This *social perception* refers to "the initial stages in the processing of information that culminates in the accurate analysis of the dispositions and intentions of other individuals" (Allison, Puce, & McCarthy, 2000). Basic social perception is a critical building block for more sophisticated social behaviors, such as thinking about the motives and emotions of others. Brothers (1990) first suggested the notion of a **social brain**, a set of interconnected neuroanatomical structures that process social information, enabling the recognition of other individuals and the evaluation their mental states (e.g., intentions, dispositions, desires, and beliefs).

The social brain is hypothesized to consist of the amygdala, the orbital frontal cortex (OFC), fusiform gyrus (FG), and the posterior superior temporal sulcus (STS) region, among other structures. Though all areas work in coordination to support social processing, each appears to serve a distinct role. The amygdala helps us recognize the emotional states of others (e.g., Morris et al., 1996) and also to experience and regulate our own emotions (e.g., LeDoux, 1992). The OFC supports the "reward" feelings we have when we are around other people (e.g., Rolls, 2000). The FG, located at the bottom of the surface of the temporal lobes detects faces and supports face recognition (e.g., Puce, Allison, Asgari, Gore, & McCarthy, 1996). The posterior STS region recognizes the biological motion, including eye, hand and other body movements, and helps to interpret and predict the actions and intentions of others (e.g., Pelphrey, Morris, Michelich, Allison, & McCarthy, 2005).

18.14.20 Current Understanding of Social Perception in ASD



Figure 18.14.13: The human brain has specialized functions to help guide our social interactions. [Image: Allan Ajifo, https://goo.gl/jv4iXf, CC BY 2.0, https://goo.gl/BRvSA7]

The social brain is of great research interest because the social difficulties characteristic of ASD are thought to relate closely to the functioning of this brain network. **Functional magnetic resonance imaging (fMRI)** and **event-related potentials (ERP)** are complementary brain imaging methods used to study activity in the brain across the lifespan. Each method measures a distinct facet of brain activity and contributes unique information to our understanding of brain function.

FMRI uses powerful magnets to measure the levels of oxygen within the brain, which vary according to changes in neural activity. As the neurons in specific brain regions "work harder", they require more oxygen. FMRI detects the brain regions that exhibit a relative increase in blood flow (and oxygen levels) while people listen to or view social stimuli in the MRI scanner. The areas of the brain most crucial for different social processes are thus identified, with spatial information being accurate to the millimeter.

In contrast, ERP provides direct measurements of the firing of groups of neurons in the cortex. Non-invasive sensors on the scalp record the small electrical currents created by this neuronal activity while the subject views stimuli or listens to specific kinds of information. While fMRI provides information about *where* brain activity occurs, ERP specifies *when* by detailing the timing of processing at the millisecond pace at which it unfolds.

ERP and fMRI are complementary, with fMRI providing excellent *spatial resolution* and ERP offering outstanding *temporal resolution*. Together, this information is critical to understanding the nature of social perception in ASD. To date, the most thoroughly investigated areas of the social brain in ASD are the superior temporal sulcus (STS), which underlies the perception and interpretation of biological motion, and the fusiform gyrus (FG), which supports face perception. Heightened sensitivity to biological motion (for humans, motion such as walking) serves an essential role in the development of humans and other highly social species. Emerging in the first days of life, the ability to detect biological motion helps to orient vulnerable young to critical sources of sustenance, support, and learning, and develops independent of visual experience with biological motion (e.g., Simion, Regolin, & Bulf, 2008). This inborn "life detector" serves as a foundation for the subsequent development of more complex social behaviors (Johnson, 2006).



Figure **18.14.14**: From an evolutionary standpoint, it was incredibly important for our survival to maintain social relationships. Therefore, it makes sense that we would be able to recognize faces within the first few days of our infancy. [Image: donnierayjones, https://goo.gl/obrI2x, CC BY 2.0, https://goo.gl/v4Y0Zv]

From very early in life, children with ASD display reduced sensitivity to biological motion (Klin, Lin, Gorrindo, Ramsay, & Jones, 2009). Individuals with ASD have reduced activity in the STS during biological motion perception. Similarly, people at increased genetic risk for ASD but who do not develop symptoms of the disorder (i.e. unaffected siblings of individuals with ASD) show increased activity in this region, which is hypothesized to be a compensatory mechanism to offset genetic vulnerability (Kaiser et al., 2010).

In typical development, preferential attention to faces and the ability to recognize individual faces emerge in the first days of life (e.g., Goren, Sarty, & Wu, 1975). The special way in which the brain responds to faces usually emerges by three months of age (e.g., de Haan, Johnson, & Halit, 2003) and continues throughout the lifespan (e.g., Bentin et al., 1996). Children with ASD, however, tend to show decreased attention to human faces by six to 12 months (Osterling & Dawson, 1994). Children with ASD also show reduced activity in the FG when viewing faces (e.g., Schultz et al., 2000). Slowed processing of faces (McPartland, Dawson, Webb, Panagiotides, & Carver, 2004) is a characteristic of people with ASD that is shared by parents of children with ASD (Dawson, Webb, & McPartland, 2005) and infants at increased risk for developing ASD because of having a sibling with ASD (McCleery, Akshoomoff, Dobkins, & Carver, 2009). Behavioral and attentional differences in face perception and recognition are evident in children and adults with ASD as well (e.g., Hobson, 1986).

18.14.21 Exploring Diversity in ASD

Because of the limited quality of the behavioral methods used to diagnose ASD and current clinical diagnostic practice, which permits similar diagnoses despite distinct symptom profiles (McPartland, Webb, Keehn, & Dawson, 2011), it is possible that the group of children currently referred to as having ASD may actually represent different syndromes with distinct causes. Examination of the social brain may well reveal diagnostically meaningful subgroups of children with ASD. Measurements of the "where" and "when" of brain activity during social processing tasks provide reliable sources of the detailed information needed to profile children with ASD with greater accuracy. These profiles, in turn, may help to inform treatment of ASD by helping us to match specific treatments to specific profiles.

The integration of imaging methods is critical for this endeavor. Using face perception as an example, the combination of fMRI and ERP could identify who, of those individuals with ASD, shows anomalies in the FG and then determine the stage of information processing at which these impairments occur. Because different processing stages often reflect discrete cognitive processes, this level of understanding could encourage treatments that address specific processing deficits at the neural level.

For example, differences observed in the early processing stages might reflect problems with low-level visual perception, while later differences would indicate problems with higher-order processes, such as emotion recognition. These same principles can be applied to the broader network of social brain regions and, combined with measures of behavioral functioning, could offer a comprehensive profile of brain-behavior performance for a given individual. A fundamental goal for this kind of subgroup approach is to improve the ability to tailor treatments to the individual.

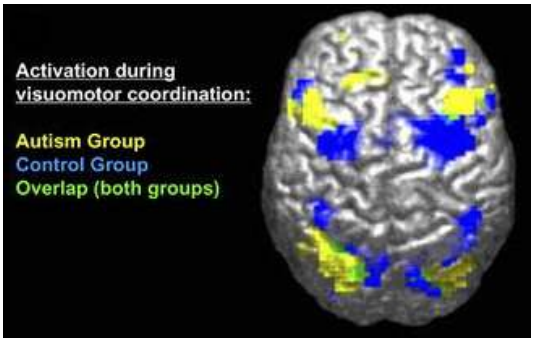


Figure **18.14.15**: Trying to diagnose the precise autism disorder can be difficult; many cases share similar symptoms. However, burgeoning technology, like the fMRI, allows clinicians a glimpse into the patient's brain and thus a better understanding of his or her disorder. [Image: Ralph-Axel Müller, https://goo.gl/WwxCV1, CC BY 2.5, https://goo.gl/0QtWcf]

Another objective is to improve the power of other scientific tools. Most studies of individuals with ASD compare groups of individuals, for example, individuals on with ASD compared to typically developing peers. However, studies have also attempted to compare children across the autism spectrum by group according to differential diagnosis (e.g., Asperger's disorder versus autistic disorder), or by other behavioral or cognitive characteristics (e.g., cognitively able versus intellectually disabled or anxious versus non-anxious). Yet, the power of a scientific study to detect these kinds of significant, meaningful, individual differences is only as strong as the accuracy of the factor used to define the compared groups.

The identification of distinct subgroups within the autism spectrum according to information about the brain would allow for a more accurate and detailed exposition of the individual differences seen in those with ASD. This is especially critical for the success of investigations into the genetic basis of ASD. As mentioned before, the genes discovered thus far account for only a small portion of ASD cases. If meaningful, quantitative distinctions in individuals with ASD are identified; a more focused examination into the genetic causes specific to each subgroup could then be pursued. Moreover, distinct findings from neuroimaging, or *biomarkers*, can help guide genetic research. **Endophenotypes**, or characteristics that are not immediately available to observation but that reflect an underlying genetic liability for disease, expose the most basic components of a complex psychiatric disorder and are more stable across the lifespan than observable behavior (Gottesman & Shields, 1973). By describing the key characteristics of ASD in these objective ways, neuroimaging research will facilitate identification of genetic contributions to ASD.

18.14.22 Atypical Brain Development Before the Emergence of Atypical Behavior

Because autism is a developmental disorder, it is particularly important to diagnose and treat ASD early in life. Early deficits in attention to biological motion, for instance, derail subsequent experiences in attending to higher level social information, thereby

driving development toward more severe dysfunction and stimulating deficits in additional domains of functioning, such as language development. The lack of reliable predictors of the condition during the first year of life has been a major impediment to the effective treatment of ASD. Without early predictors, and in the absence of a firm diagnosis until behavioral symptoms emerge, treatment is often delayed for two or more years, eclipsing a crucial period in which intervention may be particularly successful in ameliorating some of the social and communicative impairments seen in ASD.

In response to the great need for sensitive (able to identify subtle cases) and specific (able to distinguish autism from other disorders) early indicators of ASD, such as biomarkers, many research groups from around the world have been studying patterns of infant development using prospective longitudinal studies of infant siblings of children with ASD and a comparison group of infant siblings without familial risks. Such designs gather longitudinal information about developmental trajectories across the first three years of life for both groups followed by clinical diagnosis at approximately 36 months.



Figure **18.14.16**: If autism is diagnosed early enough, treatments have developed to the point that children with ASD can learn and grow to have more intensive social interactions. [Image: hepingting, https://goo.gl/TIoAcY, CC BY-SA 2.0, https://goo.gl/rxiUsF]

These studies are problematic in that many of the social features of autism do not emerge in typical development until after 12 months of age, and it is not certain that these symptoms will manifest during the limited periods of observation involved in clinical evaluations or in pediatricians' offices. Moreover, across development, but especially during infancy, behavior is widely variable and often unreliable, and at present, behavioral observation is the only means to detect symptoms of ASD and to confirm a diagnosis. This is quite problematic because, even highly sophisticated behavioral methods, such as eye tracking (see Figure 1), do not necessarily reveal reliable differences in infants with ASD (Ozonoff et al., 2010). However, measuring the brain activity associated with social perception can detect differences that do not appear in behavior until much later. The identification of biomarkers utilizing the imaging methods we have described offers promise for earlier detection of atypical social development.

ERP measures of brain response predict subsequent development of autism in infants as young as six months old who showed normal patterns of visual fixation (as measured by eye tracking) (Elsabbagh et al., 2012). This suggests the great promise of brain imaging for earlier recognition of ASD. With earlier detection, treatments could move from addressing existing symptoms to preventing their emergence by altering the course of abnormal brain development and steering it toward normality.

18.14.23 Hope for Improved Outcomes

The brain imaging research described above offers hope for the future of ASD treatment. Many of the functions of the social brain demonstrate significant *plasticity*, meaning that their functioning can be affected by experience over time. In contrast to theories that suggest difficulty processing complex information or communicating across large expanses of cortex (Minshew & Williams, 2007), this malleability of the social brain is a positive prognosticator for the development of treatment. The brains of people with ASD are not wired to process optimally social information. But this does not mean that these systems are irretrievably *broken*. Given the observed plasticity of the social brain, remediation of these difficulties may be possible with appropriate and timely intervention.

18.14.23.1 Outside Resources

Web: American Psychiatric Association's website for the 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders*

http://www.dsm5.org

Web: Autism Science Foundation - organization supporting autism research by providing funding and other assistance to scientists and organizations conducting, facilitating, publicizing and disseminating autism research. The organization also provides information about autism to the general public and serves to increase awareness of autism spectrum disorders and the needs of individuals and families affected by autism.

http://www.autismsciencefoundation.org/

Web: Autism Speaks - Autism science and advocacy organization

http://www.autismspeaks.org/

18.14.23.2 Discussion Questions

1. How can neuroimaging inform our understanding of the causes of autism?
2. What are the ways in which neuroimaging, including fMRI and ERP, may benefit efforts to diagnosis and treat autism?
3. How can an understanding of the social brain help us to understand ASD?
4. What are the core symptoms of ASD, and why is the social brain of particular interest?
5. What are some of the components of the social brain, and what functions do they serve?

18.14.23.3 Vocabulary

Endophenotypes

A characteristic that reflects a genetic liability for disease and a more basic component of a complex clinical presentation. Endophenotypes are less developmentally malleable than overt behavior.

Event-related potentials (ERP)

Measures the firing of groups of neurons in the cortex. As a person views or listens to specific types of information, neuronal activity creates small electrical currents that can be recorded from non-invasive sensors placed on the scalp. ERP provides excellent information about the timing of processing, clarifying brain activity at the millisecond pace at which it unfolds.

Functional magnetic resonance imaging (fMRI)

Entails the use of powerful magnets to measure the levels of oxygen within the brain that vary with changes in neural activity. That is, as the neurons in specific brain regions "work harder" when performing a specific task, they require more oxygen. By having people listen to or view social percepts in an MRI scanner, fMRI specifies the brain regions that evidence a relative increase in blood flow. In this way, fMRI provides excellent spatial information, pinpointing with millimeter accuracy, the brain regions most critical for different social processes.

Social brain

The set of neuroanatomical structures that allows us to understand the actions and intentions of other people.

18.14.23.4 References

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Science*, *4*(7), 267–278.

Bentin, S., Allison, T., Puce, A., Perez, E., et al. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8*(6), 551–565.

Brothers, L. (1990). The social brain: A project for integrating primate behavior and neurophysiology in a new domain. *Concepts in Neuroscience*, *1*, 27–51.

Dawson, G., Meltzoff, A. N., Osterling, J., Rinaldi, J., & Brown, E. (1998). Children with autism fail to orient to naturally occurring social stimuli. *Journal of Autism & Developmental Disorders*, *28*(6), 479–485.

Dawson, G., Webb, S. J., & McPartland, J. (2005). Understanding the nature of face processing impairment in autism: Insights from behavioral and electrophysiological studies. *Developmental Neuropsychology*, *27*(3), 403–424.

Elsabbagh, M., Mercure, E., Hudry, K., Chandler, S., Pasco, G., Charman, T., et al. (2012). Infant neural sensitivity to dynamic eye gaze is associated with later emerging autism. *Current Biology*, *22*(4), 338–342.

Geschwind, D. H., & Levitt, P. (2007). Autism spectrum disorders: Developmental disconnection syndromes. *Current Opinion in Neurobiology*, *17*(1), 103–111.

Goren, C. C., Sarty, M., & Wu, P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics, 56*(4), 544–549.

Gottesman I. I., & Shields, J. (1973) Genetic theorizing and schizophrenia. *British Journal of Psychiatry, 122*, 15–30.

Hobson, R. (1986). The autistic child's appraisal of expressions of emotion. *Journal of Child Psychology and Psychiatry, 27*(3), 321–342.

Johnson, M. H. (2006). Biological motion: A perceptual life detector? *Current Biology, 16*(10), R376–377.

Kaiser, M. D., Hudac, C. M., Shultz, S., Lee, S. M., Cheung, C., Berken, A. M., et al. (2010). Neural signatures of autism. *Proceedings of the National Academy of Sciences of the United States of America, 107*(49), 21223–21228.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child, 2*, 217–250.

Klin, A., Lin, D. J., Gorrindo, P., Ramsay, G., & Jones, W. (2009). Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature, 459*(7244), 257–261.

Maestro, S., Muratori, F., Cavallaro, M. C., Pei, F., Stern, D., Golse, B., et al. (2002). Attentional skills during the first 6 months of age in autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*(10), 1239–1245.

McCleery, J. P., Akshoomoff, N., Dobkins, K. R., & Carver, L. J. (2009). Atypical face versus object processing and hemispheric asymmetries in 10-month-old infants at risk for autism. *Biological Psychiatry, 66*(10), 950–957.

McPartland, J. C., Dawson, G., Webb, S. J., Panagiotides, H., & Carver, L. J. (2004). Event-related brain potentials reveal anomalies in temporal processing of faces in autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 45*(7), 1235–1245.

McPartland, J. C., Webb, S. J., Keehn, B., & Dawson, G. (2011). Patterns of visual attention to faces and objects in autism spectrum disorder. *Journal of Autism and Develop Disorders, 41*(2), 148–157.

Minshew, N. J., & Williams, D. L. (2007). The new neurobiology of autism: Cortex, connectivity, and neuronal organization. *Archives of Neurology, 64*(7), 945–950.

Osterling, J., & Dawson, G. (1994). Early recognition of children with autism: A study of first birthday home videotapes. *Journal of Autism and Developmental Disorders, 24*, 247-257.

Osterling, J. A., Dawson, G., & Munson, J. A. (2002). Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation. *Development & Psychopathology, 14*(2), 239–251.

Ozonoff, S., Iosif, A. M., Baguio, F., Cook, I. C., Hill, M. M., Hutman, T., et al. (2010). A prospective study of the emergence of early behavioral signs of autism. *Journal of the American Academy of Child and Adolescent Psychiatry, 49*(3), 256–266.

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism & Developmental Disorders, 32*(4), 249–261.

Schultz, R. T., Gauthier, I., Klin, A., Fulbright, R. K., Anderson, A. W., Volkmar, F., et al. (2000). Abnormal ventral temporal cortical activity during face discrimination among individuals with autism and Asperger syndrome. *Archives of General Psychiatry, 57*(4), 331–340.

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences, 105*(2), 809–813.

de Haan, M., Johnson, M. H., & Halit, H. (2003). Development of face-sensitive event-related potentials during infancy: A review. *International Journal of Psychophysiology, 51*(1), 45–58.

18.14.23.5 Attributions

Adapted by Kenneth A. Koenigshofer, PhD, from Pelphrey, K. A. (2021). Autism: insights from the study of the social brain. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology.* Champaign, IL: DEF publishers. Retrieved from http://noba.to/yqdepwgt

18.14.23.6 **Authors**



- Dr. Kevin Pelphrey is the Harris Professor in the Yale Child Study Center and director of the Center for Translational Developmental Neuroscience and Center for Excellence in Autism Research and Treatment. His research focuses on the application of cognitive neuroscience and genetics to understanding the systems biology of neurodevelopmental disorders.

18.14.23.7 Creative Commons License

Autism: Insights from the Study of the Social Brain by Kevin A. Pelphrey is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Permissions beyond the scope of this license may be available in our Licensing Agreement.

This page titled 18.14: Chapter 14- Mirror Neurons, Theory of Mind, Social Cognition, and Neuroscience of Its Disorders is shared under a mixed 1.3 license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

18.15: Chapter 15- A Gene Essential for Speech

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Explain the first evidence that the FOXP2 gene was involved in human speech
2. Describe evidence that FOXP2 is involved in vocalization in other species
3. Describe the single nucleotide mutation on the FOXP2 gene, and the resulting alteration of the FOXP2 protein, which causes language difficulties
4. Describe the general location of the FOXP2 gene in humans by identifying the chromosome where it is found
5. Identify the differences between the human FOXP2 gene and the same gene in representative mammals and non-human primates

18.15.1 Overview

FOXP2 was the first gene discovered to be essential for human speech. It is expressed in many areas of the brain, including the basal ganglia and inferior frontal cortex, where it is essential for brain maturation and speech and language development (Enard et al., 2002). FOXP2 is found in many vertebrates ranging mouse to alligator and is involved in the vocalizations of many animals. For example, it is found in song birds and is important in their production of birdsong. In bats, FOXP2 is involved in echolocation. The gene has been highly conserved in mammals and few differences in the gene exist among mammal species. The protein that the gene produces is nearly identical in mice and primates.

18.15.2 The FOXP2 Gene

The FOXP2 (short for “forkhead box P2”) gene is the first gene that scientists ever associated with the human ability to speak (Nudel & Newbury, 2013). FOXP2 was first discovered in the KE family (medical name for this British family), about half of whose members had specific language impairments. Towards the end of 1980's, seven children of the family attended a special educational needs unit at a primary school in London. The head of the special needs unit discovered that the family had a speech disorder for three generations. Of the 30 family members, about half suffer from severe language deficiency, some are affected mildly, and few are unaffected (Watkins, et al., 1999). Their faces show rigidity at the lower half, and most cannot complete pronouncing a word. Many of them have severe stuttering and limited vocabulary. In particular, they have difficulty with consonants, and omit them, such as "boon" for "spoon", "able" for "table", and "bu" for "blue." Linguistic deficiency is also noted in written language both in reading and writing. They are characterized by lower nonverbal IQ in addition to their language difficulties. The first scientific report on the family's disorder, by Hurst, et al. (1990), showed that 16 family members were affected by severe abnormality, though their hearing was normal and some had normal intelligence, and that the condition was genetically inherited and was autosomal dominant.



Figure 18.15.1: FOXP2 protein and DNA. **Forkhead box protein P2** (**FOXP2**) is a protein that, in humans, is encoded by the **FOXP2 gene**. FOXP2 is a member of the forkhead box family of transcription factors, proteins that regulate gene expression by binding to DNA. It has multiple effects and is expressed in the brain, heart, lungs and digestive system. Ribbon diagram of forkhead box P2 (FOXP2) protein. (Image from Wikimedia Commons; File:FOXP2 (2as5).png; https://commons.wikimedia.org/wiki/F...XP2_(2as5).png; by SWISS-MODEL, based on 2as5 from PDB; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

Using positron emission tomography (PET) and magnetic resonance imaging (MRI), Vargha-Khadem, et al. (1998) found that some brain regions were underactive (compared to baseline levels) in the KE family members and that some were overactive, when compared to normal people. The underactive regions included motor neurons that control face and mouth regions. The areas that were overactive included Broca's area, the speech center (which might indicate that the speech center is having to work much harder than in normal people in order to produce even faulty speech).

Fisher, et al. (1998) identified the exact location of the gene on the long arm of chromosome 7 (7q31). The chromosomal region (locus) was named *SPCH1* (for speech-and-language-disorder-1), and it contains 70 genes. Using the known gene location of speech disorder from a boy, designated CS, of an unrelated family, they discovered in 2001 that the **main gene responsible for speech impairment** in both KE family and CS was **FOXP2**, and that this gene plays **a major role in** the **origin and development of language** (Lai, et al., 2001). Mutations in the gene result in speech and language problems (Vargha-Khadem & Liegeois, 2007) as seen in the KE family.

The exact problems caused by mutations in this gene remain hard to identify. This is not surprising when you consider the family of genes to which this one belongs. **The FOX family of genes are transcription factors, which means that they produce proteins that can regulate the expression of a number of other genes** by binding directly to their DNA. (The binding ability of these particular proteins comes from their forked shape, from which the gene family gets its name.) The **FOXP2 gene** would appear to play an **important role in orchestrating the establishment of the neural pathways during embryonic development, some of which are required for normal vocalizations in several species** studied, **including humans.** Surprisingly, this gene is extremely **well preserved phylogenetically** (across species): the **protein** that it produces is **almost identical in mice and in primates**, which are separated by some 130 million years of evolution.

The protein that the FOXP2 gene produces in humans differs by only two or three amino acids from the protein that it produces in other species. It is very **likely these two or three amino acids make the difference between animals that cannot**

speak and humans who can. Moreover, the mutations that caused this difference are estimated to have occurred between 100,000 and 200,000 years ago, roughly the time that articulate language may have first emerged in human beings.

By performing a detailed analysis of **the defective FOXP2 gene** sequence in several members of the KE family, scientists were also able to identify the precise site of the mutation that caused this gene to malfunction in these individuals. This **mutation** occurs on exon 14 **of the FOXP2 gene**, when the **guanine in a nucleotide is replaced by an adenine.** As it happens, the part of the gene where this mutation occurs is precisely the one that codes for the "forkhead" portion of the protein—the part that binds to the DNA on other genes. **This change in a single nucleotide on the FOXP2 gene** has a **direct impact on this protein, causing the amino acid arginine to be replaced with a histidine.**

In the hundreds of **normal subjects** tested, **the protein produced by FOXP2 always has an arginine** at this particular site, **while in the members of the KE family who suffered from specific language impairments,** it **always had a histidine.** Hence there is not a shadow of a doubt about the mutation that causes this disorder. That said, it is still amazing to think that the **mutation of a single one of the 2,500 nucleic bases in the FOXP2 gene** is **sufficient to impair language**!

FOXP2 is also required for proper lung development, not only in humans but in other species as well. Additional research showed that Knockout mice with only one functional copy of the FOXP2 gene have significantly reduced vocalizations as pups (Shu et al., 2005). Knockout mice with no functional copies of FOXP2 are runted, display abnormalities in brain regions such as the Purkinje layer, and die an average of 21 days after birth from inadequate lung development.

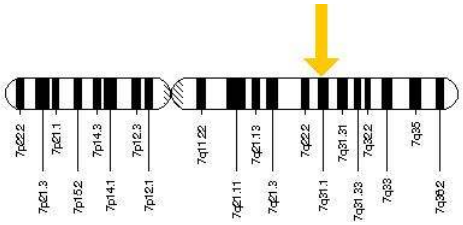


Figure 18.15.2: The FOXP2 gene is located on the long (q) arm of chromosome 7 at position 31. More precisely, the FOXP2 gene is located from base pair 114,086,309 to base pair 114,693,771 on chromosome 7 (Image and caption from Wikimedia Commons; File:FOXP2 location.png; https://commons.wikimedia.org/wiki/F...2_location.png; by U.S. National Library of Medicine; in the public domain).

18.15.3 Evolution

18.15.3.1 The FOXP2 Gene is Found in Many Diverse Species

The **FOXP2** gene is **highly conserved in mammals** (Enard et al., 2002). Conserved, in the evolutionary context, means that a trait or gene and its variants are found in a range of diverse species over long periods of evolutionary time suggesting that once it appeared it was retained as new species diverged from original ancestral species. The human gene differs from that in non-human primates by the substitution of just two amino acids, a threonine to asparagine substitution at position 303 (T303N) and an asparagine to serine substitution at position 325 (N325S) (Preuss, 2012). In mice it differs from that of humans by three substitutions, and in zebra finch by seven amino acids (Enard et al., 2002; Haesler et al., 2004; Teramitsu et al., 2004). One of the two amino acid differences between human and chimps also arose independently in carnivores and bats (Li et al., 2007; Shu et al., 2007). Similar FOXP2 proteins can be found in songbirds, fish, and reptiles such as alligators (Webb et al., 2005; Scharff & Haesler, 2005).

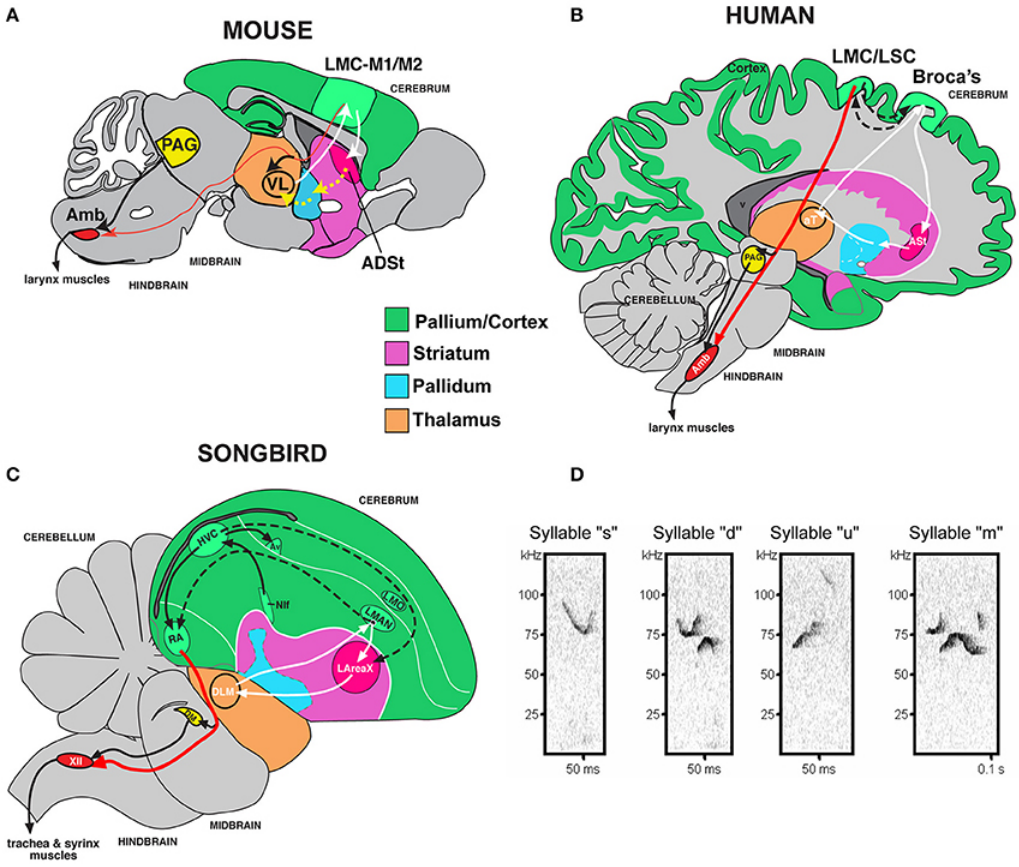


Figure 18.15.3: Mouse song system anatomy and syllable types. **(A)** Proposed anatomy of the rudimentary mouse forebrain vocal communication circuit based on Arriaga et al. (2012). Not shown are other connected brainstem regions, the amygdala, and insula. **(B)** Comparison with human, based on Arriaga et al. (2012) and Pfenning et al. (2014). **(C)** Comparison with songbird. **(D)** Sonograms of examples syllables of the four syllable categories quantified from a C57 male mouse USV song, labeled according to pitch jumps. Anatomical abbreviations: ADSt, anterior dorsal striatum; Amb, nucleus ambiguous; ASt, anterior striatum; aT, anterior thalamus; Av, nucleus avalanche; HVC, a letter-based name; LArea X, lateral Area X; LMO, lateral mesopallium oval nucleus; LMAN, lateral magnocellular nucleus of the nidopallium; LMC, laryngeal motor cortex; LSC, laryngeal somatosensory cortex; M1, primary motor cortex; M2, secondary motor cortex; NIf, interfacial nucleus of the nidopallium; PAG, periaqueductal gray; RA, robust nucleus of the arcopallium; T, thalamus; VL, ventral lateral nucleus of the thalamus; XIIts, 12th vocal motor nucleus, tracheosyringeal part (Figure and caption from Chabout et al., 2016).

18.15.3.2 The FOXP2 Gene Found in Neanderthal Fossils

DNA sampling from *Homo neanderthalensis* bones indicates that their FOXP2 gene is a little different though largely similar to those of *Homo sapiens* (i.e. humans) (Krause et al., 2005; Zimmer, 2016). Previous genetic analysis had suggested that the *H. sapiens* FOXP2 gene became fixed in the population around 125,000 years ago (Benítez-Burraco et al., 2008). Some researchers consider the Neanderthal findings to indicate that the gene instead swept through the population over 260,000 years ago, before our most recent common ancestor with the Neanderthals (Benítez-Burraco et al., 2008). Other researchers offer alternative explanations for how the *H. sapiens* version would have appeared in Neanderthals living 43,000 years ago (Benítez-Burraco et al., 2008).

18.15.3.3 Positive Selection for FOXP2

According to a 2002 study (Enard, et al., 2002), the FOXP2 gene showed indications of recent positive selection (Enard, et al., 2002; Toda et al., 1992). Some researchers have speculated that positive selection is crucial for the evolution of language in humans (Enard, et al., 2002).

The figure below (Figure 14.15.4) shows three types of selection. The top shows positive selection. Positive selection is also known as Darwinian selection and is the type of selection that Darwin envisioned as the primary mechanism of evolution (for more on the

types of selection, see Module 3.5). The original population distribution, before selection, is shown in red. The population distribution after selection is shown in blue. In directional evolution, the population values for the trait move toward one extreme of the trait. By positive selection, a new adaptive trait can sweep through a population. However, "though advantageous mutations are of great interest, they are difficult to detect and analyze because of the fact that neutral and deleterious mutations predominate them in frequency" (Thomas Lab, University of Washington).

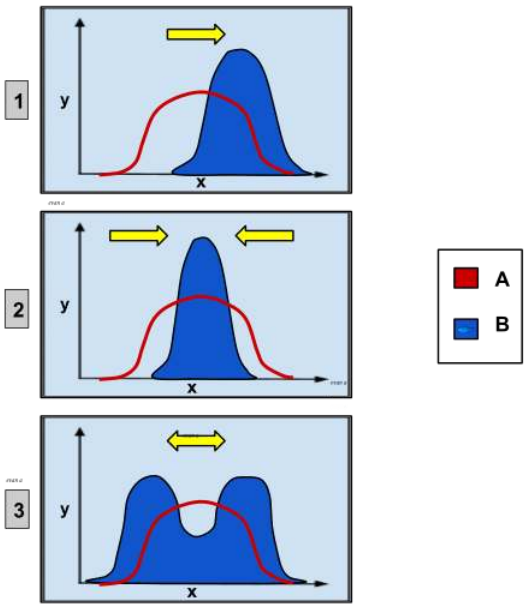


Figure **18.15.4**: These charts depict the different types of genetic selection. On each graph, the x-axis variable is the type of phenotypic trait and the y-axis variable is the number of organisms. Group A is the original population and Group B is the population after selection. Graph 1 shows directional selection, in which a single extreme phenotype is favored. Graph 2 depicts stabilizing selection, where the intermediate phenotype is favored over the extreme traits. Graph 3 shows disruptive selection, in which the extreme phenotypes are favored over the intermediate (Image and caption from Wikimedia Commons; File:Genetic Distribution.svg; https://commons.wikimedia.org/wiki/F...stribution.svg; by Ealbert17; licensed under the Creative Commons Attribution-Share Alike 4.0 International license).

18.15.3.4 Recent Positive Selection for FOXP2 Disputed

Others, however, were unable to find a clear association between species with learned vocalizations and similar mutations in FOXP2 (Webb et al., 2005; Scharff & Haesler, 2005). A 2018 analysis of a large sample of globally distributed genomes confirmed there was no evidence of positive selection, suggesting that the original signal of positive selection may be driven by sample composition (Atkinson et al., 2018; Williams, 2020). Insertion of both human mutations into mice, whose version of FOXP2 otherwise differs from the human and chimpanzee versions in only one additional base pair, causes changes in vocalizations as well as other behavioral changes, such as a reduction in exploratory tendencies, and a decrease in maze learning time. A reduction in dopamine levels and changes in the morphology of certain nerve cells are also observed (Enard et al., 2009).

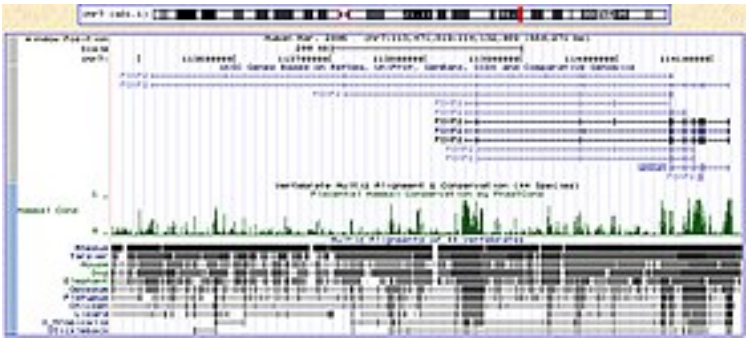


Figure 18.15.5: Human *FOXP2* gene and evolutionary conservation is shown in a multiple alignment (at bottom of figure) in this image from the UCSC Genome Browser. Note that conservation tends to cluster around coding regions (exons).

18.15.3.5 Migration of Human Groups and Genetic and Linguistic Relationships

Around the end of World War II, the Italian population geneticist Luca Cavalli-Sforza began constructing genealogical trees that established relationships among populations throughout the world. By cross-tabulating data on several dozen genes, Cavalli-Sforza established a relationship between American Indians and Asians. This finding is consistent with the most common theory about how the New World was populated: by peoples who crossed from Siberia to Alaska when the Bering Strait was frozen over during the last great Ice Age, some 30,000 years ago.

Cavalli-Sforza's findings assumed even greater importance when he correlated them with analogous studies on languages. When Cavalli-Sforza compared the genealogical trees established by geneticists with those established by linguists, with just a few exceptions, he found that the people who speak each of the 15 major families of languages are genetically related as well. The explanation for this remarkable concordance is that when a population migrates to a new territory, it takes its genes along as well as its language.

But there have been many criticisms of Cavalli-Sforza's approach, and in particular his way of defining a population. In the works of Cavalli-Sforza and his followers, the first step is to define a population, by linguistic criteria among others. Correlations are then established between these populations and their languages, which seems like a dangerously circular approach. It has also been noted that these studies are more convincing on a small scale or a large scale, but far less so on an intermediate one. The reason is that it is easier to distinguish Inuit from Bantu, for example, than to differentiate the various populations that speak Bantu languages.

In addition, the DNA samples used in many studies come from blood banks, and the accompanying records may be biased or false, because for various reasons, when people give blood, they may report their ethnicity as different from what it actually is. This points to the fact that errors in interpretation may be found when controls are applied to both linguistic and genetic data.

18.15.4 Summary

The FOXP2 gene is found widely in the animal kingdom and the human version of the gene differs little from that in other species, yet the human variant appears to have made language possible. Mutation of a single one of the 2,500 nucleic bases in the FOXP2 gene is sufficient to impair language. Though many researchers believe that the FOXP2 gene played an important role in the evolution of human language, evidence for positive selection for it is still controversial.

18.15.5 References

Atkinson EG, Audesse AJ, Palacios JA, Bobo DM, Webb AE, Ramachandran S, Henn BM (September 2018). "No Evidence for Recent Selection at FOXP2 among Diverse Human Populations". *Cell*. **174** (6): 1424–1435.e15

Benítez-Burraco A, Longa VM, Lorenzo G, Uriagereka J (November 2008). "Also sprach Neanderthalis... Or Did She?". *Biolinguistics*. **2** (2): 225–232.

Cavalli-Sforza, L. L. & Edwards, A. W. F. (1963) in Genetics Today, Proceedings of the 11th International Congress of Genetics, The Hague, The Netherlands, ed. Geerts, S. J. (Pergamon, New York), No. 3, pp. 923-933.

Cavalli-Sforza, L. L. (1986) in Evolutionary Perspectives and the New Genetics, eds. Gershowitz, H., Rucknagel, D. L. & Tashian, R. E. (Liss, New York), pp. 13-30.

Cavalli-Sforza, L. L., Piazza, A., Menozzi, P., & Mountain, J. (1988). Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *Proceedings of the National Academy of Sciences*, *85* (16), 6002-6006.

Chabout, J., Sarkar, A., Patel, S. R., Radden, T., Dunson, D. B., Fisher, S. E., & Jarvis, E. D. (2016). A Foxp2 mutation implicated in human speech deficits alters sequencing of ultrasonic vocalizations in adult male mice. *Frontiers in behavioral neuroscience*, *10*, 197.

Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, et al. (August 2002). "Molecular evolution of FOXP2, a gene involved in speech and language" (PDF). Nature. 418 (6900): 869–72.

Enard W, Gehre S, Hammerschmidt K, Hölter SM, Blass T, Somel M, et al. (May 2009). "A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice". *Cell*. **137** (5): 961–71

Fisher, S. E.; Vargha-Khadem, F.; Watkins, K. E.; Monaco, A.P.; Pembrey, M. E. (1998). "Localisation of a gene implicated in a severe speech and language disorder". Nature Genetics. 18 (2): 168–170. doi:10.1038/ng0298-168

Haesler S, Wada K, Nshdejan A, Morrisey EE, Lints T, Jarvis ED, Scharff C (March 2004). "FoxP2 expression in avian vocal learners and non-learners". *The Journal of Neuroscience*. 24 (13): 3164–75.

Hurst, J. A.; Baraitser, M.; Auger, E.; Graham, F.; Norell, S. (1990). "An extended family with a dominantly inherited speech disorder". *Developmental Medicine & Child Neurology*. 32 (4): 352–355. doi:10.1111/j.1469-8749.1990.tb16948.x

Krause J, Lalueza-Fox C, Orlando L, Enard W, Green RE, Burbano HA, et al. (November 2007). "The derived FOXP2 variant of modern humans was shared with Neandertals". *Current Biology*. 17 (21): 1908–12.

Lai, C. S. L.; Fisher, S. E.; Hurst, J. A.; Vargha-Khadem, F.; Monaco, A. P. (2001). "A forkhead-domain gene is mutated in a severe speech and language disorder". *Nature*. 413 (6855): 519–523

Li G, Wang J, Rossiter SJ, Jones G, Zhang S (September 2007). Ellegren H (ed.). "Accelerated FoxP2 evolution in echolocating bats". *PLOS ONE*. **2** (9): e900.

Nudel, R., & Newbury, D. F. (2013). Foxp2. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4* (5), 547-560.

Preuss TM (June 2012). "Human brain evolution: from gene discovery to phenotype discovery". *Proceedings of the National Academy of Sciences of the United States of America*. 109 (Suppl 1): 10709–16.

Scharff C, Haesler S (December 2005). "An evolutionary perspective on FoxP2: strictly for the birds?". *Current Opinion in Neurobiology*. 15 (6): 694–703

Shu, W., Cho, J. Y., Jiang, Y., Zhang, M., Weisz, D., Elder, G. A., ... & Buxbaum, J. D. (2005). Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. *Proceedings of the National Academy of Sciences*, *102*(27), 9643-9648.

Shu W, Lu MM, Zhang Y, Tucker PW, Zhou D, Morrisey EE (May 2007). "Foxp2 and Foxp1 cooperatively regulate lung and esophagus development". *Development*. 134 (10): 1991–2000.

Teramitsu I, Kudo LC, London SE, Geschwind DH, White SA (March 2004). "Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction". *The Journal of Neuroscience*. 24 (13): 3152–63.

Thomas, J.H., Thomas Lab. Genome Sciences, University of Washington; http://depts.washington.edu/jtlab/. Detection and analysis of positive selection. Retrieved 5/11/2022.

Toda M, Okubo S, Ikigai H, Suzuki T, Suzuki Y, Hara Y, Shimamura T (1992). "The protective activity of tea catechins against experimental infection by Vibrio cholerae O1". *Microbiology and Immunology*. 36 (9): 999–1001.

*Vargha-Khadem, F.; Watkins, K. E.; Price, C. J.; Ashburner, J.; Alcock, K. J.; Connelly, A.; Frackowiak, R. S. J.; Friston, K. J.; Pembrey, M. E.; Mishkin, M.; Gadian, D. G.; Passingham, R. E. (1998). "Neural basis of an inherited speech and language disorder". Proceedings of the National Academy of Sciences. **95** (21): 12695–12700. doi:10.1073/pnas.95.21.12695. PMC 22893. PMID 9770548*

Vargha-Khadem, F. & Liegeois, F. (2007). "From speech to gene: The KE family and the FOXP2". In Braten, Stein (ed.). *On Being Moved : From Mirror Neurons to Empathy*. Amsterdam: John Benjamins Publication. p. 11.

Watkins, Kate E; Gadian, David G.; Vargha-Khadem, Faraneh (1999). "Functional and Structural Brain Abnormalities Associated with a Genetic Disorder of Speech and Language". *The American Journal of Human Genetics*. **65** (5): 1215–1221. doi:10.1086/302631. PMC 1288272. PMID 10521285.

Webb, DM, Zhang J (2005). "FoxP2 in song-learning birds and vocal-learning mammals". *The Journal of Heredity*. 96 (3): 212–6.

Williams, S. (2020). *"Language Gene Dethroned". The Scientist Magazine®. Retrieved 5/11/22.*

Zimmer C (17 March 2016). "Humans Interbred With Hominins on Multiple Occasions, Study Finds". The New York Times. Retrieved 17 March 2016.

18.15.6 Attributions

Contributed by Kenneth A. Koenigshofer, PhD. adapted from Genes that are essential for speech by Bruno Dubuc under a Copyleft license.

Some text and images from Wikipedia, KE Family, and FOXP2, retrieved September 4, 2021. Text for section, Evolution, from Wikipedia, FOXP2.

This page titled 18.15: Chapter 15- A Gene Essential for Speech is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

18.16: Chapter 15- The Origins of Language

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Learning Objectives

1. Distinguish between continuity and discontinuity theories of the origins of language; provide examples
2. Distinguish between innate and cultural theories of the origins of language; provide examples
3. Define polygenism and monogenism
4. Define "spandrel" and "exaptation"
5. Contrast Chomsky's views with Deacon's views on the origins of language
6. Briefly describe the evidence related to whether Neanderthals had spoken language

18.16.1 **Overview**

In this section, we review a number of ideas about the origins of human language. Two prominent approaches are continuity and discontinuity theories. Continuity theorists assume that language evolved gradually from earlier forms of communication in non-human animals and hominids from as recently as 40,000 years ago, according to some theorists, to over 2 million years ago in Homo habilis, according to others. Discontinuity theorists believe that human language is so unique that it must have appeared relatively quickly in human evolution without being derived from any form of animal communication. Some theories propose that language is mostly innate, determined primarily by genes, while others hypothesize that human language has cultural origins, resulting from learning in social interaction, with limited and only general contributions from genetics.

18.16.2 **Evolution of Language**

18.16.2.0.1 From Primate Origins to a Language Ready Human Brain

The **origin of language** (spoken, signed, and written) and its relationship to human evolution are complex subjects requiring inferences from the fossil record, archeological evidence, contemporary language similarities and differences, studies of language acquisition, and comparisons between human language and communication in other animals (particularly other primates).

Language clearly depends upon the human brain having acquired features that made it capable of the production and understanding of vocal symbols. However, how human language and a language-ready brain evolved is not known. Nevertheless, most theorists assume that language must have evolved from earlier, more primitive forms of communication.

One integrative approach to the complex issue of the evolutionary origins of human language and a brain capable of human language comes from "comparative neuroprimatology." This is the study of the brains, behaviors and communication systems of monkeys, apes and humans in order to investigate "the biological and cultural evolution of the human language-ready brain" (Arbib et al., 2018, p. 371). The brains of many animals, including non-human primates, show left lateralization of vocalization in the brain, just as the dominant hemisphere for language in most humans is the left. This suggests a long evolutionary history of human language from earlier forms of vocalization.

18.16.3 Theoretical Approaches to the Origins of Language

Approaches to the origin of language can be sub-divided according to some underlying assumptions (Ulbaek, 1998):

- **"Continuity theories"** work from the assumption that language exhibits so much complexity that it could not have developed from nothing in its final form; therefore it must have evolved from earlier pre-linguistic systems among humans' primate ancestors.
- **"Discontinuity theories"** take the opposite approach—that language is such a unique trait that it cannot be compared to anything found among non-humans, and that it must have appeared fairly suddenly during the course of human evolution.
- **Innate theories**: some theories consider language mostly as an innate faculty—largely genetically encoded.
- **Cultural theories**: other theories regard language as a mainly cultural system—learned through social interaction.

18.16.3.1 Continuity Approaches

A majority of linguistic scholars believe continuity-based theories, but they vary in how they hypothesize language development. Among those who consider language as mostly innate, some—notably, Steven Pinker (Pinker & Bloom, 1990)—avoid speculating about specific precursors in nonhuman primates, stressing simply that the language faculty must have evolved in the usual gradual way (Pinker, 1994) as Darwin proposed for most traits. Others in this intellectual camp—notably Ib Ulbæk (1998)—hold that language evolved not from primate communication but from primate cognition, which is significantly more complex.

Those who consider language as learned socially, such as Michael Tomasello, propose that it developed from the cognitively controlled aspects of primate communication, these being mostly gestural as opposed to vocal (Pika & Mitani, 2006; Tomasello, 1996). Regarding the vocal precursors of human language, many continuity theorists hypothesize that language evolved from early human capacities for song (Dunn, et al., 2011; Vaneechoutte, 2014).

18.16.3.2 Discontinuity Approaches

Noam Chomsky, at the Massachusetts Institute of Technology, a proponent of discontinuity theory, argues that a single chance mutation occurred in one individual in the order of 100,000 years ago, installing the language faculty (a hypothetical component of the mind-brain) in "perfect" or "near-perfect" form (Chomsky, 1996). However, it seems unlikely that this would produce adaptive advantage unless a sufficiently large number of others also had similar capacities for communication at the same time.

18.16.3.3 When Did Language Develop?

Clearly, there are many theories about the origins of language, and the dates cited for its first appearance vary greatly from one author to another. They range from the time of **Cro-Magnon man**, about **40,000 years ago**, to the time of **Homo habilis**, about **2 million years back**.

The hypothesis that language dates as far back as the time of Homo habilis is supported by the **resilience of tool cultures** in Homo habilis and later hominid species. Tool making techniques (see module on Material Culture) must be passed from generation to generation to be sustained over long spans of time. This can be accomplished by imitation (younger members of the group learning by watching more skillful and experienced tool makers) or by verbal instruction or by a combination of both. Homo habilis "retained their tool cultures despite many climate change cycles at the timescales of centuries to millennia each, [suggesting that Homo habilis and later] species had sufficiently developed language abilities [including grammar] to verbally describe complete procedures" (Model, 2010, p. 7) for toolmaking. Research with non-human primates shows that toolmaking skills based on imitation alone, without verbal instruction, are lost under environmental changes like the changes in climate referred to above. "Chimpanzees, macaques and capuchin monkeys are all known to lose tool techniques under such circumstances" (Model, 2010, p. 7). Many experts content that the **resilience of tool culture in Homo habilis supports** the view that **language existed in these early human ancestors**.

18.16.3.4 Monogenism or Polygenism?

Regardless of how and when language emerged, another question arises immediately: did it do so once, or many times? In other words, do all languages have a common origin, a proto-language that gave rise to all the rest, or did several different dialects emerge, at various places in the world?

Those who argue for the multiple origins, or **polygenism**, of language, say that the first modern humans did not share the potential for the faculty of speech, and that only after they dispersed through migration did actual languages develop independently among various groups of *Homo sapiens*.

The proponents of polygenism base their arguments on events and behaviors that would have had little chance of occurring without spoken language, such as great migrations that would have required major planning and organizing efforts. From this premise, the polygenists have deduced, for example, that the peoples who left Africa and arrived in Australia about 60,000 years ago must have spoken a complex language before those who migrated to the Middle East.

The alternative view, the theory of **monogenism,** proposes that all languages have a common origin, a proto-language at one location that gave rise to language once, and from that original language all the world's languages developed.

Monogenists were greatly influenced by Meritt Ruhlen's *On the Origin of Languages*, which posited the existence of a single proto-language over 50,000 years ago. Ruhlen's work was based, among other things, on analyses of population genetics that

showed a high correlation between the genetic diversification of human populations and the diversification of the languages that they spoke. But other studies have shown the the correspondences between genetic classifications of populations and genealogical classifications of languages are more uncertain than was once believed. The fact remains that even though Ruhlen's work has been questioned on linguistic grounds, many people still endorse the key idea in his book: that all languages had a common origin. Among these proponents of monogenism, there are two major schools of thought. there are two major schools of thought.

18.16.4 Two Major Views within Monogenism

18.16.4.0.1 A Chance Mutation and Spandrels

Following Chomsky (see above), the first major view starts from the premise that the **human species** as we know it **arose from** an unlikely **genetic mutation** that occurred **about 100,000 years ago**, in which certain of the brain's circuits were reorganized. This reorganization gave rise to the human **"language instinct,"** thus paving the way for the explosive growth in all the cognitive abilities that the powerful communication tool of language provides. On this view, **language** is an **innate** component of human brain organization, which includes a **"universal grammar,"** which all humans inherit within their innate brain organization. This universal grammar is therefore a species-specific human trait and is expressed in similarities in the grammars of all languages of the world. This universal grammar organizes and guides language learning regardless of the human language being acquired. This view makes it hard to imagine any intermediate form of language that could function without all the grammatical structures found in languages today.

This discontinuity view of the origins of language has been criticized by some experts as anti-evolutionist, but several renowned scholars of evolution have ideas consistent with the uniqueness of human language and with discontinuity views of its origins. For example, paleoanthropologist Ian Tattersall, writes that Homo sapiens sapiens "is not simply an improved version of its ancestors—it's a new [development], qualitatively distinct from them." For Tattersall and many other scientists, the mechanism that gave rise to language involved the **relatively sudden combination of pre-existing elements** that had **not** been **selected specifically** to produce this attribute but that, together, made it possible. On this view, characteristics evolved for other purposes make a new capability, like language, possible, so that a trait like language itself was at least initially not selected for by natural selection but rather emerged from other evolved capabilities.

This type of evolutionary mechanism is thought to have come into play many times in the course of evolution; the Harvard paleontologist Stephen Jay **Gould calls it exaptation.** Exaptation means that a trait, feature, or structure that **evolved for one function takes on a different function**; for example, feathers originally evolved to keep ancestral birds warm, but then in later descendants became essential for flight.

Steven Jay Gould calls the features that result from exaptation, such as language, **"spandrels. "** In evolutionary biology, a spandrel is **a phenotypic trait that is a byproduct of the evolution of some other characteristic**, **rather than a direct product of adaptive selection**. These ideas were brought into biology by Steven Jay Gould and Richard Lewontin in a 1979 scientific paper in which they sought to temper the influence of **adaptationism**, the **view that** sees **most organismal traits** as **adaptive products of natural selection**. Gould and Lewontin argued that chance and other non-selection factors played a larger role in evolution than adaptationists claimed. They believed that many traits were evolved for other purposes become "recruited" to perform other unrelated functions during the course of evolution. These recruited "spandrels" are examples of exaptation.

Like Noam Chomsky, Steven Gould also believed that human language is so different from anything else in the animal kingdom that he did not see how it could have developed from ancestral cries or gestures, but he did imagine its having emerged as a side effect of the explosive growth of human cognitive abilities.

18.16.4.0.1 A More Adaptationist View

The **second major school of monogenism** posits a concept of the evolution of Homo sapiens in which **language developed from cognitive faculties** that were **already well established, and that once it was present in some earlier form, it was then naturally selected for**. In this view, the birth of language was **triggered not by a random mutation** (as the first view states), but simply by the availability of an increasingly powerful cognitive tool. Bit by bit, by natural selection, those groups of hominids who developed an articulate **language that let them discuss past and imaginary events** would thereby have gradually supplanted those groups that as yet had only a proto-language. The emphasis here on natural selection for language ability makes this approach more of an adaptationist view.

This second school of monogenism is identified with the linguist and psychologist, **Steven Pinker**, who believes that language may very well have been the target that evolution was "aiming for" (this phrase is not to be taken literally since evolution has no purpose or goal but happens automatically; see Chapter 3; natural selection selected for language ability only in the sense that those who were better at language survived and reproduced more offspring). Pinker argues that the **brain has a general capacity for language**—a concept often associated with connectionist theory in cognitive science (see module on connectionist networks in the chapter on learning and memory). Pinker invokes the **Baldwin effect**, for example, as a major evolutionary force that could have led to modern language (see discussion of Baldwin Effect below). The **ability to learn language would therefore have become a target of natural selection**, thus **permitting the selection of language-acquisition devices that were genetically pre-wired** into the brain's circuits.



Figure **18.16.1**: The Tower of Babel by Pieter Bruegel the Elder (1563). According to the Biblical story of the Tower of Babel, everyone originally spoke the same language, but then God changed things so that everyone spoke different languages. As a result, the tower, which was supposed to reach to the heavens, was never completed, because the people building it could no longer understand one another (Image and caption: public domain via Wikimedia Commons; https://commons.wikimedia.org/wiki/F...t_-_edited.jpg).

This theory of monogenism favored by Steven **Pinker also implies intermediate forms of language** that eventually led to our own. For example, Derek **Bickerton**, a linguist renowned for his work on the evolution of language, suggests that human **language abilities evolved in two stages**. In the **first**, humans would have used a **proto-language** of symbolic representations that took the concrete form of vocal and/or gestural signs. This stage might have lasted nearly 2 million years. Then, **about 50 000 years ago**, humans would have developed a **more formal syntax** that let them exchange ideas with significantly more precision and clarity. With syntax, people could not only label things ("leopard paw print", "danger", etc.), but also join several labels together to express even more meaning ("When you see a leopard paw print, watch out!").

Thus, if **symbolic representations**, already present **in the proto-languages**, made the construction of the **first mental models of reality** possible, it was the emergence of **syntax** that gave human language the great richness that it has today. To give some idea of how the transition from symbolic representations to syntax may have occurred, Bickerton cites the example of the pidgin languages of the colonial period. These rudimentary languages were developed by people of different cultural origins who needed to communicate. Though the pidgin languages themselves had no grammar at all, when they were learned by a second generation, they became what are known as creoles: new, grammatical languages derived from multiple mother tongues.

Another important scholar of the origins of language, **anthropologist Terrence Deacon**, takes exception to the primacy of grammar, believing instead that the **essential feature of language is its use of symbols**. According to Deacon, the so-called symbols that some authors say animals use are actually only indexes. He says that people who try to teach language to chimpanzees

always ensure that the things designated by the words or icons being taught are present in the animal's environment, which makes these words or icons mere indexes. Deacon associates this inferior level of language, based on signs and icons, with that used by children in their earliest years. By contrast, says Deacon, articulate adult language depends on the specificity of the symbols, which in turn depends on the logical connections that **each symbol** in a language **has with the others**. For Deacon, it is this **network of relationships**, far more than the mere occurrence of arbitrary signs, that **characterizes** the **symbols used by human** beings.

Deacon therefore thinks that we must try to **understand the evolution of language not in terms of innate grammatical functions**, **but rather** in terms of the manipulation of symbols and of **relationships among symbols**. There is certainly a human predisposition for language, but this predisposition would be the result of the **co-evolution of the brain and of language**. What is **innate, according to Deacon, is a set of mental abilities** that give us certain natural tendencies, which are expressed in the same universal language structures. Thus Deacon offers a different concept from Chomsky, who associates the origins of universal grammar with a language-specific innovation in the brain.

Deacon sees this co-evolution of the brain and language as being **rooted in the complexity of humans' social lives**, which involved not only a high degree of co-operation between the men and women of a community to acquire resources, but also exclusive monogamous relationships to ensure proper care for very young children who were greatly dependent on adults. This highly explosive mixture is **not found in any other species** (the great apes, for example, gather their food individually). To ensure the stability of the group, rituals and restrictions were required: in other words, abstractions that could be comprehended only if the individuals **involved** could understand and **use symbols**.

18.16.5 Universal Human Language Circuitry

The first changes in the neurons of the left hemisphere that accompanied the development of language faculties during hominization may have occurred about 100,000 years ago, or even earlier. But the truly explosive growth in these faculties most likely began with the evolution of the **angular gyrus**, about 50,000 years ago (see modules 14.10 and 14.11).

Together, the angular and supramarginal gyri constitute a multimodal associative area that receives auditory, visual, and somatosensory inputs. The neurons in this area are thus very well positioned to process the phonological and semantic aspect of language that enables us to identify and categorize objects.

The language areas of the brain are distinct from the circuits responsible for auditory perception of the words we hear or visual perception of the words we read. The auditory cortex lets us recognize sounds, an essential prerequisite for understanding language. The visual cortex, which lets us consciously see the outside world, is also crucial for language, because it enables us to read words and to recognize objects as the first step in identifying them by a name.

Scientists believe that articulate language as we now know it must have already appeared 50,000 or 60,000 years ago, because it was then that the various human ethnic groups became differentiated. But all these groups still retain the ability to learn any language spoken anywhere in the world. Thus a Polish or Chinese immigrant to New York City ends up speaking with a New York accent, and vice versa, which just goes to show that all of us have inherited the same linguistic potential.

18.16.6 Pidgin

A **pidgin** is a language created spontaneously from a mixture of several languages, so that the people who speak them can communicate. The people who develop a pidgin language agree on a limited vocabulary and employ only a rudimentary grammar. For example, in Franco-Vietnamese pidgin, this results in sentences such as "Moi faim. Moi tasse. Lui aver permission repos. Demain moi retour campagne." [Me hunger. Me lie down. He have permission rest. Tomorrow me return country.]

The first documented pidgin, the Lingua Franca, was used by Mediterranean merchants in the Middle Ages. Another well known pidgin was developed from a mixture of Chinese, English, and Portuguese to facilitate trade in Canton, China during the 18th and 19th centuries. Another classic example is the pidgin developed by slaves in the Caribbean, whose cultural origins were too diverse for their own languages to survive after their forced transplantation.

Children who grow up together and learn a pidgin tend to spontaneously impose a lexical structure on it to create a creole: a true language whose vocabulary comes from other languages. But this does not happen with all pidgins, and some are lost or become

obsolete.

According to researchers such as Derek Bickerton, people who find themselves in the particular circumstances described above revert to an older form of communication, what Bickerton calls a proto-language, of which pidgin would be the modern manifestation.

18.16.7 Baldwin Effect

In 1896, American psychologist James Mark Baldwin proposed an evolutionary mechanism that soon came to be known as the "**Baldwin effect**". It is a process whereby a **behavior that originally had to be learned can eventually become innate**, that is, fixed in the genetic programming of the species concerned (Sznajder, et al., 2012). The effectiveness of the learning plays a key role in the Baldwin effect, which **distinguishes it from Lamarckian inheritance of acquired characteristics.**

The idea behind the Baldwin effect is that individuals who are able to learn a given kind of behavior more effectively may over the course of their lives acquire advantages that individuals whose brains are less plastic will not. Natural selection will therefore tend to favor those who always learn faster until, at some point in evolution, the behavior will no longer need to be learned at all: it will have become instinctive.

It should be noted that the Baldwin effect assumes that the environment remains relatively stable, because if it changed too much, there would be no selection against plasticity, which would become an important adaptive factor. But if the environment remains stable for a long time, natural selection may favor a mutation that makes the behavior innate and hence more robust and efficient.

The Baldwin effect, as an evolutionary mechanism that targets learning abilities, has been successfully simulated with many computer programs. Many scientists believe that it may have played a decisive role in the evolution of language, nevertheless the existence of the effect is controversial and many evolutionary biologists dismiss the idea (French & Messinger, 1994).

18.16.8 Neanderthal man

It was long believed that Neanderthal man could not communicate verbally—that Neanderthals must have had some primitive form of language, but could not produce the complete range of sounds of human language. According to a hypothesis advanced by American linguist Philip Lieberman, Neanderthals' larynxes had not yet descended so low as those of Homo sapiens, so they would have had a great deal of difficulty in pronouncing the three main vowels present in the majority of the world's languages (ee as in "beet", oo as "boot" and a as in "aha!").

However, some authors argue that to speak a rudimentary language, one need not master all of the vowels, so long as the language has a sufficient number of consonants.

Moreover, recent research has raised questions about Lieberman's hypothesis. Many researchers find it hard to believe that Neanderthals, who produced sophisticated tools, adorned their bodies with bracelets and necklaces, buried their dead, and produced works of art, had little or no ability to communicate verbally.

Some authors even believe that the skull on which Lieberman based his work was not truly representative of Neanderthal man. Contrary to his findings, reconstructions of other Neanderthal skulls have shown that their base would have allowed the existence of a vocal tract very similar to that of modern humans. For example, the discovery in 1989 of the 60,000-year-old skull of a male Neanderthal with a hyoid bone (the bone that supports the larynx) even led some researchers to say that he had probably been able to speak.

One thing is certain: Neanderthals disappeared about 28,000 years ago, leaving the Earth to their rivals, Homo sapiens sapiens, who had everything they needed to use an articulate symbolic language with elaborate syntax. We should not discount the possibility that Neanderthals also had developed the ability to speak and that language may have played a significant role in their lives as well.

18.16.9 Summary

It is assumed by most theorists that language must have evolved from earlier, more primitive forms of communication such as pre-linguistic systems used by our primate ancestors. Some theorists such as Steve Pinker consider language to be mostly innate. Explosive growth of language capacities in humans, according to some theorists, began about 50,000 years ago with the evolution

of the angular gyrus, a multimodal area of cortex. It is speculated that human language must have evolved at least 50,000 to 60,000 years ago when human ethnic groups differentiated yet all retained the ability to learn any human language with ease that they are exposed to in their early experience.

18.16.10 References

Arbib, M. A., Aboitiz, F., Burkart, J. M., Corballis, M., Coudé, G., Hecht, E., ... & Wilson, B. (2018). The comparative neuroprimatology 2018 (CNP-2018) road map for research on How the Brain Got Language. *Interaction Studies*, *19* (1-2), 370-387.

Chomsky, N, 1996. *Powers and Prospects. Reflections on human nature and the social order*. London: Pluto Press, p 30.

Dunn, M.; Greenhill, SJ.; Levinson, SC.; Gray, RD. (May 2011). "Evolved structure of language shows lineage-specific trends in word-order universals". *Nature*. **473** (7345): 79–82.

French, R. M., & Messinger, A. (1994). Genes, Phenes and the Baldwin Effect: Learning and Evolution in. In *Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems* (Vol. 4, p. 277). MIT Press.

Pika, S. & Mitani, J. (2006). "Referential gestural communication in wild chimpanzees (*Pan troglodytes*)". *Current Biology*. **16** (6): R191–R192. doi:10.1016/j.cub.2006.02.037.

Pinker, S. & Bloom, P. (1990). "Natural language and natural selection". *Behavioral and Brain Sciences*. **13** (4): 707–784. doi:10.1017/S0140525X00081061. S2CID 6167614.

Pinker, S. (1994). *The language instinct*. New York: W. Morrow and Co.

Sznajder, B., Sabelis, M. W., & Egas, M. (2012). How adaptive learning affects evolution: reviewing theory on the Baldwin effect. *Evolutionary biology*, *39*(3), 301-310.

Tomasello, Michael (1996). B M Velichkovskiĭ; Duane M Rumbaugh; Universität Bielefeld Zentrum für Interdisziplinäre Forschung (eds.). *The cultural roots of language*. *Communicating meaning : the evolution and development of language*. Mahwah, N.J.: L. Erlbaum

Ulbæk, I. (1998). James R Hurford; Michael Studdert-Kennedy; Chris Knight (eds.). *The origin of language and cognition*. *Approaches to the evolution of language : social and cognitive base*. Cambridge, UK; New York: Cambridge University Press. pp. 30–43.

Vaneechoutte, M. (2014). "The Origin of Articulate Language Revisited: The Potential of a Semi-Aquatic Past of Human Ancestors to Explain the Origin of Human Musicality and Articulate Language" (PDF). *Human Evolution*. **29**: 1–33.

18.16.11 Attributions

Adapted by Kenneth A. Koenigshofer, PhD., from The origins of language by Bruno Dubuc, The Brain from Top to Bottom, under a Copyleft license. Some text also adapted from: Model, E. P. (2010, 2020). Origin of Language; usilacs.org. http://usilacs.org/wp-content/upload...-Wikipedia.pdf; licensed under the Creative Commons Attribution-ShareAlike License; retrieved 4/2/2022.

This page titled 18.16: Chapter 15- The Origins of Language is shared under a mixed license and was authored, remixed, and/or curated by Kenneth A. Koenigshofer (ASCCC Open Educational Resources Initiative (OERI)) .

Index

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

1 A

ADH

12.1: Maintaining Water Balance

aldosterone

12.1: Maintaining Water Balance

androgen

12.1: Maintaining Water Balance

antidiuretic hormone

12.1: Maintaining Water Balance

2 C

corticosteroid

12.1: Maintaining Water Balance

cortisol

12.1: Maintaining Water Balance

3 D

diabetes insipidus

12.1: Maintaining Water Balance

diabetes mellitus

12.1: Maintaining Water Balance

4 E

Electroencephalography

2.3: Non-Invasive Techniques: Direct Functional Imaging Techniques

5 G

glucagon

12.4: Digestion and the Hormonal Control of Metabolism

gluconeogenesis

12.4: Digestion and the Hormonal Control of Metabolism

glycogenolysis

12.4: Digestion and the Hormonal Control of Metabolism

6 H

hormonal regulation

12.1: Maintaining Water Balance

hyperglycemia

12.4: Digestion and the Hormonal Control of Metabolism

hypoglycemia

12.4: Digestion and the Hormonal Control of Metabolism

7 I

insulin

12.4: Digestion and the Hormonal Control of Metabolism

8 M

magnetoencephalography

2.3: Non-Invasive Techniques: Direct Functional Imaging Techniques

mineralocorticoid

12.1: Maintaining Water Balance

9 N

norepinephrine

12.1: Maintaining Water Balance

10 O

osmoreceptor

12.1: Maintaining Water Balance

11 R

renin

12.1: Maintaining Water Balance

Glossary

This page is a draft and under active development. Please forward any questions, comments, and/or feedback to the ASCCC OERI (oeri@asccc.org).

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: Biopsychology (OERI) - DRAFT for Review

Webpages: 137

Applicable Restrictions: Noncommercial

All licenses found:

- Unknown License: 91.2% (125 pages)
- Undeclared: 2.2% (3 pages)
- CC BY 4.0: 2.2% (3 pages)
- CC BY-SA 4.0: 2.2% (3 pages)
- CC BY-NC-SA 4.0: 1.5% (2 pages)
- CC BY-NC-SA 3.0: 0.7% (1 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)

- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)

- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)
- Unknown License: 0% (0 page)

By Page

- Biopsychology (OERI) - DRAFT for Review - *Unknown License*
  - Front Matter - *Unknown License*
    - TitlePage - *Unknown License*
    - InfoPage - *Unknown License*
    - Table of Contents - *Undeclared*
    - About the Authors - *Unknown License*
    - Licensing - *Undeclared*
  - 1: Biopsychology as a Course of Study - *Unknown License*
    - 1.1: What is Biopsychology? - *CC BY-NC-SA 3.0*
    - 1.2: History of Biopsychology - *CC BY-NC-SA 4.0*
    - 1.3: Connecting Biology to Behavior - *CC BY 4.0*

- 1.4: What are the Divisions of Biopsychology? - *Unknown License*

- 2: Research Methods and Ethical Considerations of Biological Psychology and Neuroscience - *Unknown License*
  - 2.1: Introduction to Research in Biological Psychology - *Unknown License*
  - 2.2: Non-Invasive Techniques: Structural Imaging - *CC BY-SA 4.0*
  - 2.3: Non-Invasive Techniques: Direct Functional Imaging Techniques - *CC BY-NC-SA 4.0*
  - 2.4: Non-Invasive Techniques: Indirect Functional Imaging Techniques - *Unknown License*
  - 2.5: Other Non-Invasive Techniques - *CC BY-SA 4.0*
  - 2.6: Invasive Techniques - *CC BY 4.0*
  - 2.7: Ethical Considerations - *CC BY-SA 4.0*
- 3: Evolution, Genes, and Behavior - *Unknown License*
  - 3.1: Evolution of Minds and Behavior - *CC BY 4.0*
  - 3.2: Evolution - Mechanisms and Evidence - *Unknown License*
  - 3.3: Human Evolution - *Unknown License*
  - 3.4: Evolutionary Theories in Psychology - *Unknown License*
  - 3.5: DNA, Chromosomes, and Gene-Environment Correlations - *Unknown License*
  - 3.6: The Nature-Nurture Question - *Unknown License*
  - 3.7: Epigenetics in Psychology - *Unknown License*
  - 3.8: Human Genetics - *Unknown License*
  - 3.9: Genes and Behavior - *Unknown License*
- 4: Nervous System Anatomy - *Unknown License*
  - 4.1: Nervous System Tissue - *Unknown License*
  - 4.2: Nervous System Structure and Terminology - *Unknown License*
  - 4.3: Nervous System Development - *Unknown License*
  - 4.4: The Central Nervous System - *Unknown License*
  - 4.5: The Peripheral Nervous System - *Unknown License*
  - 4.6: Communication and the Endocrine System - *Unknown License*
- 5: Communication within the Nervous System - *Unknown License*
  - 5.1: Neurons and their Basic Functions - *Unknown License*
  - 5.2: Neurons Generate Voltage Changes to Code Information - *Unknown License*
  - 5.3: Synaptic Transmission - *Unknown License*
- 6: The Effects of Psychoactive Drugs - *Unknown License*
  - 6.1: Psychopharmacology and Psychoactive Drug Classification - *Unknown License*
  - 6.2: Mechanisms of Psychoactive Drug Action - *Unknown License*
  - 6.3: Substance Use and Addiction - Theories, Brain Mechanisms, and Treatment - *Unknown License*
  - 6.4: Specific Drug Effects, Misuse, and Risks - *Unknown License*
  - 6.5: Psychotherapeutics - *Unknown License*
- 7: Senses - *Unknown License*
  - 7.1: Senses Overview - *Unknown License*
  - 7.2: Vision - *Unknown License*
  - 7.3: Hearing - *Unknown License*
  - 7.4: Touch - *Unknown License*
  - 7.5: Taste and Smell - *Unknown License*
- 8: Sensation and Perception - *Unknown License*
  - 8.1: Sensation vs Perception - *Unknown License*
  - 8.2: Psychophysical Methods - *Unknown License*
  - 8.3: Perception and Action - *Unknown License*
  - 8.4: Gestalt Principles of Perception - *Unknown License*
- 9: Movement - *Unknown License*
  - 9.1: Classification of Muscle Types and Functions - *Unknown License*
  - 9.2: Synaptic Control of Movement- Neuromuscular Junction and Motor Units - *Unknown License*
  - 9.3: Spinal Reflexes - *Unknown License*
  - 9.4: Motor Circuitry- Neural Structures and Pathways - *Unknown License*
  - 9.5: Impact of Disease States and Neural Damage on Motor Control - *Unknown License*
- 10: Learning and Memory - *Unknown License*
  - 10.1: Learning, Genes, and Adaptation - *Unknown License*
  - 10.2: Types of Learning and Biological Adaptation - *Unknown License*
  - 10.3: Introduction to Memory and the Brain - *Unknown License*
  - 10.4: Synaptic Mechanisms of Learning and Memory - *Unknown License*
  - 10.5: Brain Mechanisms of Memory Disorders - *Unknown License*
  - 10.6: Localization of Types and Stages of Memory - *Unknown License*
  - 10.7: Hormones, Learning, and Memory - *Unknown License*
- 11: Wakefulness and Sleep - *Unknown License*
  - 11.1: Consciousness - *Unknown License*

- 11.2: What is Sleep? - *Unknown License*
- 11.3: Stages of Sleep - *Unknown License*
- 11.4: The Brain and Sleep - *Unknown License*
- 11.5: Theories of Sleep - *Unknown License*
- 11.6: Sleep Disorders - *Unknown License*

- 12: Ingestive Behaviors - Eating and Drinking - *Unknown License*
  - 12.1: Maintaining Water Balance - *Unknown License*
  - 12.2: Triggering Drinking Behavior - Osmometric and Volumetric Thirst - *Unknown License*
  - 12.3: Ingestive Behavior Basics - *Unknown License*
  - 12.4: Digestion and the Hormonal Control of Metabolism - *Unknown License*
  - 12.5: Hunger - Theories, Detectors, and the Hypothalamus - *Unknown License*
  - 12.6: Ingestive Behavior in Context - *Unknown License*
- 13: Sexuality and Sexual Development - *Unknown License*
  - 13.1: Sex Hormones, Sexual Motivation, and Reproductive Behaviors - *Unknown License*
  - 13.2: Sexual Development, Anatomy and Physiology - *Unknown License*
  - 13.3: Differences of Sexual Development - *Unknown License*
  - 13.4: Gender - *Unknown License*
  - 13.5: Sexual Orientation - *Unknown License*
  - 13.6: Sex Differences and Evolutionary Interpretations - *Unknown License*
- 14: Intelligence and Cognition - *Unknown License*
  - 14.1: Overview of Brain Mechanisms in Intelligence, Language, and Cognition - *Unknown License*
  - 14.2: Cognition, Intelligence, and Language as Psychological Adaptations - *Unknown License*
  - 14.3: High Intelligence-Nature or Nurture; Learning Disabilities - *Unknown License*
  - 14.4: Cognition- Categories, Concepts, Schemas and the Brain - *Unknown License*
  - 14.5: Brain Mechanisms and Intelligence - *Unknown License*
- 15: Language and the Brain - *Unknown License*
  - 15.1: Introduction to Language - *Unknown License*
  - 15.2: Broca's Area, Wernicke's Area, and Additional Language-Processing Areas in the Brain - *Unknown License*
  - 15.3: Other Brain Models of Spoken and Written Language - *Unknown License*
  - 15.4: Handedness, Language, and Brain Lateralization - *Unknown License*
  - 15.5: The Right Hemisphere's Contribution to Language - *Unknown License*
- 16: Emotion and Stress - *Unknown License*
  - 16.1: The Neurological Bases of Emotions - *Unknown License*
  - 16.2: Valence of Emotions - *Unknown License*
  - 16.3: Fight or Flight - *Unknown License*
  - 16.4: Social Neuroscience - *Unknown License*
  - 16.5: Stress and health - *Unknown License*
- 17: Biological Bases of Psychological Disorders - *Unknown License*
  - 17.1: Biological Factors in Psychological Disorders- An Introduction - *Unknown License*
  - 17.2: Biological Basis of Mood Disorders- Major Depressive and Bipolar Disorders - *Unknown License*
  - 17.3: Biological Basis of Anxiety Disorders - *Unknown License*
  - 17.4: Biological Basis of Schizophrenia Spectrum Disorder - *Unknown License*
- 18: Supplemental Content - *Unknown License*
  - 18.1: Chapter 3- Population Evolution (and the Modern Synthesis) - *Unknown License*
  - 18.2: Chapter 3- Phenotypic Variation and Population Genetics - *Unknown License*
  - 18.3: Chapter 3- Adaptive Evolution - *Unknown License*
  - 18.4: Chapter 3- Modern Primates and Their Evolution - *Unknown License*
  - 18.5: Chapter 3- Material Culture - *Unknown License*
  - 18.6: Chapter 10- Computer Models and Learning in Artificial Neural Networks - *Unknown License*
  - 18.7: Chapter 10- Acetylcholine, Enriched Experience, and Memory - *Unknown License*
  - 18.8: Chapter 10- Memory Types and Stages - *Unknown License*
  - 18.9: Chapter 10- The Psychology of Memory Encoding, Storage, and Retrieval - *Unknown License*
  - 18.10: Chapter 10- Psychological Factors Affecting Learning, Memory, and Forgetting - *Unknown License*
  - 18.11: Chapter 14- Traditional Models of Human Intelligence - *Unknown License*
  - 18.12: Chapter 14- Problem Solving, Categories and Concepts - *Unknown License*
  - 18.13: Chapter 14- The Brain and Evolution of Cognition - *Unknown License*

- 18.14: Chapter 14- Mirror Neurons, Theory of Mind, Social Cognition, and Neuroscience of Its Disorders - *Unknown License*
- 18.15: Chapter 15- A Gene Essential for Speech - *Unknown License*
- 18.16: Chapter 15- The Origins of Language - *Unknown License*

- Back Matter - *Unknown License*
  - Index - *Unknown License*
  - Glossary - *Unknown License*
  - Detailed Licensing - *Undeclared*