OPEN NEUROSCIENCE INITIATIVE

Austin Lim, PhD

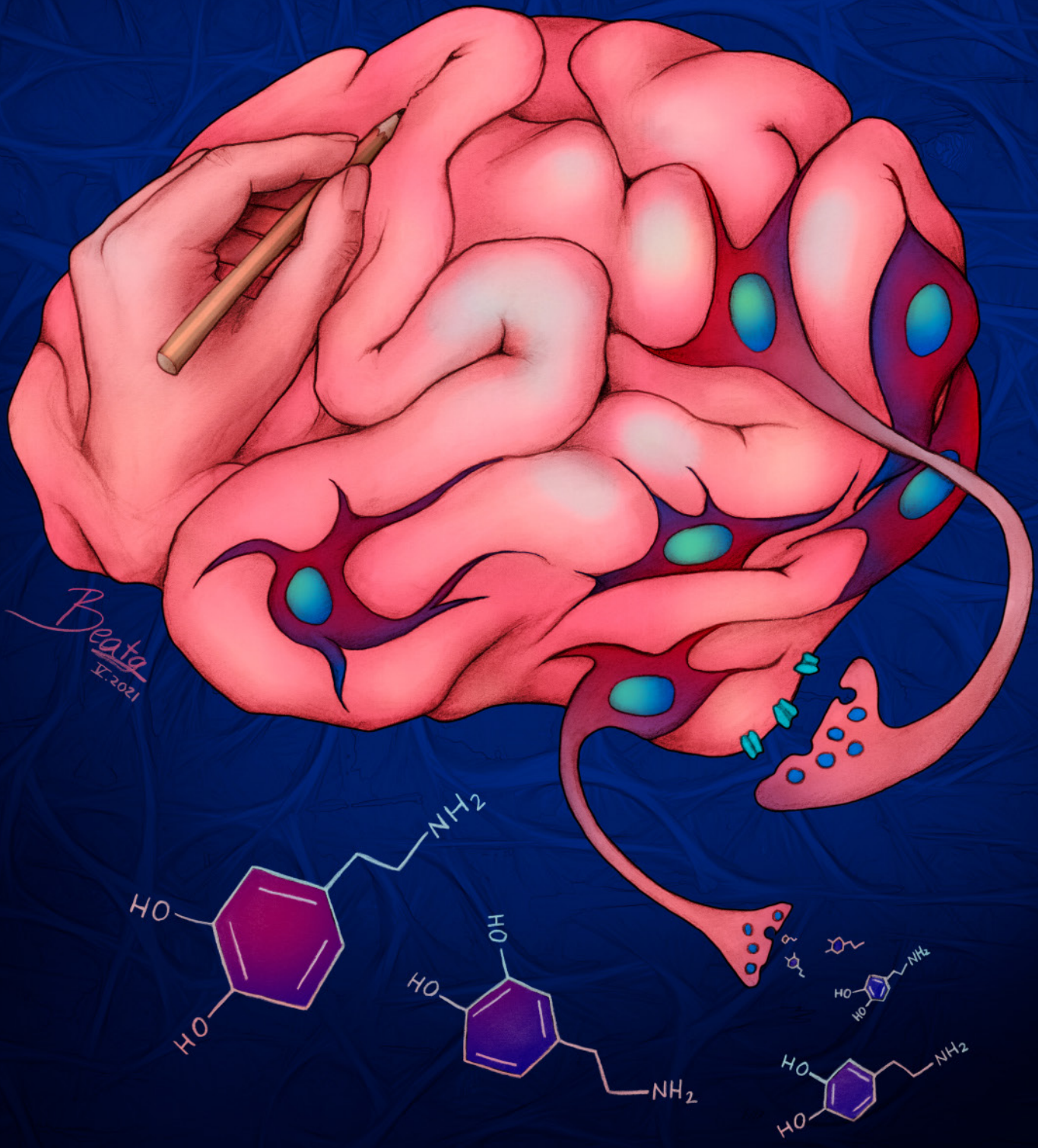


Illustration by Beata Science Art

Open Neuroscience Initiative

First Edition
2021

Table of Contents

Preface

For years, I had used old college textbooks to prop up my computer monitor. Of course I gained valuable information from these texts at one point. But every time I sat down to work, I was reminded of just how expensive college can be.

This project began as a means to overcoming the financial burden that face undergraduate neuroscience students when buying textbooks. By compiling and writing a completely free-to-access textbook that covers the foundations of a typical college introduction to neuroscience course, students would have one less obstacle to overcome in their educational career, allowing them to focus their valuable time and attention on learning rather than finances.

To make this project a reality, I began with a humble tweet in May 2019 that managed to gain a tiny bit of traction among the neuroscience Twitter-sphere. A handful of retweets and comments later, several experts who shared my ideology for open educational resources contacted me directly, looking for ways to be involved in the project. Ultimately, almost half of the contributing authors or content expert editors were recruited through Twitter, while the other half were personal connections I had made through my time as a neuroscientist.

My choice to distribute this text freely over the internet was deliberate. We have been living in a digital era where most (if not all) of us have completely unfettered access to the entirety of the world's knowledge using a desktop, laptop, or cell phone. Why shouldn't a formal neuroscience education be any different?

My hope for this project is to share my passion of neuroscience with the world. Additionally, widespread adoption of open educational resources like this means increased access and decreased inequity across the realm of higher education, opening the doors for a more inclusive and diverse future.

Note: I did not receive any financial compensation through the writing of this textbook, and I am hosting the complete document on my personal website. If you would like to support this endeavor and future projects directly, you can send any donations to the following bitcoin address:

3NNosyLrZEVf6yoB7uEsj37xoCCiEKuEFZ

I acknowledge a major limitation present in all textbooks: that science is rapidly-changing and ever-evolving. To our knowledge, the information is current as of 2021. The authors have worked with the editors and a growing team of researchers to verify that the statements made here are backed by peer-reviewed studies. However, I would encourage that additional sources be used to verify the most up-to-date numbers or statistics and the current state of knowledge.

Acknowledgements

I would like to thank the following contributing authors for lending their time, energy, and expertise to this project:

Dr. Ben Marcus
Dr. Andrea Gaede
David Graykowski
Dr. Alexander Rajan
Claire Sun

Additionally, I am very grateful for the valuable editorial comments provided by the following experts:

Dr. Harriet de Wit
Dr. Alexandrina Guran
Dr. Monica Javidnia
Dr. Helena Ledmyr
Dr. Ben Marcus
Dr. Allison Morales
Dr. Alexander Rajan
Dr. Dana Simmons
Dr. Sandra Virtue
Dr. Katie Willis

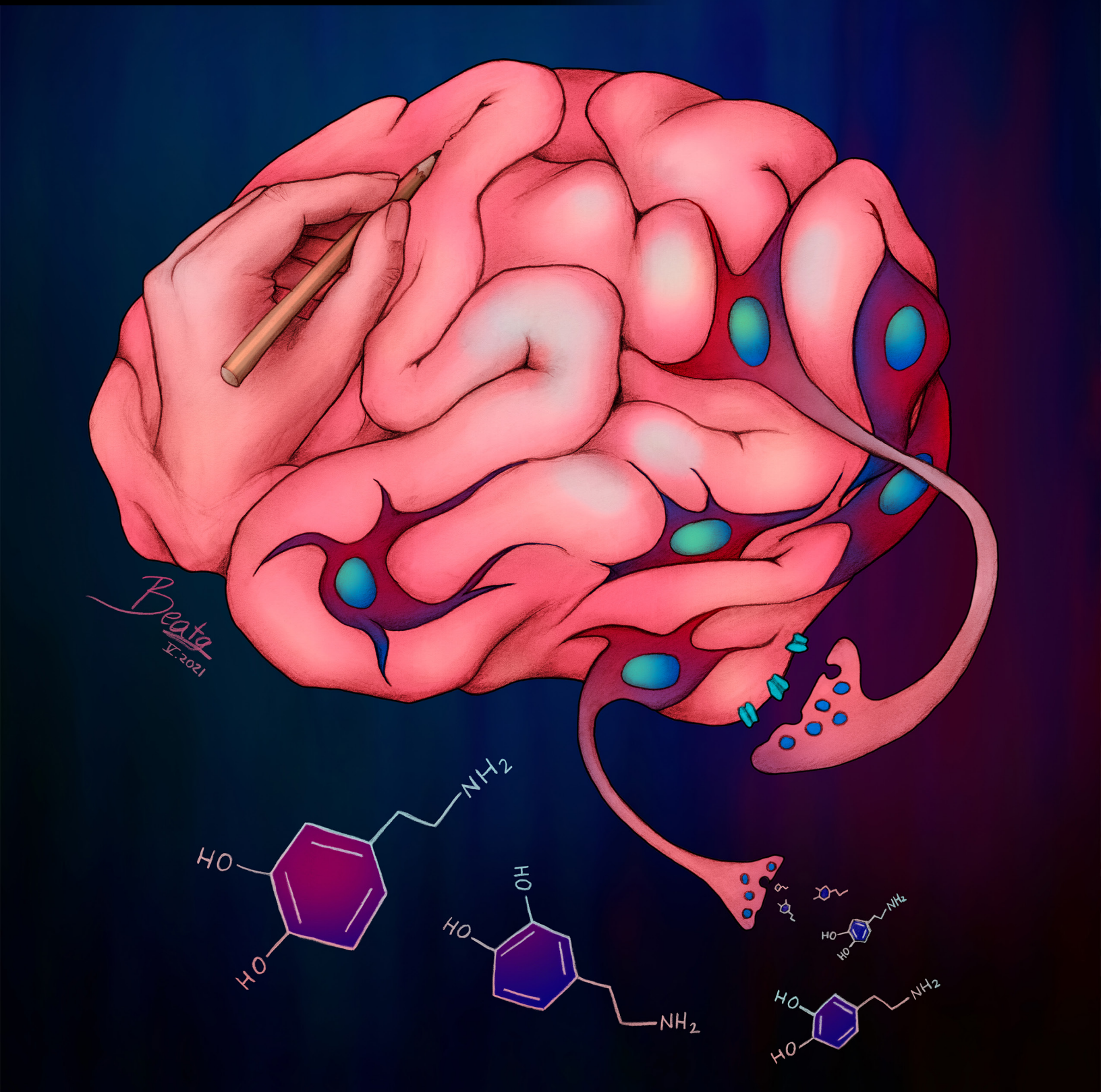
A big thank you to Dr. Beata Mierzwa, the artist who designed the cover image. This illustration demonstrates the wide range of topics that neuroscientists concern themselves with, spanning the level of molecules, to neurons, all the way up to complex behaviors such as drawing. More of her work can be found online (http://www.beatascienceart.com/).

This project was supported by the Vincentian Endowment Fund Grant through DePaul University.

Finally, thank you to my parents for their limitless love and support, to my friends who regularly endure my frequent conversation interruptions with neuroscience facts, and to my wife for her patience through this three-year long project.

Chapter 1:

Introduction



Austin Lim, PhD (DePaul University)

Editor: Dana Simmons, PhD (University of Chicago)

1.1 What is Neuroscience?

In short, neuroscience is the study of the nervous system, the collection of nerve cells that interpret all sorts of information, which allows the body to coordinate activity in response to the environment.

The study of neuroscience has taught us that the brain is a complicated organ with several connection routes, both between different bodily organs and within itself. Some of those connections communicate information down towards the body, such as signals that allow us to control the movements of our muscles, or to change the activity of our internal organs. Other connections ascend into the brain, conveying all sorts of information from the world around us into a representation of our surroundings. Still, other routes communicate between brain areas, such as when the sudden detection of a threat passes through our visual system and turns into a "get ready" signal that then prepares the rest of our body for conflict. Because of this complex system of communication, the nervous system can be thought of as a series of highways and roads that connect different cities (organs.)

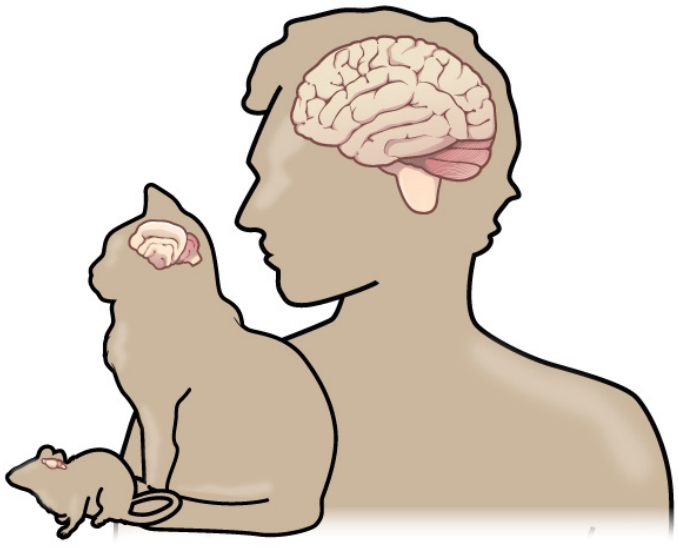


Figure 1.1 Neuroscience is the study of the nervous system and the way the brain performs its many functions.

The nervous system conveys all of these different types of information using a combination of electrical and chemical signals. The main active cellular units of the nervous system, the **neurons**, are highly sensitive to changes in their environment. Similar in the way that computers do all their work using a highly coordinated binary signal of 0s and 1s, the electrical output of many neurons is an all-or-nothing response called an **action potential**. A wide variety of chemicals called **neurotransmitters** is responsible for passing information between neurons.

The brain is the main computational powerhouse of the body, much more complex and

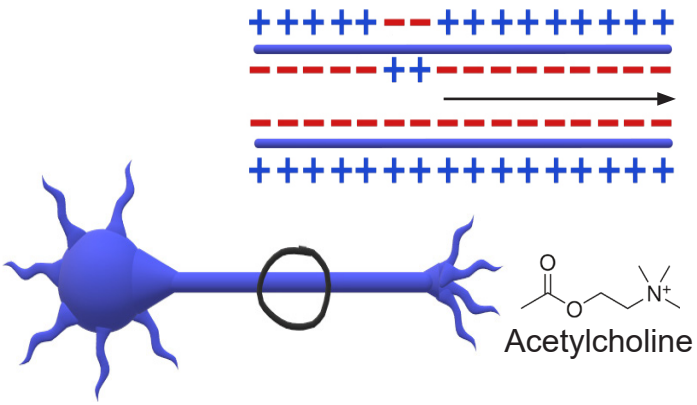


Figure 1.2 Neurons temporarily change their electric properties during an action potential, which allows for the release of neurotransmitters.

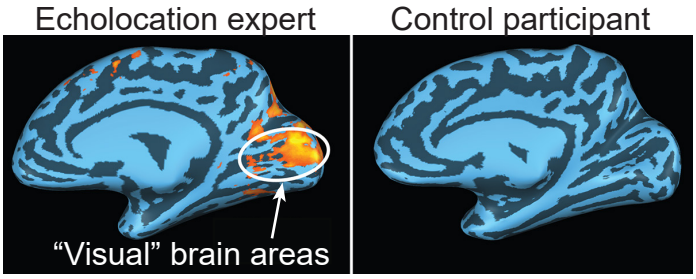


Figure 1.3 The brain is remarkably flexible, and blind people can use visual areas during echolocation (left, activity in yellow and orange).

intricate than any artificially created system thus far. Estimations of the computational power of the brain suggest that we can handle somewhere around 10^28 operations per second, processing power that is many orders of magnitude faster than any supercomputer to date. While it's true that computers can do large, mathematical calculations that (most) humans can't, the real strength of our brain is its flexibility: brains are capable of changing and adapting to a wide variety of circumstances. Blind people use their visual areas of the brain while echolocating , stroke survivors can regain lost motor functions using the unaffected brain circuits, and babies can effortlessly learn two languages simultaneously in a bilingual household.

The brain is also responsible for the most abstract and unique human functions, such as the origin of consciousness, the place where our thoughts, fears, and desires are born, and the endless creativity of our species. All the musical works of Mozart, the literary genius of Shakespeare, and the philosophical theories of Aristotle were produced through some complex interaction of neurons that we may never understand.

1.2 How do we learn about Neuroscience?

Generally speaking, there are three main research designs that have been used - and will likely continue to be used - to learn about the brain.

Experimental design

The gold standard in science is the use of **experimental design**. In an experiment, the scientist uses a stepwise process of developing a research question and hypothesis, then answering that question by performing tests. The main goal of an experiment is to establish a causal relationship between one factor that is being changed, the **independent variable**, and the factor that is influenced, the **dependent variable**. A well-designed experiment has variables that are carefully controlled, which minimizes the influence of extraneous variables, often called **confounding variables**. The influence of confounding variables can be eliminated by comparing the experimental group with a **control group**, a group that is as similar as possible in every way except for the manipulation of the independent variable. Importantly, subjects or patients are generally assigned to the experimental or control group at random.

For example, consider the research question: "Does studying more increase performance on exams?" Here, the independent variable is the number of hours studied, while the dependent variable is the grade on the exam. A potential confounding variable could be the number of hours slept before the exam, since poor sleep causes poor memory recall performance,

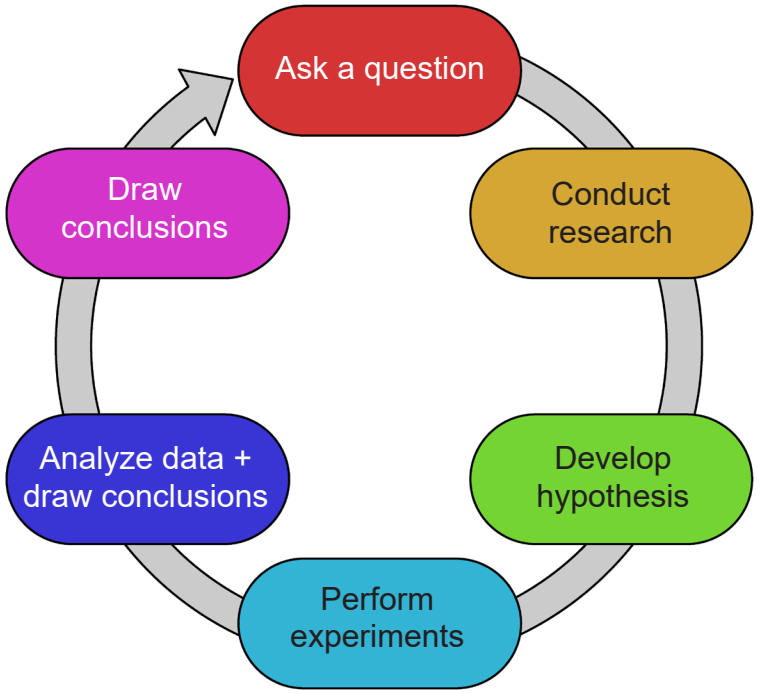


Figure 1.4 The scientific method is a circular, stepwise process used to establish causality between variables.

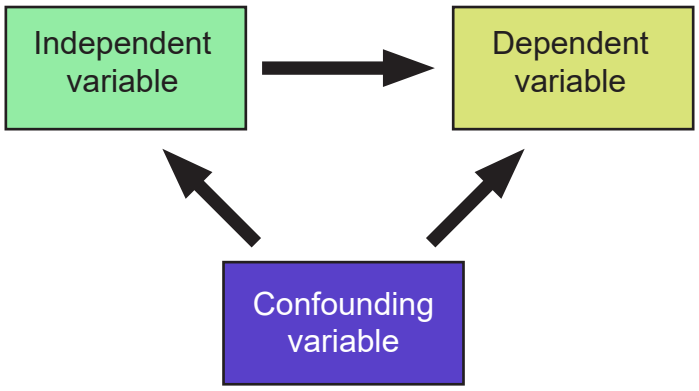


Figure 1.5 A good experiment establishes causality between an independent variable and a dependent variable by eliminating the influence of confounding variables.

and students may choose to study instead of sleep. To eliminate this confounding variable, it would be good to only compare grades for the students who slept for roughly the same amount of time. A control group in this experiment would be a group of students who are given the test without the opportunity to study. Ideally, these students will be as similar to the experimental group as possible: roughly the same age, gender distribution, educational history, and so on.

The strength of a well-designed experiment is that it establishes **causality**: a change in the independent variable causes a change in the dependent variable. Because of this, assuming that the sample population is **representative** (the distribution of the characteristics in the sample is proportionally similar to the distribution in the total population,) experiments allow us to extrapolate findings to a larger population.

Patients who participate in an experiment are often placed into an artificial environment or unnatural circumstances, which can affect their performance. For example, imagine you were participating in the "studying / test score experiment." Having the added pressure of knowing you were in a study could cause you to perform worse. Alternatively, knowing that you are in the experimental group might cause you to focus more intently during your study time, which could increase the group's average. To ameliorate this unintended effect, it's important for participants in the experimental group and the control group to be subjected to the exact same circumstances.

Observational study

A second way to gain scientific information is through an observational study, one type of which is a **quasiexperimental design**. These studies do not benefit from separation of patients into groups randomly, and therefore may have several uncontrolled variables. Quasiexperimental studies are usually done when conducting an experimental study may be too impractical or otherwise unethical.

An example of this type of study might aim to answer the question "Do people with traumatic brain injury have worse hand-eye coordination?" Performing an experiment with this question would be wildly unethical, since you cannot intentionally give your patients head injuries (It would also be challenging to find people willing to get hit on the head for science!). Instead of conducting this unethical experiment, you could answer this question with a quasiexperimental study. In the study, there would be two different populations of patients, an experimental group containing people who have already been diagnosed with head injury, and a control group that is demographically similar, but without a diagnosis of head injury. You could then ask both groups to

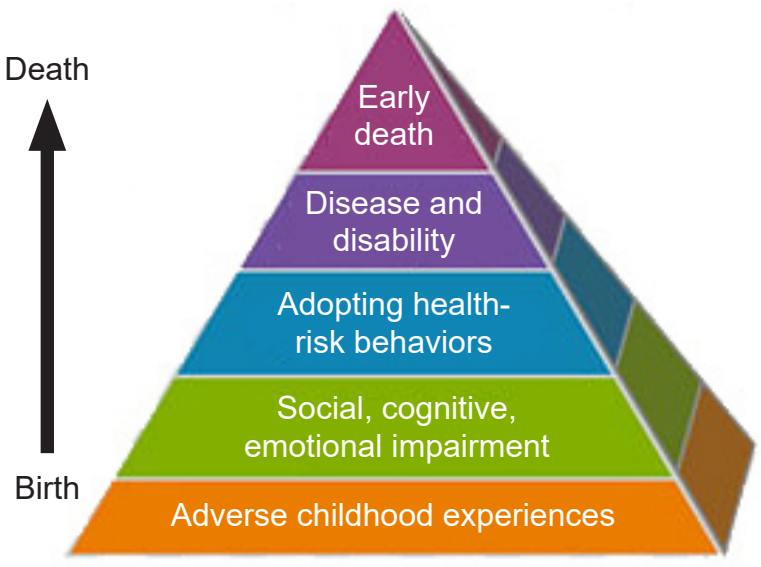


Figure 1.6 The Adverse Childhood Experiences (ACEs) was a large-scale quasi-experimental CDC study examining 17,000 people, linking childhood trauma with premature death.

perform a hand-eye coordination task and see if there are differences in their performance.

The weakness here is that there is no true randomness in the groups, thus the influence of confounding variables may affect our interpretation of the data. If the head injury group performed worse on the hand-eye coordination task, you would be able to demonstrate correlation, but not causation, which may present as a "chicken or the egg" problem. Perhaps the head injury group already had poor hand-eye coordination, which led to a car / workplace / sports accident, resulting in their head injury. In this example, we cannot conclude whether the brain injury directly causes a worsening of hand-eye coordination.

Case Study

A third strategy is the **case study**, a highly detailed description of a single patient and their condition. A case study documents the details regarding a specific deficit or enhancement, and is an opportunity to examine individuals with very rare conditions, which are useful for informing about the functions of different brain structures. Examining millions of healthy people may not give us the same insight as studying just one person with a specific injury or disorder. For example , case studies have been instrumental in teaching us about the brain structures involved with memory (Patient HM; chapter 13), language (Patient Tan; chapter 14), and fear processing (Patient SM; chapter 15).

Figure 1.7 A study of a single individual with special circumstances, such as Phineas Gage (left) and his brain damage (right), is called a case study.

Perhaps the most famous case study in all of neuroscience is the 1848 story of the railroad worker **Phineas Gage**. An unfortunate workplace accident left him with significant brain damage, largely to his frontal lobe. The subsequent changes in his personality taught us that one of the functions of this area of the brain is regulating our inhibitions (chapter 2).

Like a quasi-experimental study, case studies only show correlation, not causation. It is difficult to generalize the findings from a case study to the population at large. Usually, case studies are descriptions of nearly-one-of-a-kind individuals: A man with memory deficits in response to his brain being punctured by a toy fencing sword, a woman who had never experienced fear in her life, or a man who became overwhelmingly light-hearted and joyful after recovering from being shot in the head.

Case studies can be helpful for the development of hypotheses that can later be tested experimentally. For example, consider Patient HM, the man who had his left and right hippocampus surgically removed and couldn't create certain types of memory. A research question based on this case study might be "Is the hippocampus needed for the creation of navigational memory?" Then, an experimental study could be performed in rodents, where we surgically remove the hippocampus (experimental group) or a different part of the brain (control group) and see how well the rodents perform on a memory task.

1.3 What Neuroscience is NOT

As complex as the brain is, naturally misconceptions make their way into popular culture. It's valuable to address these myths about neuroscience and explain the evidence that refutes these statements .

"We only use 10% of our brain."

This wildly-inaccurate statistic has been the foundation for several fictional movies, TV shows, and books. The truth is that we use every part of the brain, and most of our brain is active most of the time - just not at the same time. Neurologist V.S. Ramachandran uses a great analogy to describe the fallacy of this myth: does a traffic light only use 33% of its lights? A properly functioning traffic light will use all three lights at very precise times. The activity of the brain is closely regulated by multiple mechanisms which prevent unusual electrical activity. In fact, if too many cells were active at the wrong times, just like a traffic light showing both green and red, chaos ensues - one cause of seizures is excessive neural activity.

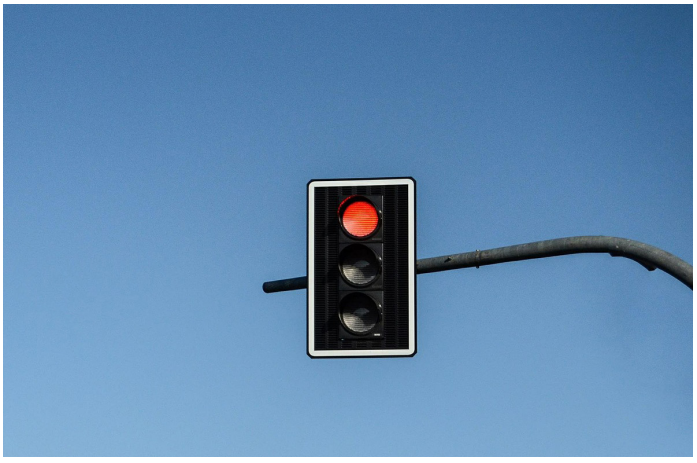


Figure 1.8 A fully-functioning brain uses nearly 100% of the component parts, but at precisely controlled times, like a traffic light.

"Forming memories causes new neurons to be born."

Another misconception is the idea that each new cell in our brain represents a new memory. While we are far from understanding the process of exactly how memories are formed in the brain, we do have a few clues. Most likely, memories are stored at the sites of close contact between neurons, called **synapses**. Changes in in ways neurons connect and communicate with one another is likely the mechanism behind how memories are formed and stored, rather than the creation of new neurons.

Even though the process of cell reproduction is halted in the majority of adult neurons, we are still capable of new neuronal growth, a process called **neurogenesis**. A few brain areas in particular, like the hippocampus (used in learning and memory functions; chapter 13) and the olfactory epithelium (used for smelling; chapter 9), do exhibit frequent birth and death of new neurons.

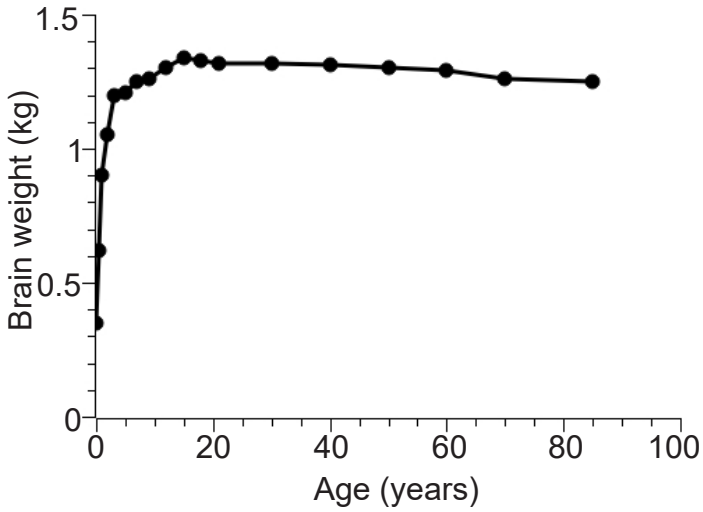


Figure 1.9 The weight of the brain does not increase much beyond the teen years, but we continue to learn throughout the rest of our lives.

"The brain cannot repair itself."

If neurons aren't being replaced in adulthood, then how do people spontaneously recover from neurological injuries like a stroke?

One of the most amazing features of the brain is the phenomenon of **plasticity**, the ability to change over time. Even if critical brain areas are damaged, it is theorized that the brain learns how to "rewire itself," essentially figuring out how to carry out these functions without using the damaged connections.

Unfortunately, there are some conditions that are **neurodegenerative**, meaning that their symptoms get progressively worse over time. Many of these disorders, like Parkinson's disease (chapter 10) and Alzheimer's disease (chapter 13), currently don't have any simple cures or treatments that don't carry risks and side effects. For people with these conditions, there isn't strong evidence that the brain can recover from the destruction caused by these diseases.

"If you are analytical, you are left brain dominant, but if you are creative, you are right brain dominant."

A common misconception is that the two hemispheres of the brain are responsible for wildly different functions. The truth is that nearly every function that the left half of the brain can do, the right half can do just as well, and vice versa. Sensory information, voluntary control of the muscles, memories, and many other behaviors can be performed equally well by both the left and right halves of the brain.

A major exception to the "left vs. right" component is the processing and production of language. For some reason unknown to scientists, these functions are heavily lateralized in the left hemisphere for most people (chapter 14).

Fascinatingly, we do have one strange quirk about signaling between the brain and the rest of the body: signaling pathways from the left brain crosses over to communicate with the right half of the body, and vice versa. This **contralateral** organization is an unintended consequence o f evolution, and is one of the major distinguishing features of the vertebrate brain.

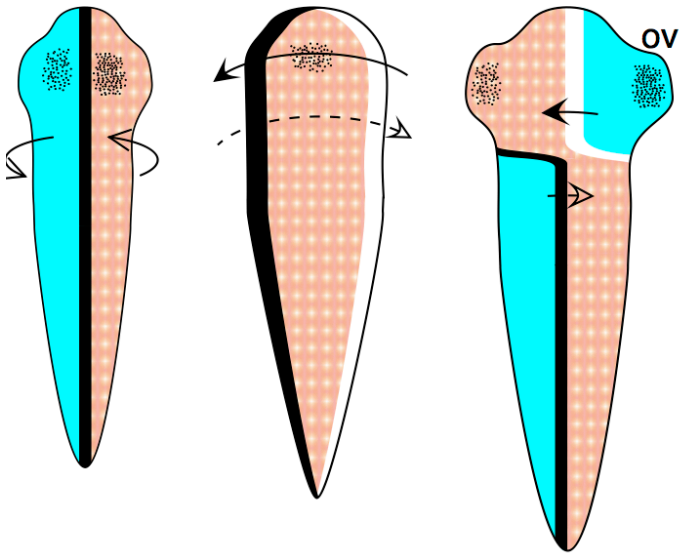


Figure 1.10 The vertebrate nervous system likely twists in development, resulting in a contralateral organization.

1.4 Neuroscience is ever changing

One of the most exciting and satisfying aspects of modern science is the rapidity of new discoveries the field. New findings are often communicated by publishing academic studies in scientific journals. More neuroscience studies were published between 2015 and 2020 than in the previous seventy years! But, advancements in neuroscience haven't always moved so quickly.

The ritualistic funerary rites of the ancient Egyptians around 2500 BCE provide a glimpse into how humankind's understanding of the brain has changed over time. When important Egyptians died, major organs including their stomach, lungs, and liver, were removed and stored in canopic jars in preparation for immortality in the afterlife. The fate of the brain, however, was much messier: Using a pair of sticks up the nose, the brain was blended up into a mush and flushed out of the skull using palm wine. The brain, apparently, wasn't needed for the afterlife.

Figure 1.11 The ancient Egyptians preserved some of the important internal organs after death (top), while the brain was scrambled using tools (bottom) and discarded.

Around 2,000 years later, ancient Greek physicians had a different understanding of the function of the brain. Aristotle developed a theory that the heart was the seat of the soul, and that blood was the life force that dictated a person's behavior. When a person was "hot blooded," they acted impulsively with no regard for consequences. In his view, the function of the brain was to cool the blood as the blood passed through it, which calmed the temper.

For the hundreds of years that followed, physicians attempted to correlate behaviors with changes in the brain. In the mid 1800s, Paul Broca was one of the first to suggest that specific areas of the brain were responsible for carrying out specific functions, which came to be called **localization theory**. Much evidence favors this line of thinking, such as the idea that language comprehension starts in a small patch of cells in the left hemisphere (Chapter 14),

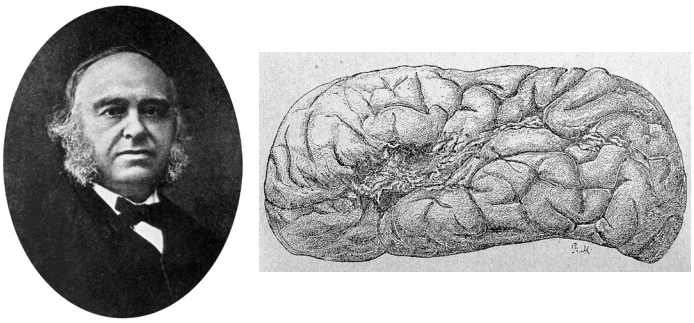


Figure 1.12 Paul Broca (left), through his study of the brains of patients with the language disorder aphasia (right), was a proponent of the localization theory of brain function.

perception of faces relies on a set of cells at the base of the brain (Chapter 7), and balance and motor coordination depends on the cerebellum (Chapter 10). On the other hand, the opposing view, called the **distributive processing theory**, suggests that behavioral functions require activation of cells across several different areas of the brain. Complex behaviors such as emotion, consciousness, or **cognition** (the act of generating knowledge through a combination of senses, memories, and thoughts) require coordinated action across distinct brain areas. Most likely, some behaviors are more localized than others, but still rely on signals from across many other brain areas. As with most fields of biology, absolutes are rare in neuroscience.

While anatomists and physicians tried to define the gross anatomical workings of the brain, they missed out on a layer of understanding at the level of cells until microscopy was widely adopted by the scientific community. In the early 1900s, a heated debate between two anatomists, Camillo Golgi and Santiago Ramon y Cajal, prompted researchers to look more closely at the neurons. Through careful drawings of their observations, they concluded that neurons had different shapes and therefore carried out different functions. This microscopic level analysis laid the foundation for understanding the cells that make up the nervous system and the way they communicate with one another (Chapter 2).

Today, we have a clearer understanding of the function of the brain, largely due to the advancements brought to us by a better understanding of animal biology and new technology. In 1954, the **electron microscope** was aimed at the space between neurons for the first time, allowing us to see a tiny anatomical component about 20 nanometers across - a thousand times smaller than the width of a human

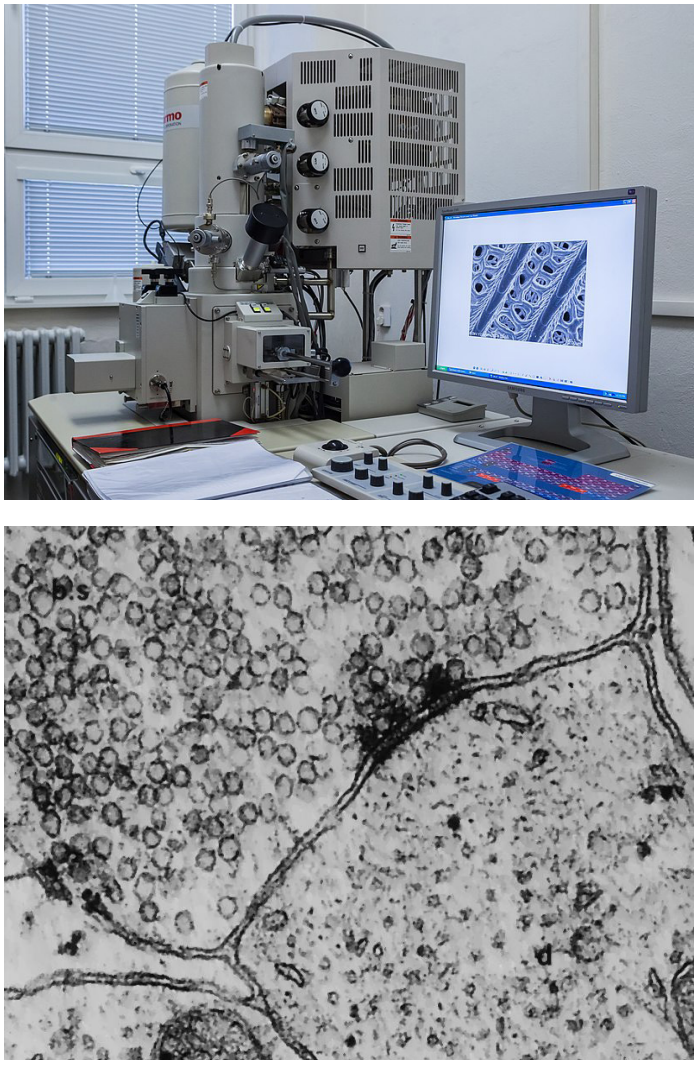


Figure 1.13 An electron microscope (top) gives us the resolution to visualize the synapse (bottom), the tens of nanometer distance between two neurons.

hair (Chapter 5). A medical diagnostic tool, the functional magnetic resonance imaging device (fMRI), made its debut to the neuroscience world in 1991, which allowed us to visualize brain activity while a person is actively engaged in behaviors, such as a decision making task, or while observing visual stimuli (Chapter 6). Today, much excitement revolves around visualization strategies like **CLARITY**, a method to render an entire brain transparent, which helps us to map out the nature of the connections that span the nervous system.

The ever-changing landscape of scientific inquiry presents a challenge. Our current understanding of the brain, as described here, is only a snapshot along the timeline of scientific discoveries. As we look to the future, many new discoveries will continue to reinforce what knowledge we have already amassed. But some discoveries, with help from not-yet-invented technology, will push the frontiers of knowledge and find compelling evidence against long-standing accepted theories in the field , prompting a shift in the paradigm.

1.5 Neuroscience is an integrative field of study

Realistically, our modern understanding of "neuroscience" is a combination of several academic disciplines, all using their strengths to understand some aspect of the nervous system. Because of this integrative nature, it is possible to study neuroscience from many different perspectives, each of them more fitting for answering different types of questions . These "angles" of analysis are described below.

At the root of the study is biology. Whenever you are studying living processes, such as learning, visual perception, or consciousness, you dip into the realm of biology. The broad field of biology can be subdivided into smaller, more precise categories. Molecular neurobiologists study proteins and gene regulation, cellular neurobiologists examine how networks of neurons communicate with one another, and cognitive neuroscientists study the underlying causes of behaviors. Understanding neuroscience involves genetics, such as the autosomal dominant neurodegenerative condition, Huntington's disease (Chapter 10.) Other biological subdisciplines such as ecology and evolution are also considered in neuroscience as well, such as the parasite Toxoplasma, which changes an animal's response to fearful stimuli, allowing the organism to reproduce as it moves through different species in the food web (Chapter 15.)

Psychology provided the earliest explanations about the brain and ideas about the origin of the mind. Some questions in this field branched off from philosophy, as people began thinking about the **"mind-body problem,"** the discussion that centered around the question if a function as complex as consciousness could result from activity of a clump of cells. Psychologists also wondered whether parts of the brain in isolation have different properties than when those parts are working together. This property, called **emergence**, is the idea that the whole is greater than the sum of its parts. Psychologists examine neuroscience from a top-down view, aiming questions at understanding the whole organism before looking at smaller components of the organism (compare this with biological approaches, often a bottom-up view that starts at the level of cells or molecules.)

Chemistry is a strong influencer of nervous system function - just ask anyone who forgot their morning cup of coffee! We use a variety of **endogenous** (originating from within the body) chemicals that act as signaling molecules, allowing communication between cells. These chemicals exist in many different structures, which determine their function. Some are acidic while others are basic; some are polar, others are fat soluble, and some are even gases (chapter 5). The nervous system is also highly sensitive to influence by **exogenous** chemicals (meaning they originate from outside the body), such as caffeine and cocaine (Chapter 11).

Many principles of physics can be observed through the functioning of neurons. For example, neurons maintain a negative electrical charge, usually measured on the scale of tens of millivolts (a millivolt is a thousandth of a volt.) The main way for neurons to send signals depends on a temporary change in this voltage; this signal is called an **action potential**. This change in voltage is brought on by the movement of charged ions across the cell membrane, and they closely follow the rules of magnetism: opposite charges attract while like charges repel (Chapter 4).

The field of **computational neuroscience** has grown from the use of mathematical modeling to describe or predict some aspect of the nervous system. If our current estimates are correct, we have around 86 billion neurons in the brain, a number so large that it is difficult to conceptualize. It would be nearly impossible to understand that many components of a system without taking advantage of the sheer mathematical strength of a computer.

Healthcare providers, like neurologists and psychiatrists, work from a different angle. They coordinate closely with researchers to apply scientific knowledge from the field or laboratory to treat patients, thus using biological principles as therapies. For example, neurologist Dr . Oliver

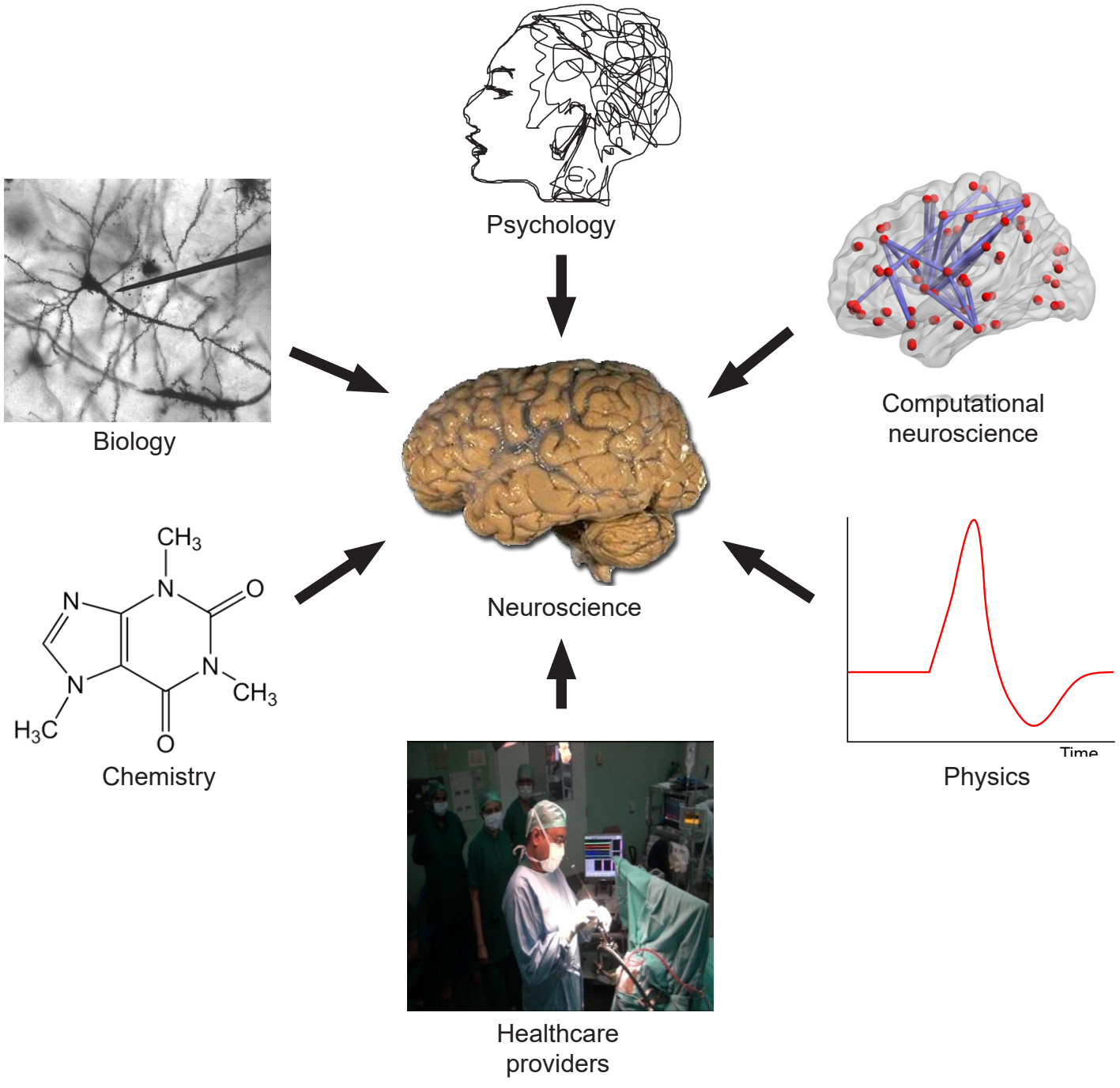


Figure 1.14 Neuroscience is made up several academic disciplines.

Sacks used his knowledge of the dopamine neurotransmitter system to treat patients with a paralysis-like condition in the 1960s, leading to the development of levadopa treatment for Parkinson's disease (chapter 5). Other healthcare providers use imaging strategies like a CT scan to assess the extent of a head injury or the location of a brain tumor, while an EEG can be helpful for the diagnosis of epilepsy (Chapter 6).

Engineers help develop the tools needed to understand questions in neuroscience, such as the patch clamp rig or electron microscope, highly specialized pieces of lab equipment. They also work closely with healthcare providers to translate science into therapy, such as the deep brain stimulator devices for the treatment of conditions such as Parkinson's disease.

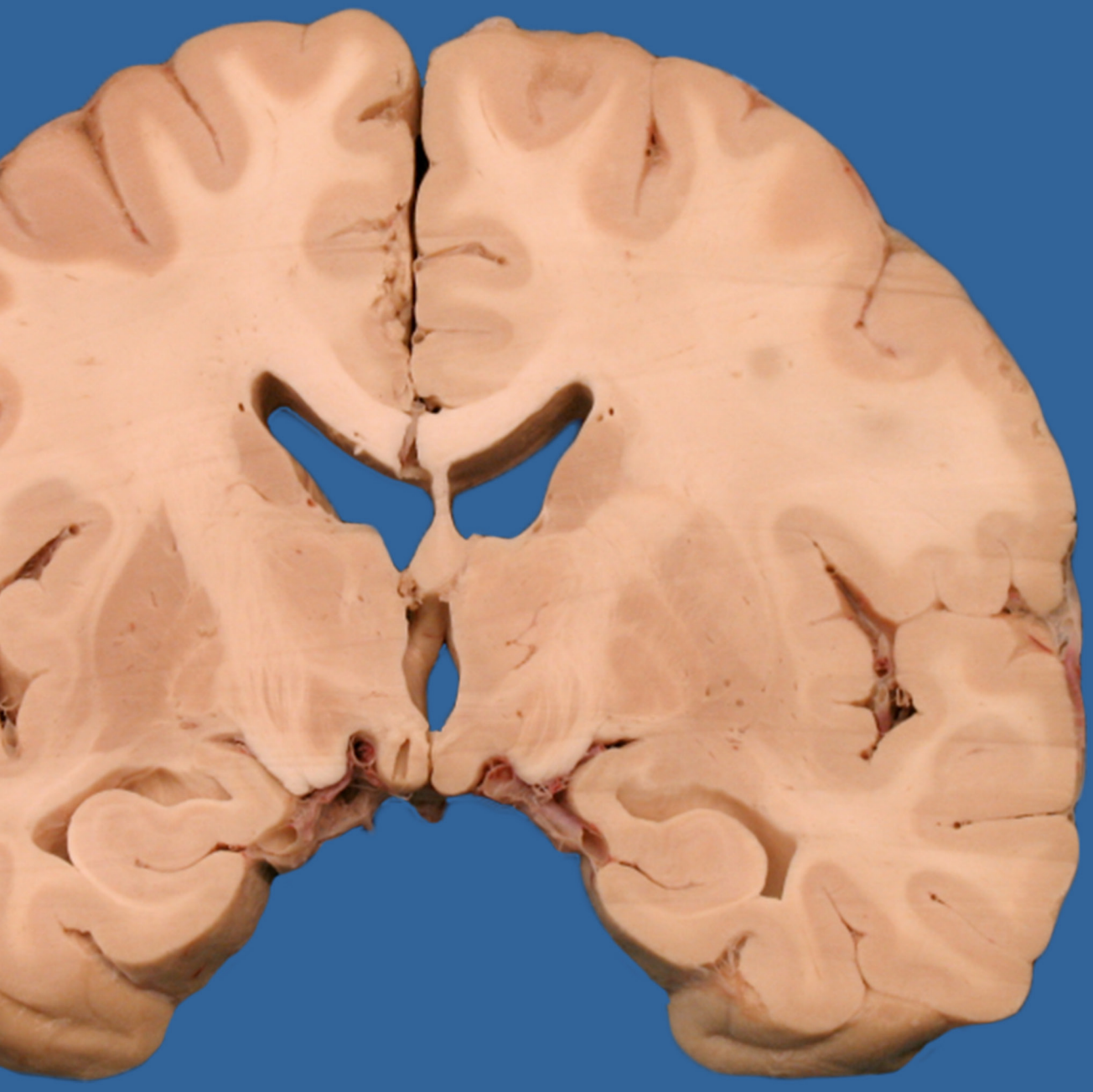
Collectively, all the people who participate in neuroscience in some way are united by their interest in the workings of the body. Because of the overwhelming complexity of the nervous system, there are many questions still unanswered. The continual appearance of new questions in neuroscience keeps us wondering, inspires curiosity, and promises a multitude of fascinating career paths for centuries to come.

The Open Neuroscience Initiative is funded by a grant from the Vincentian Endowment Fund of DePaul University.

Image credits
Cover: Beata Science Art (https://beatascienceart.com/)
1.1 https://commons.wikimedia.org/wiki/File:Figure_35_03_05_Brain_size_Vertical.png
1.2 https://commons.wikimedia.org/wiki/File:Action_potential_basic_shape.svg Modified by Austin Lim https://commons.wikimedia.org/wiki/File:Acetylcholine.svg https://commons.wikimedia.org/wiki/File:Resting_Potential.png Modified by Austin Lim https://commons.wikimedia.org/wiki/File:Glutamate_2,3-aminomutase_general_reaction.png Modified by Austin Lim
1.3 https://commons.wikimedia.org/wiki/File:Brain_image_of_blind_echolocator.tif Modified by Austin Lim
1.4 https://commons.wikimedia.org/wiki/File:The_Scientific_Method.svg Modified by Austin Lim
1.5
1.6 https://commons.wikimedia.org/wiki/File:Adverse-childhood-experiences-study-psychological-trauma-childhood-trauma-strengthen-prevention-thumbnail.jpg Modified by Austin Lim
1.7 https://commons.wikimedia.org/wiki/File:Phineas_Gage_CGI.jpg https://commons.wikimedia.org/wiki/File:Phineas_Gage.jpg
1.8 https://pixabay.com/photos/light-red-stop-street-427961/
1.9 https://commons.wikimedia.org/wiki/File:Brain_weight_age.gif Modified by Austin Lim
1.10 https://en.wikipedia.org/wiki/Contralateral_brain#/media/File:AxialTwistDevelopment.png
1.11 https://commons.wikimedia.org/wiki/File:WLANL_-_andrevanb_-_kist_uit_de_27-_31e_dynastie_(4).jpg https://commons.wikimedia.org/wiki/File:Three_of_four_cranial_crochets,_copies,_used_for_removing_th_Wellcome_L0058415.jpg
1.12 https://commons.wikimedia.org/wiki/File:Paul_Broca_2.jpg https://commons.wikimedia.org/wiki/File:Pierre_Marie,_Travaux_et_memoires._Wellcome_L0028667.jpg
1.13 https://commons.wikimedia.org/wiki/File:Scanning_electron_microscope_-_UFCH_JH_(2020)_01.jpg https://commons.wikimedia.org/wiki/File:Synapse_neuro-neuronale.png
1.14 https://commons.wikimedia.org/wiki/File:Pyramidal_hippocampal_neuron_with_fake_microelectrode.jpg https://pixabay.com/illustrations/psychology-mind-thoughts-thought-2422442/ https://commons.wikimedia.org/wiki/File:Caffeine_and_adenosine.png https://commons.wikimedia.org/wiki/File:ActionPotential.png https://commons.wikimedia.org/wiki/File:Brain_network.png https://upload.wikimedia.org/wikipedia/commons/4/40/Deep_Brain_Stimulation_surgery.png https://commons.wikimedia.org/wiki/File:Human_brain_NIH.png?wprov=srpw1_219

Chapter 2:

Anatomy of the Nervous System



Austin Lim, PhD (DePaul University)
Editor: Helena Ledmyr, PhD, INCF (Karolinska Institutet)

To physically take something apart is a great way to learn about the relationship between structure and function. Engineers and electricians use schematics to help them figure out what parts connect to where, and architects may use blueprints to aid them in the creation of a building. You can learn a lot about a system by understanding the function of the individual components and how they interact with one another. The whole is often greater than the sum of their parts, and a failure of one part to perform its duty can have wide reaching consequences.

The nervous system is one of the most complex systems that we know of. Parts of this system malfunction frequently, and the results are a wide range of neurological disorders that affect humans, from injury to genetic disorders.

The gross anatomy of the nervous system is an important foundation to the studies of other aspects of neuroscience. This chapter covers some of the major anatomical structures in the nervous system, starting with the brain, working down the spinal cord, all the way out to non-nervous tissue. The end of this chapter will also cover the many support structures that allow the nervous system to carry out their jobs.

Broadly speaking, the nervous system can be divided into two main categories: The **central nervous system (CNS)** and the **peripheral nervous system (PNS)**. Simply put, the CNS is the brain and spinal cord, while the PNS is all the other nerve cells in the body. The two systems are not isolated from each other; information passes rapidly between the PNS and the CNS, and vice versa. When a signal that originates in the PNS moves to the CNS, we sometimes say that the signal is *incoming* or *ascending*, while the CNS to PNS direction is *outgoing* or *descending*. Information that arrives into the CNS is also

Figure 2.1 Anatomical structures of the CNS and the PNS.

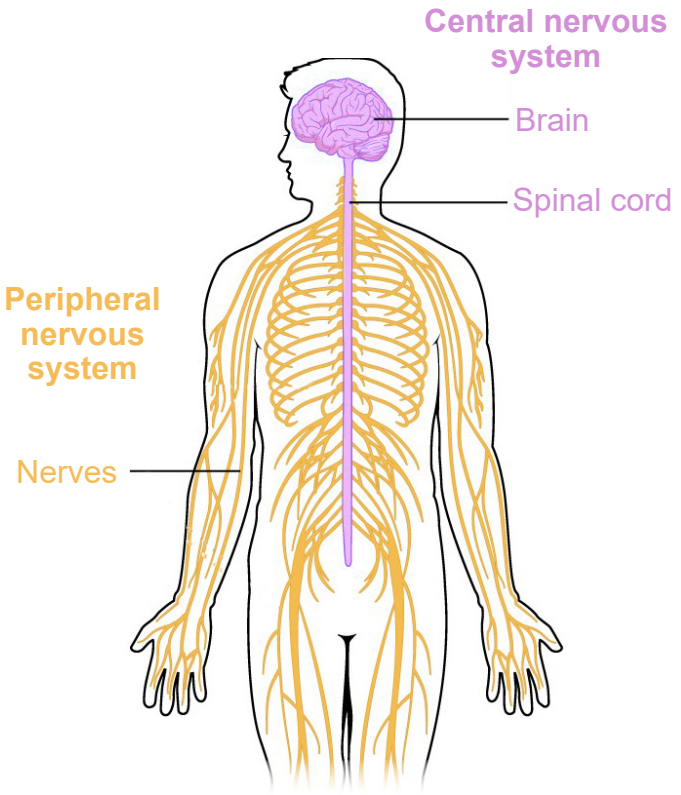
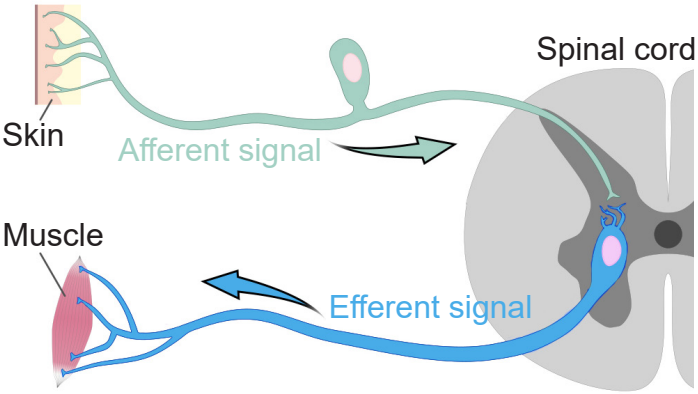


Figure 2.2 Afferent versus efferent signals.



called an **afferent** signal, while information leaving the CNS is an **efferent** signal. These two terms are frequently confused, but you can use the knowledge of other words that start with "e" to remember that an "exit" or an "escape" is something that moves away. Alternatively, an efferent signal is something that has an *effect* on the outside world, while an afferent signal *affects* the person.

Chapter 2 outline

2.1 Central nervous system (CNS)

Anatomically, the CNS consists of two organs, the brain and the spinal cord.

The brain is the main organ where movement originates, where thoughts and plans develop, and where consciousness is housed. The brain is what pushes us to act on our drives and desires, where language begins, and where memories are stored. At 160 mm (~6 in) long and 90 mm tall (~3.5 in), it has a total volume of about 1400 cubic centimeters and would fill about a third of a gallon. The intact adult brain weighs about 1.5 kg (3 pounds), which is barely 2% of total body weight. Despite this relatively small size, it is extremely power hungry, and uses up about one-fifth of the body's total energy expenditure.

Anatomical language of the brain

When talking about parts of the brain, it is helpful to have a set of words that can describe the location of various anatomical structures unambiguously. Think of the following three pairs of anatomical words as the "north, east, south, west" of the brain.

Consider a person who is facing sideways. Here, we'd like to use language that can describe anatomical structures that are along the "front of the head" to the "back of the head" axis. Parts of the brain that are more forward are called **rostral**, which comes from the Latin word meaning beak. To describe structures towards the back part of the brain, we use **caudal**, the opposite of rostral. Another pair of words that describes this same directional axis is **anterior** and **posterior**: ante- meaning "before" and post- meaning "after". Often, rostral and anterior are used interchangeably, as are caudal and posterior.

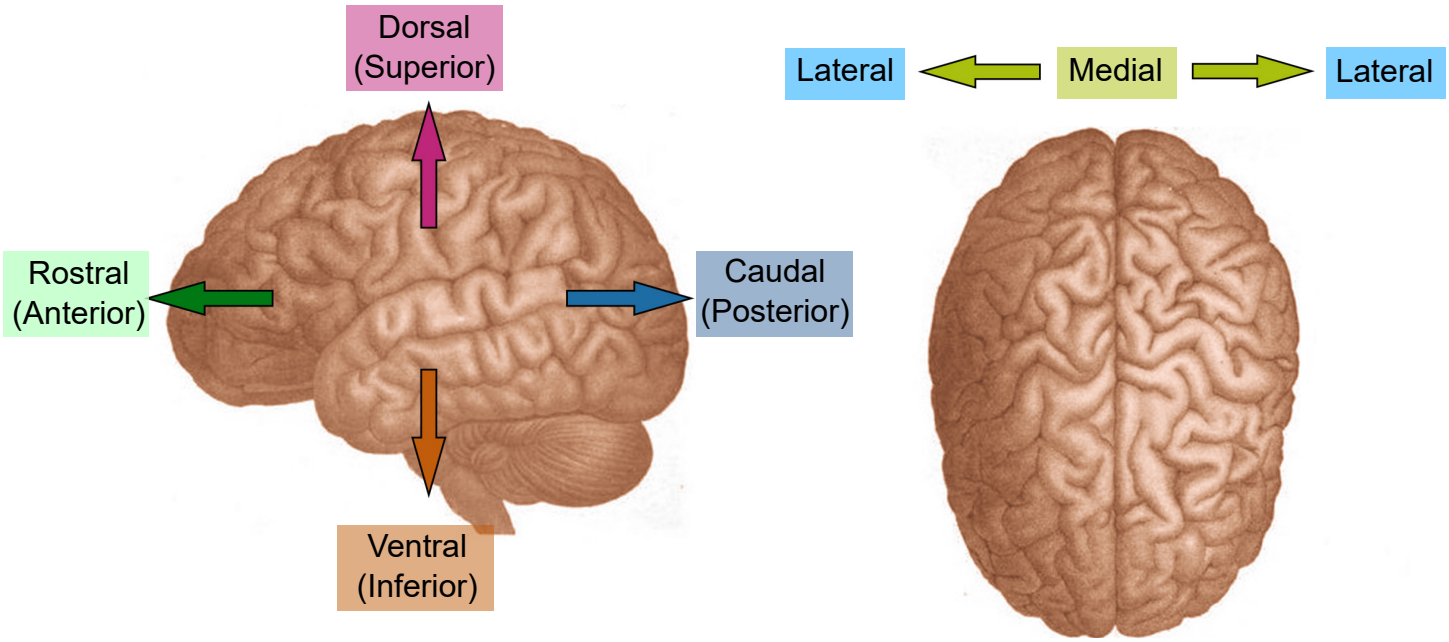
Another axis that should be considered is the top-to-bottom direction. Brain structures that are above, or closer to the top of the head, are described as being **dorsal**, while the structures that are below are **ventral**. As before, you may see another pair of words that share the same meaning. Sometimes, **superior** is used to describe a structure on top, while **inferior** is used to describe a structure below.

Now, imagine that the person has turned to face towards you. The third axis can be

demonstrated seen in this orientation. Brain structures that are closer to the center of the brain are described as being **medial**, while structures that are closer to the sides of the brain are **lateral**.

Neuroanatomists use these words to describe the relationship of one structure to another. For example, in Fig 2.7 which shows the four lobes of the brain, the frontal lobe is anterior or rostral to the parietal lobe, and the parietal lobe is dorsal to the temporal lobe. These anatomical words can also be combined to subdivide complex brain regions. The thalamus has many small subsections, such as the dorsomedial nucleus or the ventropostero-lateral nucleus. Naming structures with this anatomical language is useful in identifying where they are located in a brain scan or autopsy, but these words tell us nothing about function.

Figure 2.3 A cartoon illustration of the brain showing anatomical langauge.



Visualizing the brain

As a three dimensional structure, the brain can be sectioned for visualization or analysis in several ways. Here, we will describe the three main orientations, each at right angles to the others.

One projection of the brain is called a **coronal** slice. Imagine that you were looking at a brain of a person who was looking to the left. If the brain was cut into several sections vertically, slicing from anterior to posterior, the brain would be cut into coronal slices. “Corona” comes from the Latin word for crown, since looking at the brain cut into slices parallel to a crown would be a coronal section.

Another way to image the brain is called a **horizontal** projection. In a horizontal slice, cuts are made along the dorsal-ventral direction, from the top of the brain to the bottom. Given that the person is standing up, horizontal slices are parallel to the plane of the ground. Another way to think of the horizontal plane is through brain imaging techniques where a person moves headfirst into a scanner. Here, the machines are capturing horizontal images where each progressive image is a snapshot of the brain slightly more ventral than the last.

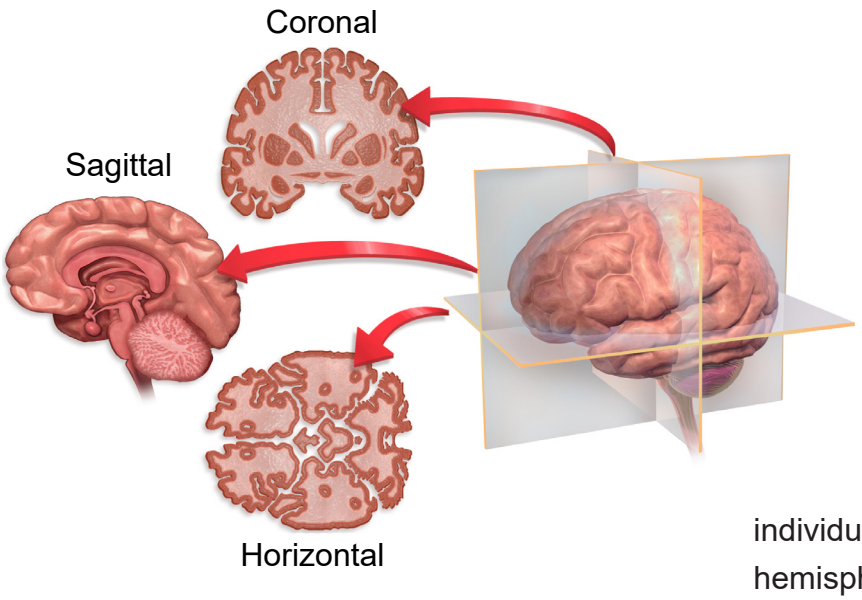


Figure 2.4 Cross sections used for visualizing the three orientations of the brain. “Sagittal” refers to the slice that divides the brain into left and right hemispheres, while slices parallel to that are called “parasagittal” sections.

The third and last direction is a **parasagittal** slice. The parasagittal section takes cuts parallel to the midline, slicing in the medial / lateral plane. If a person was facing you, and brain sections were taken across from left to right, these would be parasagittal sections. Notably, because the parasagittal section only samples from one hemisphere of the brain at a time, parasagittal slices are never symmetrical. The word *-sagittal* comes from the Latin word meaning arrow, like the zodiac sign Sagittarius (If you were an archer drawing back a bow and arrow, a parasagittal slice would be parallel to that plane).

In a sliced section of the brain, you might notice that brain tissue has different colored areas. Some brain tissue is pale and almost white, and these areas are described as **white matter**. Generally, white matter represents pathways of communication. For neurons to send signals rapidly, the cells can be modified in a way that adds several layers of fatty lipids called **myelin**. This modification causes light to reflect, causing it to appear white to the eye (think of the strips of fat that you might see in a steak). Other sections of brain tissue have a darker pink / gray color, appropriately called **gray matter**. These areas are usually dense with cell bodies.

The brain has two very similar halves, the left and right hemisphere. Oftentimes, in the neurotypical individual, information passes between both hemispheres rapidly: what one hemisphere senses or learns, so does the other hemisphere. It is the white matter tracts that allow for this transfer of information. When a white matter pathway crosses from one hemisphere to another, we call it a decussation. In coronal or horizontal brain slices, you may be able to observe the main white matter tract that allows for the passage of information between the two hemispheres, called the **corpus callosum**.

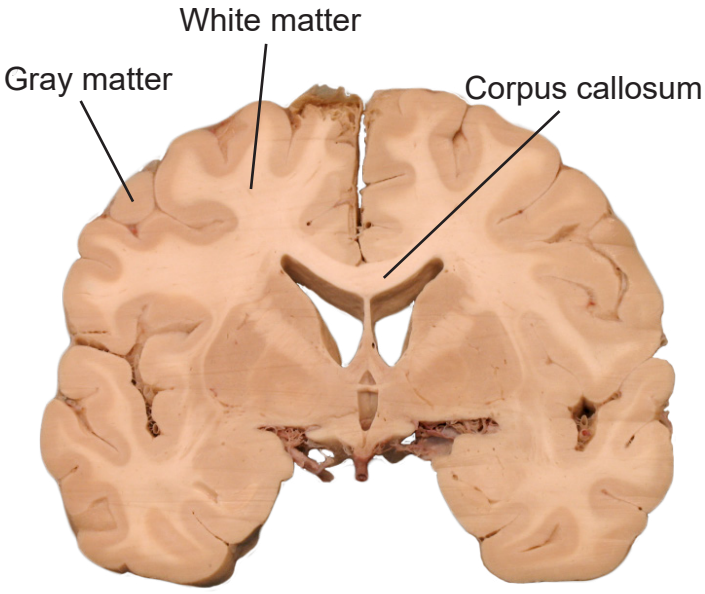


Figure 2.5 Coronal brain section showing examples of white matter, gray matter, and the **corpus callosum**, the major communication tract between the left and right hemispheres.

Brain structures through development

The brain is often divided into five groups based on their developmental origin. When the embryo first starts to form, cells are classified into three main germ layers: the ectoderm, mesoderm, and endoderm. Of the three, the ectoderm eventually develops into the nervous system. The ectoderm layer folds into itself, and after merging at the surface, creates the **neural tube**. This neural tube forms during the third to fourth week of gestation. The cells of the neural tube will eventually become the components of the CNS.

Early in development, the neural tube has three distinct compartments. At this time, the undeveloped nervous system is appropriately called the "three-vesicle stage." One week later, parts of this progenitor nervous system divide; this becomes the "five-vesicle stage." The names of the five vesicles can be used to describe either the stages through development, or a grouping of structures that eventually form in adulthood. From posterior to anterior, they are:

1. Rhombencephalon, or **hindbrain**

Evolutionarily speaking, the rhombencephalon represents the oldest part of the CNS. These structures likely evolved some 570 million years ago. As these structures develop into the five-vesicle stage, they become subdivided into two regions:

a. Myelencephalon

The myelencephalon develops into the **medulla oblongata**, a structure that is found at the posterior end of the brain stem. The medulla contains many clumps of neurons that are responsible for functions that an organism carries out unconsciously, such as breathing or changes in heart rate and blood pressure. It also contains areas that can detect toxins in the blood that come from dietary sources, triggering vomiting.

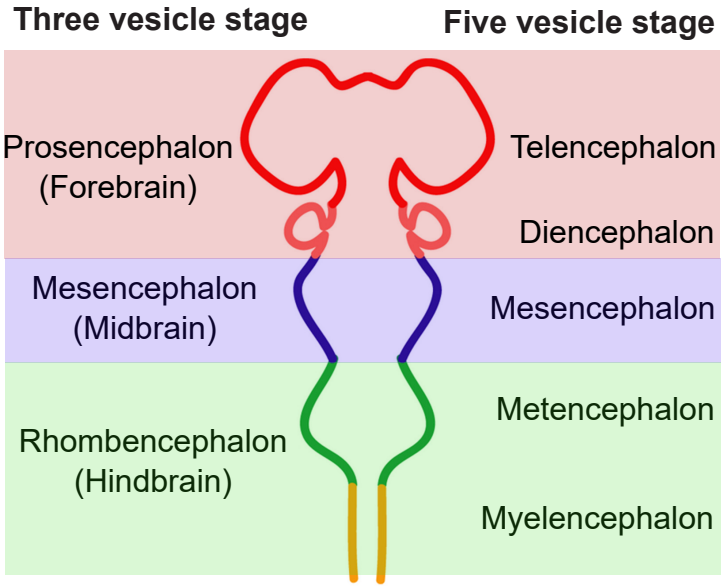
b. Metencephalon

The metencephalon develops into two structures, the **pons** and **cerebellum**. The pons, like the medulla, helps us perform involuntary functions like breathing. It also contains several areas that help us hear sounds and taste foods. The cerebellum, or "little brain", is best known as a structure that enables motor control functions, such as balance, coordination, posture, and learning physical actions.

2. Mesencephalon, or **midbrain**

The midbrain structures do not change much from the three-vesicle stage to the five-vesicle stage. There are many structures in the midbrain, and they can perform a wide variety of functions. For example, the periaqueductal gray allows us to respond to painful stimuli, the red

Figure 2.6 The future structures of the developing nervous system.



nucleus and substantia nigra coordinate complex movements, the tectum allows us to respond to incoming visual stimuli, and the ventral tegmental area is important for the processing of reward and motivation.

3. Prosencephalon, or **forebrain**

The prosencephalon eventually develops into the "higher order" brain regions, including the cerebral cortex. Most of the time, when you see an image of an intact brain from the side or the top, the structures that are visible to you are the forebrain structures.

a. Diencephalon

The diencephalon contains a few major structures. The **thalamus** is often referred to as a "relay station" in the brain, since almost every sensory modality (sight, taste, touch, hearing, and more) passes information through the thalamus. The **hypothalamus** is also within the diencephalon, and this structure serves as a communication route to the body's endocrine system. Neural signals originating in the hypothalamus have the capability to influence the chemistry and function of the entire body.

b. Telencephalon

Structures of the telencephalon are the basal ganglia and cerebral cortex. The **basal ganglia** are made up of a series of brain structures that are used for such behaviors as motor and habit learning, emotional processing, and action selection. The **cerebral cortex** makes up the outermost layer of the brain: the word cortex itself comes from the word meaning "bark," the outer layer of a tree. Here, the brain processes behaviors such as attention, memory, and language.

It may be useful to think of brain structures as roughly being organized in a phylogenetic "timeline". The more basic features of a creature are generally controlled by posterior brain structures, and the more complex functions are carried out by structures towards the anterior end of the brain. For example, the brain stem of the hindbrain contains networks of cells for basic survival, such as respiration and simple locomotion. Next, the midbrain is important for motivation and more coordinated movements. And finally, the forebrain carries out the highest order functions, such as personality, intentional inhibition of actions, and planning out long term actions. All of the structures work together simultaneously to produce the whole range of animal activities.

Major lobes of the cortex

Most of what we see when we imagine the brain is **cortex**, the bumpy outer surface that is made up of raised ridges (**gyri**, singular **gyrus**) and grooved indentations (**sulci**, singular **sulcus**; sometimes also called a **fissure**). Although each gyrus and sulcus has a name that either identifies its function or location, there are only three sulci that we will introduce here to help orient us around the neuroanatomical features of the cortex. The **longitudinal fissure** is the most obvious fissure in the brain. It divides the two hemispheres, running along the anterior-posterior axis, visible from a dorsal view of the brain. If you were to cut along the longitudinal fissure completely, you would get two symmetrical portions of brain, the left and right hemispheres. The **central sulcus** is a large fissure that starts at the dorsal part of the brain at about the halfway point on the anterior-

posterior axis. In a sagittal view, the central sulcus runs ventrally about half the length of the brain. The other groove worth noting is the **lateral fissure**. This one runs roughly along the anterior to posterior direction, and curves gently dorsally. Again in a sagittal view, it is roughly seen in the middle third of the brain in the anterior-posterior axis.

The cortex is roughly divided into 4 major lobes, which are named after the bones of the skull that surround each section of brain. The lobes are paired, meaning that the whole brain contains two of each, a left and a right. In general, the structures are roughly symmetrical. The four lobes of the cortex and their functions, approximately from posterior to anterior, are:

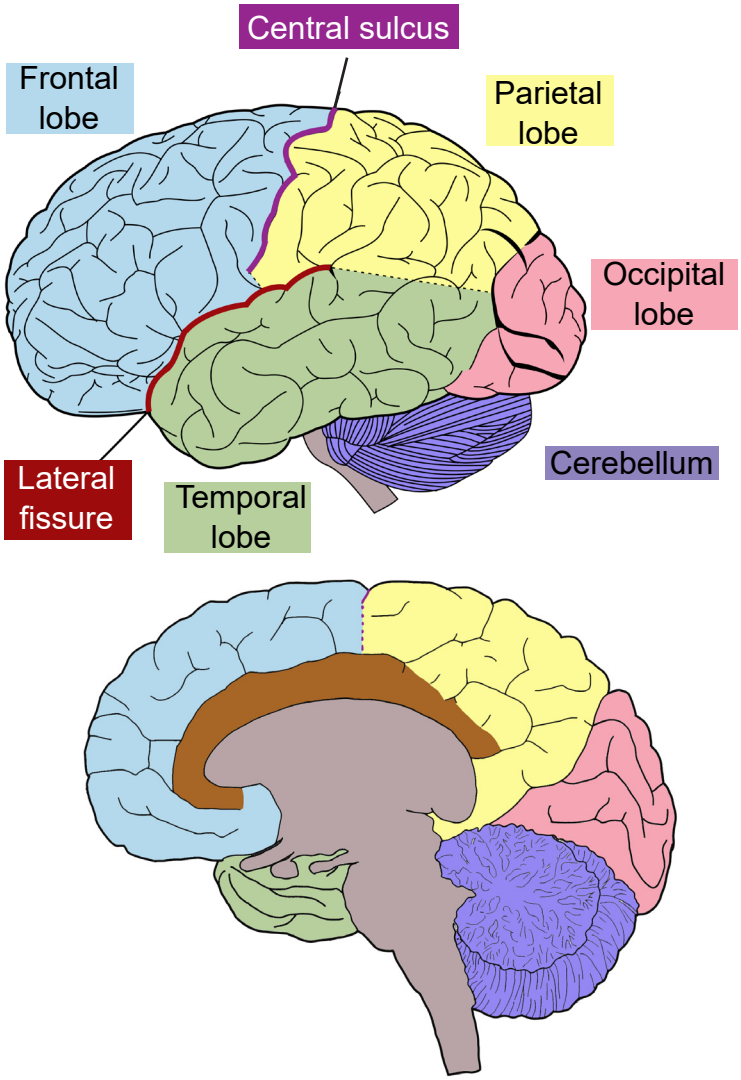


Figure 2.7 Lobes of the telencephalon and two of the major sulci from a lateral view (top) and a midsagittal view (bottom). The insular cortex is not visible from the outside.

1. Occipital lobe

The occipital lobe is the posterior most section of the brain. Anatomically, there is not an obvious border that separates the occipital lobe from adjacent areas of the cortex. The occipital lobe is the smallest of the four lobes.

The main function of the occipital lobe is for processing of visual stimuli. Our eyes are capable of capturing light and converting that light into signals. The **primary visual cortex** of the occipital lobe, also called **V1**, interprets those signals into a representation of the visual world. Other vision-related stimuli, such as objects in motion, object orientation, and color are also processed by neurons in the occipital lobe.

2. Temporal lobe

The temporal lobe is the ventral-most lobe of the brain, and the lateral fissure marks its dorsal border. It is anterior to the occipital lobe. It is the lobe of the brain that is immediately behind the temple, the structure on the lateral aspect of the skull. The name comes from the Latin word meaning time: the passage of time in adults is often marked by graying hair, and these gray hairs may first appear at the temples.

The auditory system allows our brain to interpret sound waves. We can distinguish between voices talking, instruments playing, and dogs barking. This is possible because of a population of cells located in the temporal lobe called the **primary auditory cortex**, or **A1**.

The ability to remember important facts depends on memory-related processes. These

functions are carried out in part by a brain structure called the **hippocampus**, which is buried medially and ventrally in the temporal lobe.

The temporal lobe also houses some structures that are important for language. Patients with injuries in parts of the temporal lobe may experience deficits in the comprehension of language, while different injuries lead to a deficit in the production of language.

3. Parietal lobe

The parietal lobe is in the dorsal aspect of the brain, and immediately anterior to the occipital lobe. The anterior-most (front) end of the parietal lobe is bordered by the central sulcus, and the ventral border is the lateral fissure.

The sense of touch is complex. With our skin, we are able to detect light touch, temperature, pain, vibration, and many other modalities. This ability to sense different tactile properties of things in the world around us with our body is one of the major functions of the parietal lobe. Another closely related sense, **proprioception**, the ability to identify where parts of your body are located, is also processed by neurons of the parietal lobe. These functions are carried out by the **primary somatosensory cortex**, or **S1** (*soma*- referring to the body).

4. Frontal lobe

The frontal lobe is the anterior most part of the brain. The posterior border of the frontal lobe is the lateral sulcus. Among mammals, it is the largest of the four lobes.

The frontal lobe contains the **primary motor cortex**, or **M1**, which is directly anterior to the central sulcus. M1 contains neurons that control movement of the body. For example, if you were to activate the dorsal most part of M1 in a person, you would see motor activity in the leg.

The frontal lobe also carries out the "higher order" functions of the brain. Our personality is influenced by the frontal lobe - an injury to these brain structures can result in a radical change in a person's behavior. Frontal lobe allows us to do mental math, to hold a string of letters in our head to be repeated backward, and to suppress socially unacceptable actions.

Clinical connection: Phineas Gage

The mid-1800s saw an expansion of industry in the United States. The most reliable and quickest way to move goods and people was with the railways that were starting to zig-zag across the growing country. The factor slowing down railway expansion was terrain: land had to be relatively flat for the tracks to be laid down. Terraforming mountains was a dangerous ordeal, and in the years before TNT, a relatively safe explosive was developed, work accidents were a risk that these demolition workers faced.

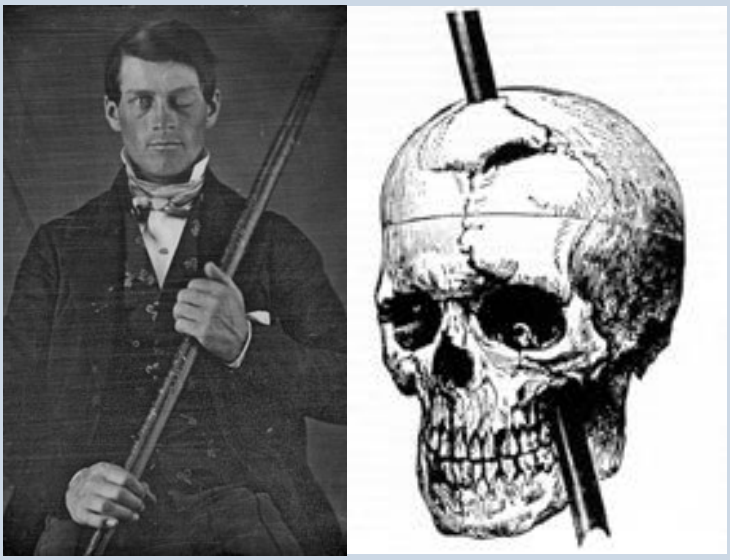


Figure 2.8 Phineas Gage with the tamping rod (left) and a drawing of the injury (right).

Phineas Gage was one of those workers. In the green mountain hills of Vermont, Gage was putting explosive powder into a crack in a mountain to clear land for a new railway. As Gage packed the explosive using a three-foot-long metal rod, a spark accidentally ignited the blast prematurely, causing the tamping iron to rocket cleanly through Gage's skull.

Miraculously, Gage survived the blast. Within a month, he had made an almost complete recovery. Gage was talking excitedly with his doctors, he was eating voraciously, and even reported experiencing no pain. While the doctors noticed that his entire frontal lobe had been destroyed, it was his friends who noticed the dramatic change in his personality: whereas he was once a friendly man, adored and respected by his coworkers, the new Gage was irreverent, generally unlikable, and prone to using profanity at the most inappropriate times. The pre-injury Gage was a shrewd businessman who followed through with his plans, but Gage now was unreliable and at times, acted almost animalistically. Because of the injuries to his frontal lobe, his friends described him as being "no longer Gage."

Spinal cord

It is sometimes easy to think of neuroscience as a focused study of the brain: How does activity of the brain contribute to behavior? In what ways does the brain change in disease? Why do the cells of the brain behave the way they do?

The truth is there are many parts of the body that also fall under the broad study of neuroscience. For example, the automatic knee-jerk reflex that a clinician examines when they tap on your patellar tendon is a test of the nervous system. The reflex is driven by sensory neurons that detect muscle stretch, motor neurons that cause the kicking response, and interneurons that prevent the opposing muscle from acting. We have neural circuits that provoke changes in the activity of our internal organs, from the beating of our heart to the digestion of food, and the study of these systems is certainly part of neuroscience as well.

Moving posterior from the brainstem is the other organ of the central nervous system, a long, thin structure of nervous tissue called the **spinal cord**. It functions to carry information

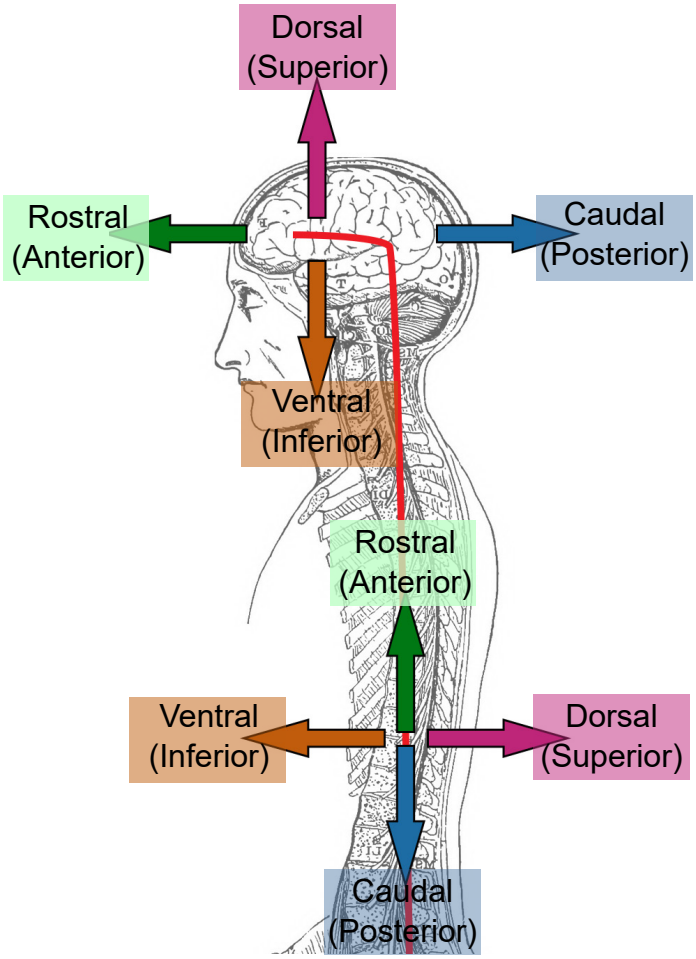


Figure 2.9 Unlike other mammals like dogs or cats, humans walk upright, giving us a "hooked" nervous system.

both upwards towards the brain, and downwards towards the body's other organs and muscles. It can also process sensations and form an appropriate motor response in the absence of brain input. The spinal cord originates roughly at the level of your neck and runs down to the small of your back, giving it a length around 44 cm (17.5 inches). The diameter of the spinal cord is not uniform all the way down, being ~13 mm (0.5 inches) at its widest and ~6.5 mm (0.25 inches) in diameter at the thinner areas (slightly smaller than the diameter of a pencil.)

The spinal cord is housed within a series of bones, called the **vertebral column**. Although the spinal cord itself is continuous, it can be divided based on the overlying vertebrae. A combination of a letter and a number is used to identify each section of the spinal cord, the letter corresponding to the vertebral section and the number referring to the number of bones down from the previous section (the smaller numbers are more anterior, larger numbers more posterior). Branching off from each section of the spinal cord are two pairs of nerves, the afferent (incoming to the CNS) sensory nerve roots which branch from the dorsal side of the spinal cord, and the efferent (outgoing from the CNS) motor nerve roots which branch from the ventral side of the spinal cord. These two branches meet and extend away from the spinal cord. After merging, they are called the **spinal nerves**, and humans have 31 pairs of these.

Moving from anterior (top) to posterior (bottom), the four regions of the spinal cord are:

1. Cervical

The cervical region corresponds to the upper 8 pairs of spinal nerves. Nerves that exit through the cervical region innervate the muscles in the neck, shoulders, arms and hands. Afferent nerves detect somatosensory inputs from these

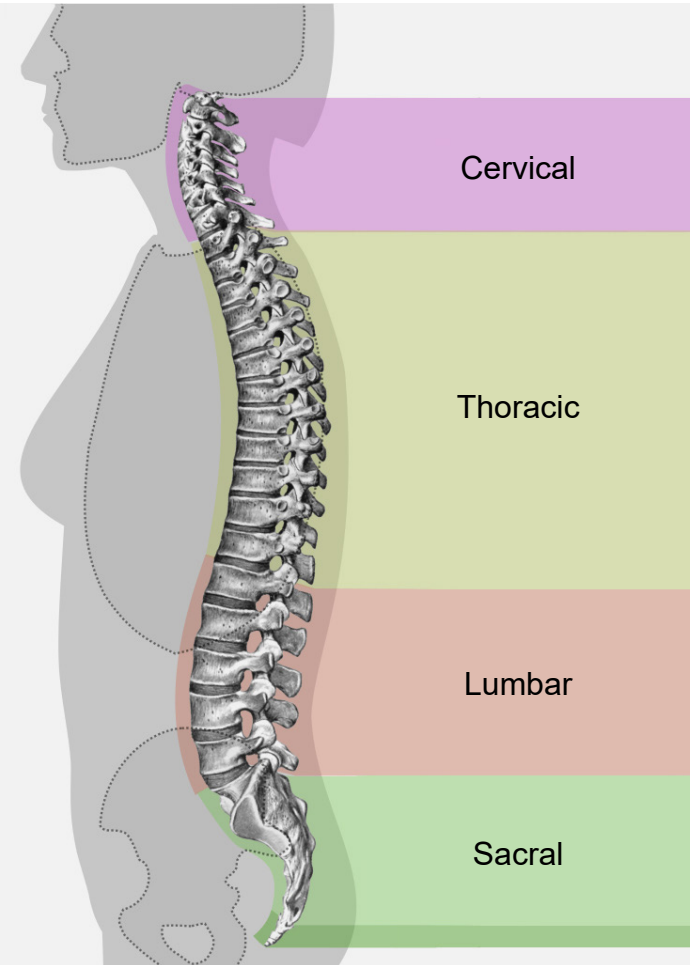


Figure 2.10 The names of the regions of the vertebral column, the bones that protect the spinal cord.

same areas. Sections C3 through C5 innervate the diaphragm, so an injury at this level or higher can quickly lead to death since the person may stop breathing. The spinal cord is at the widest diameter at the cervical area, as it has a swelling that corresponds to the many inputs and outputs to the arms.

2. Thoracic

There are 12 pairs of spinal nerves that make up the thoracic area of the spinal cord. These regions innervate the middle trunk area, the intercostal muscles between the ribs, and abdominal muscles. Branches off the spinal nerves in the thoracic areas are responsible

for changing the activity of the various internal organs during a fight-or-flight response (more on the autonomic nervous system in section 2.2).

3. Lumbar

There are 5 pairs of lumbar spinal nerves. These pathways carry motor command information to the hips, thighs, and knees . Afferent lumbar inputs detect sensory information from the ventral side of the legs, such as the top of the thigh or the shin bone. As in the cervical region, the lumbar region has a swelling that increases the diameter of this section of spinal cord compared to the thoracic or sacral areas.

4. Sacral

At the posterior-most end of the spinal cord is the sacral region, which consists of 5 pairs of nerves. Sacral spinal nerves control flexing of the toes. These nerves detect sensory information around the genital organs and the dorsal aspects of the legs, like the buttocks and the back of the thighs. There are also parasympathetic nerves that come from the sacral region, and these innervate the colon, bladder, and genital organs (again, more on the autonomic nervous system in 2.2).

Since information must pass through the anterior regions of the spinal cord to reach the posterior parts of the body, the more anterior an injury, the more parts of the body that are affected. The disease that affected former U.S. President Franklin Delano Roosevelt likely damaged posterior spinal cord structures, which explains why function of his legs was lost while function of his arms were left intact. However, the injury actor

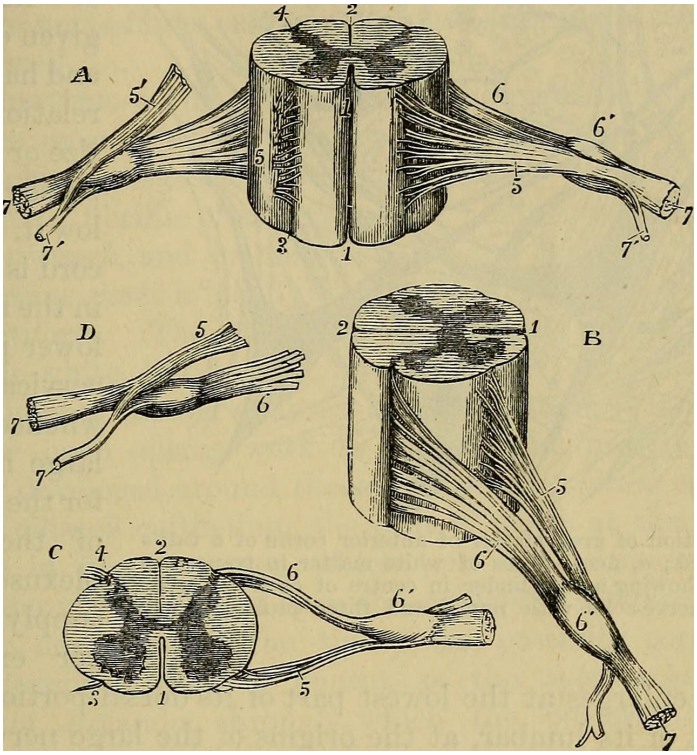


Figure 2.11 Drawing of the spinal cord and the connecting spinal nerves.

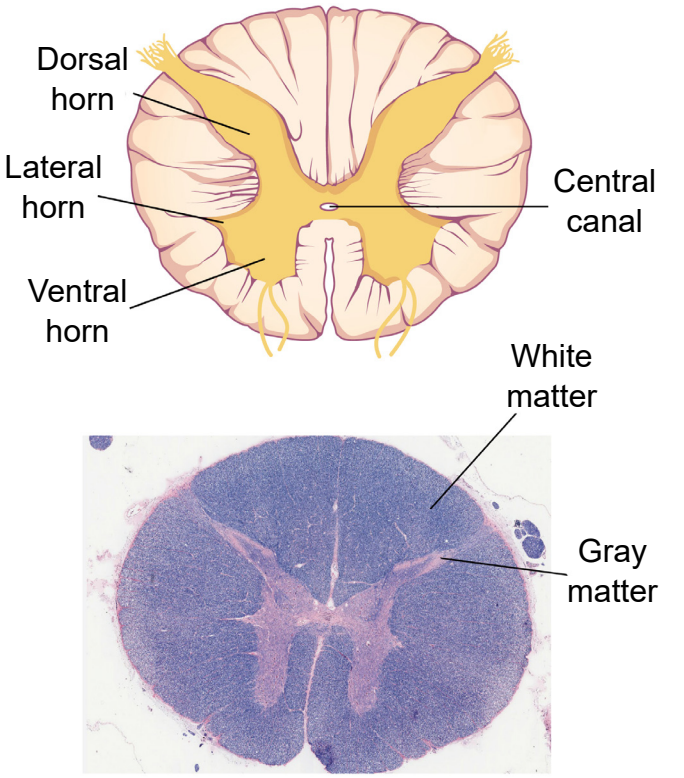


Figure 2.12 Cross-section (transverse) of the spinal cord showing a few anatomical features (top). Section of spinal cord stained with LFB, which dyes myelin in blue, which is why white matter looks more blue than gray matter (bottom).

Christopher Reeves sustained while horseback riding destroyed his spinal cord at the anterior-most level of C1, causing complete paralysis and lack of sensation from the neck down.

Unlike in the brain, there is basically only one truly meaningful projection of the spinal cord, and that is by cutting sections in the **transverse** plane. These sections are parallel to the ground if the spinal cord was oriented vertically, like a person standing up. Sometimes, this projection is also called a **cross-section**.

No matter which level of the spinal cord you look at, all cross-sections have a few similarities. Most notably, the inner portion of a section has a butterfly-shaped structure of gray matter surrounded by a border of white matter. Gray matter contains mostly cell bodies while white matter contains mostly pathways of communication. Therefore, the ascending sensory tracts and descending motor tracts run along the outer portions of a spinal cord section.

Another feature that is seen across all levels of the spinal cord are the entry points for the afferent sensory nerves and the exit points for the efferent motor nerves. Somatosensory information arrives into the spinal cord from the dorsal side. Most of the sensory neurons have their cell bodies outside of the spinal cord in a large clump of nervous tissue close to the dorsal side called the **dorsal root ganglion**. On the other hand, efferent motor nerves exit the spinal cord on the ventral side.

One of the main differences across different sections of the spinal cord is the ratio of white matter to gray matter. In general, the ratio leans towards more white matter at the anterior regions of the spinal cord compared to the posterior parts of the spinal cord.

The CNS ends at the spinal nerves. The remaining nerves that affect the muscles and beyond are the beginning of the peripheral nervous system.

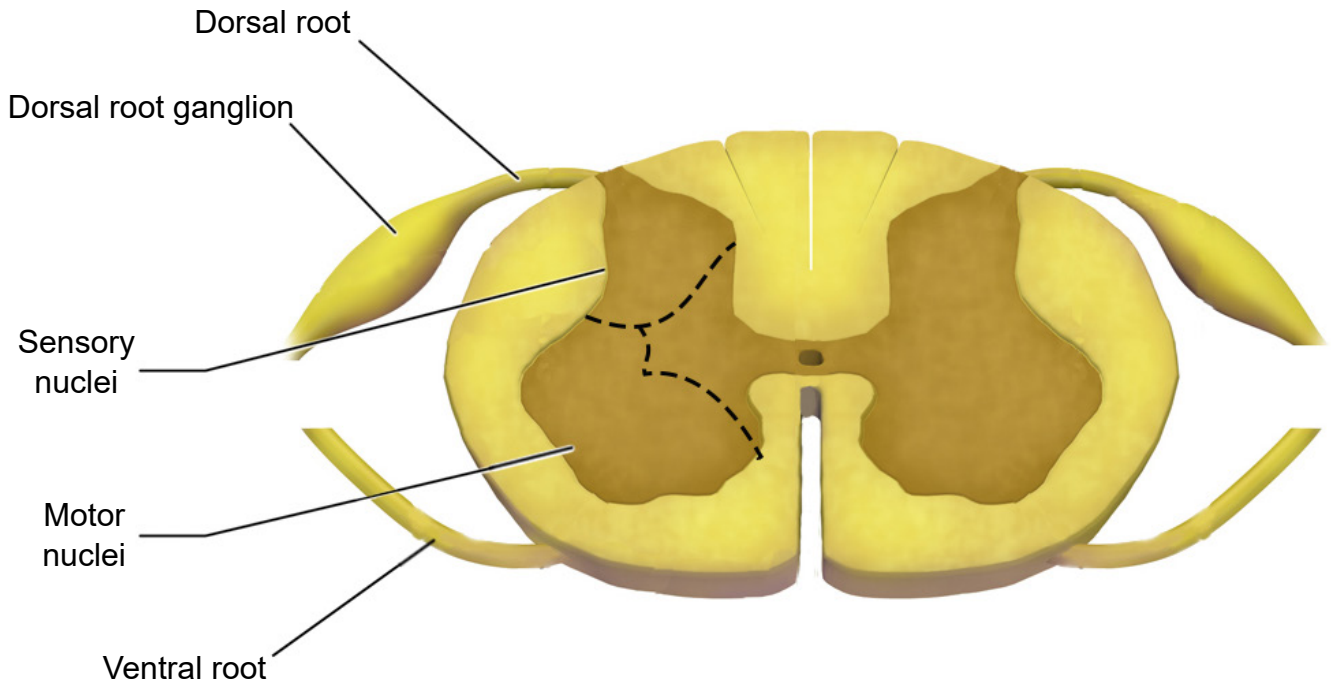


Figure 2.13 Cross-section of the spinal cord showing the dorsal root ganglion, the clump of cell bodies from the incoming sensory neurons.

2.2 Peripheral nervous system (PNS)

The PNS functions as the intermediary between the CNS and the rest of the body, including the skin, internal organs, and muscles of our limbs.

When talking about the PNS, another pair of anatomical words will be introduced, **proximal** and **distal**. A part of the body that is more proximal is something that is in close *proximity* to the CNS, whereas something that is distal is farther away, or more *distant*, from the CNS. For example, the efferent nerves that project to the neck and shoulder muscles are proximal to the nerves that project to the hand muscles.

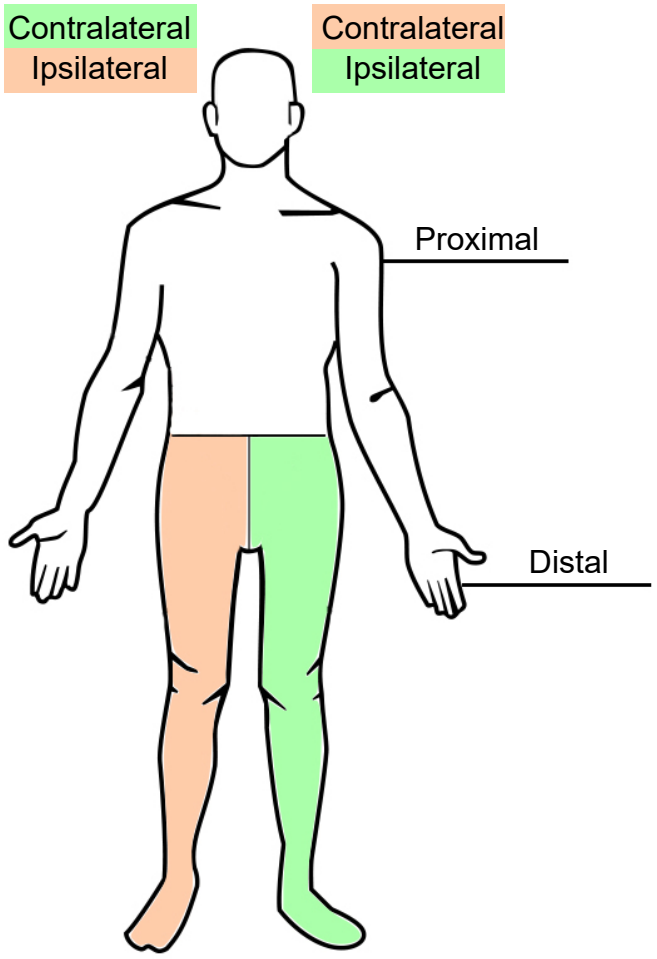


Figure 2.14 Anatomical language used in describing relationships from parts of the nervous system.

Another pair of anatomical terms that is useful in describing the nervous system is **contralateral** and **ipsilateral**. When the vertebrate nervous system evolved, there was a preference for a "crossed system," one where the left half of the brain generally controlled and received information from the right half of the body, and vice versa. This type of communication is called a contralateral connection. Most parts of the nervous system have this contralateral organization, where stuff on the left side of our vision first goes into the primary visual cortex of the right brain, physical sensations on the left hand first goes into the somatosensory cortex of the right brain, and so on. The opposite of contralateral is ipsilateral, and you would use this word to describe a part of the body that is on the same side as that half of the nervous system. For example, the right hand is ipsilateral to the right hemisphere of the spinal cord.

The PNS can be divided into three main branches:

1. Somatic nervous system

The somatic nervous system represents all the parts of the PNS that are involved with the outside environment, either in sensing the environment or acting on it. For example, the nerves that detect pressure or pain on the foot are part of the afferent somatic nervous system. We also think of the PNS as the branch that sends signals to our skeletal muscles. The nerves that innervate the muscles of the legs as we run are part of the efferent somatic nervous system. The somatic nervous system is also called the

"voluntary nervous system" since it is used to cause muscle movement related to intentional actions.

2. Autonomic nervous system

The autonomic nervous system encompasses all the branches of the PNS that deal with the internal environment. As with the somatic nervous system, the autonomic nervous system is comprised of nerves that detect the internal state as well as nerves that influence the internal organs. The body carries out all sorts of functions and responses unconsciously without any intentional control. It can do so by sending signals to smooth muscles and glands. The signals that cause us to sweat when it is hot, our pupils to dilate when it is dark, and our blood pressure to adjust when we stand up too quickly are all driven by the nerves of the autonomic nervous system.

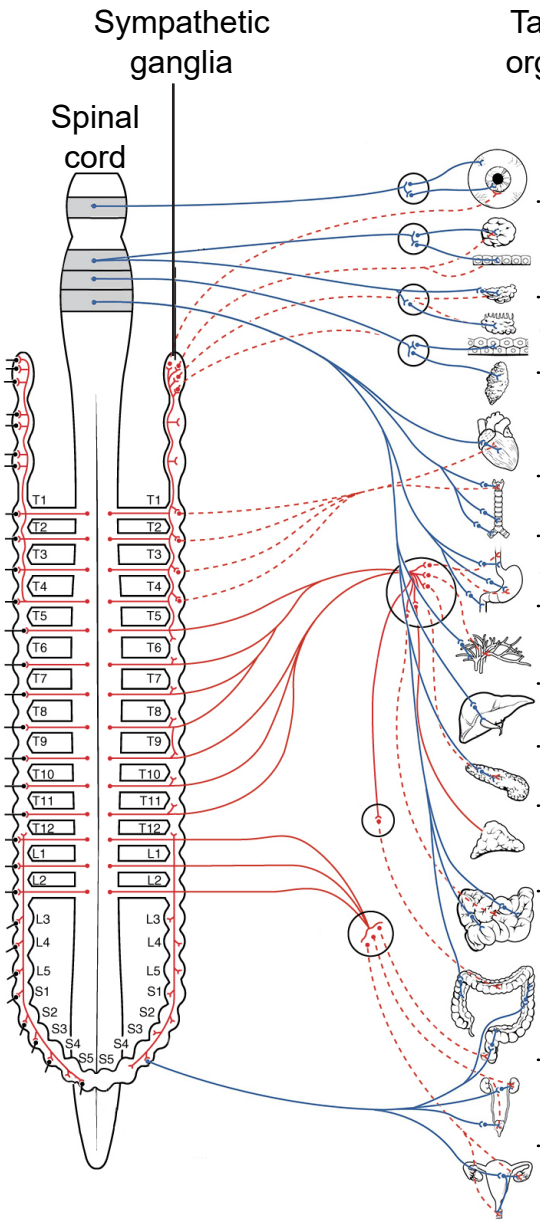
Imagine what would happen to your body if you, while walking through the quad, turned a corner and encountered a huge feral tiger, somehow set free from the Lincoln Park Zoo. You would notice a sudden fear response: your heart rate would skyrocket, you would start breathing more rapidly, and your body temperature might increase. Furthermore, your body will undergo additional reflexive responses that you might not even notice: your pupils will dilate, the bronchioles in your lungs will dilate, and your liver and kidneys will start to activate a variety of enzymes. All of these responses happen within seconds of seeing the tiger.

This set of physiological changes to a threat is sometimes called the **fight-or-flight response**, which is driven by one of the two branches of the autonomic nervous system, the **sympathetic nervous system**. The sympathetic nervous system activates when we are faced with a threat, either perceived or real. All of these rapid bodily responses result in the body preparing to attack or defend itself. Increased respiration allows the body to take in more oxygen, and dilation of blood vessels in the muscles allows that oxygen to get to the muscles, which is needed for muscle activation.

Now, consider a completely opposite scenario. You've just gorged yourself on a huge dinner of deep dish pizza, Italian beef sandwiches, bacon cheese fries, and tiramisu for dessert. You would probably feel relaxed, satisfied, and more than a little sluggish. A different physiological response is happening, a behavior called the **rest-and-digest** response. These physiological changes are driven by the other main branch of the autonomic nervous system, called the **parasympathetic nervous system**.

Both the sympathetic nervous system and parasympathetic nervous systems influence the internal organs simultaneously. At all times, the heart is getting signals from the sympathetic nervous system which increase heart rate, and signals from the parasympathetic nervous system which decreases heart rate. However, this seesaw-like balance can shift quickly in either direction, such as inducing a sympathetic response if a fearful stimulus is encountered, like the runaway tiger.

The sympathetic and parasympathetic nervous systems differ anatomically as well. The nerves that drive the sympathetic response branch off the spinal cord at the thoracic and lumbar levels, so sometimes we can use thoracolumbar to describe the site of origin. The nerves form a chain of many clumps of cells that run alongside the spinal cord, called the **sympathetic ganglion**. On the other hand, parasympathetic drive originates predominantly in the cervical spinal cord, with some signals originating in



Target organs	Sympathetic effect	Parasympathetic effect
Eye	Pupillary dilation	Pupillary constriction
Lacrimal gland Mucous membranes	↓ secretion	↑ secretion
Oral glands	↓ saliva production	↑ saliva production
Heart	↑ heart rate	↓ heart rate
Respiratory system	Dilate bronchi	Constrict bronchi
Stomach	↓ activity	↑ activity
Abdominal blood vessels	Vasoconstriction	Vasodilation
Liver	↑ glucose utilization	↓ glucose utilization
Pancreas	↓ activity	↑ activity
Adrenal gland	↑ norepinephrine release	
Small intestine Large intestine	↓ motility	↑ motility
Kidney + bladder	Inhibit urination	Stimulate urination
Gonads + genitalia	↑ Muscle constriction	↑ blood flow

Figure 2.15 Diagram showing innervation of different organs by the sympathetic and parasympathetic nervous systems and the corresponding effect. Note that the adrenal gland does not have parasympathetic activity.

the sacral areas. The parasympathetic nervous system usually receives signals from several cranial nerves. CN X, also called the **vagus nerve** innervates multiple bodily organs in the midsection of the body (vagus comes from the same root word as *vagrant*, appropriate since this nerve wanders all around the body.)

3. Enteric nervous system

The internal organs that carry out digestive functions, such as the esophagus, stomach, and intestines, are surrounded by a dense mesh of neurons that regulate their activity. Consisting of half a billion nerve cells, this net of neurons cause the digestive tract to increase or decrease the rate of these processes depending on the body's demands.

The enteric nervous system receives signals from both the sympathetic and parasympathetic nervous systems, and functions without our conscious knowledge. Historically, these digestive functions have been classified as part of the autonomic nervous system, but these responses do not share the same reflex pathway, and the enteric signals can work entirely independent of the vagus nerve, for example.

Cranial nerves

Many sensory and motor nerves originate in the spinal cord, but there are 12 pairs of nerves that exit from the brain. These **cranial nerves** can control motor functions, carry out general or specialized sensory functions, or both. The cranial nerves usually deal with parts related to the head, but not exclusively.

Table 2.16 The twelve cranial nerves and their function.

Cranial Nerve	Function	Description
CN I Olfactory nerve	Sensory	Sense of smell
CN II Optic nerve	Sensory	Sense of vision
CN III Oculomotor nerve	Motor	Control of extraocular muscles which allow movement of eyeballs; constriction of pupils; changing of lens shape
CN IV Trochlear nerve	Motor	Control of the superior oblique muscle of the eye that moves the eyeball down and lateral
CN V Trigeminal nerve	Sensory + motor	Tactile and pain sensory information from the face and mouth; Control of muscles used in chewing
CN VI Abducens nerve	Motor	Control of the lateral rectus muscle of the eye that moves the eyeball outward laterally
CN VII Facial nerve	Sensory + motor	Control of the muscles that allow for facial expressions; Taste sensation on the anterior 2/3rds of the tongue
CN VIII Vestibulocochlear nerve	Sensory	Detection of sound information and head positional (vestibular) information
CN IX Glossopharyngeal nerve	Sensory + motor	Detection of somatic sensory in the middle ear and posterior 1/3rd of the tongue; Taste sensation on the posterior 1/3rd of the tongue; Controls the stylopharyngeal muscle that allows swallowing;
CN X Vagus nerve	Sensory + motor	Control of the internal organs by autonomic nervous system using parasympathetic activity
CN XI Accessory nerve	Motor	Control of the sternocleidomastoid and trapezius muscles of the neck and shoulders
CN XII Hypoglossal nerve	Motor	Control of the muscles of the tongue

2.3 Support structures of the nervous system

Although we mostly think about nerve cells as being the main characters of the nervous system, there are many other anatomical features that play important supporting roles. These are often non-neuronal structures, but are still critically important in allowing the nervous system to do what it needs to do.

Brain circulation and cerebral blood flow

Like every other organ in the body, the brain requires oxygen and nutrients to function. In humans, this function is accomplished by the blood that is pumped around the body using a network of blood vessels called the circulatory system. The brain has a very high demand for oxygen and nutrients: at only 2% of total body weight, it receives about 15% of total cardiac output.

Oxygenated blood moves anteriorly through the aorta into the brain via two different pairs of arteries, the vertebral arteries and the internal carotid arteries. The left and right vertebral arteries merge into a single basilar artery, and along with the left and right internal carotid arteries, feed into a loop-like circular blood vessel called the **circle of Willis**. According to one theory, the circle of Willis is an anatomical redundancy that allows an organism to maintain cerebral blood flow even if one part of the arterial system is narrowed or blocked. From here, the circle of Willis has paired "exit" arteries that distribute blood to different areas of the brain: the **anterior cerebral arteries** provide blood to dorsomedial cortical structures and deep brain structures, and the **posterior cerebral arteries** provide blood to the occipital lobe. The other major arteries are called the **middle cerebral arteries** which branch off the internal carotid artery and deliver blood to the lateral cortices.

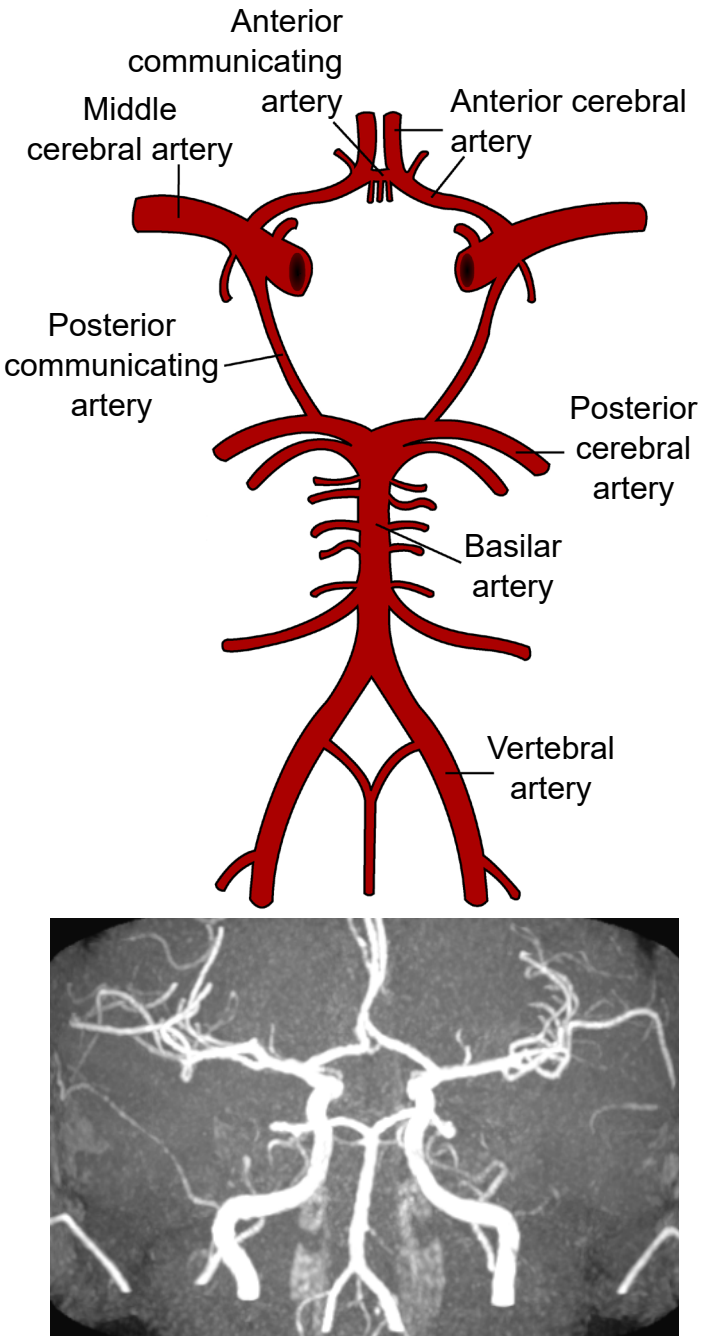


Figure 2.17 Diagram of the circle of Willis (top) and an angiogram (bottom) showing the structure in real life.

Clinical connection: Stroke

Stroke is an extremely common, life-threatening medical condition that results in a loss of blood flow to the brain. According to 2016 statistics from the World Health Organization, stroke is the second highest cause of death worldwide. The number one risk factor for stroke is high blood pressure.

There are two common types of strokes that a person may experience. More than 80% of all strokes are **ischemic strokes** (pronounced *is-keemik*), which happens when normal blood flow is interrupted, causing cell death by deprivation of brain tissue of oxygen and nutrients. Generally, this type of injury can happen when a blood clot forms, travels through the circulatory system and gets lodged in a tiny brain blood vessel, thus blocking the passage of blood.

The other 20% of strokes are **hemorrhagic strokes**, which results from a burst blood vessel that causes bleeding into the brain. The presence of uncontrolled blood inside the brain causes an increase in intracranial pressure, which can be lethal. Many brain cells may die since they cannot take up oxygen directly from the blood. Additionally, blood has dramatically different properties than the normal solution which the brain cells live in, and this can cause the neurons to trigger a self-destruction program. Generally, hemorrhagic stroke is more deadly than ischemic stroke.

Because the different blood vessels of the brain's circulatory system are responsible for providing blood to specific areas of the brain, it is possible to diagnose the specific area where the stroke is happening based on the presentation of symptoms. For example, if the middle cerebral artery blood is occluded by an ischemic stroke, the left hemisphere motor cortex will lose blood flow. Because of the contralateral organization of the descending motor pathway, the patient may therefore present with paralysis or weakness in the right half of the body.

It is vitally important to correctly diagnose and differentiate between the two types of strokes. An ischemic stroke may be treated with injection of a "clot-busting" drug, a substance that helps the body break down the offending blockage. However, these clot-busters could make the bleeding from a hemorrhagic stroke even worse.

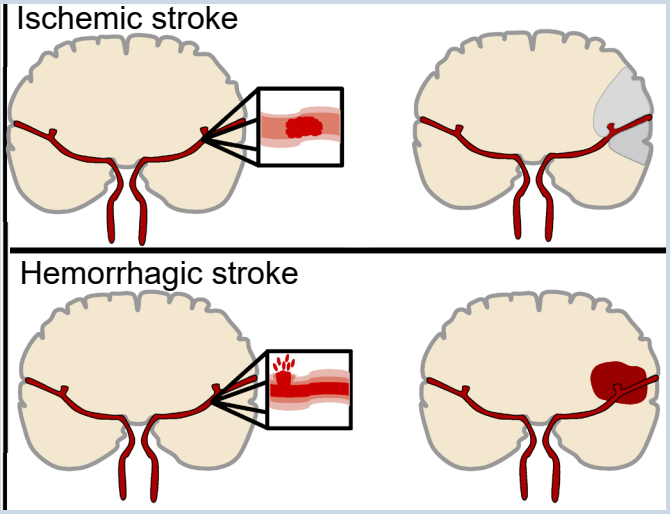


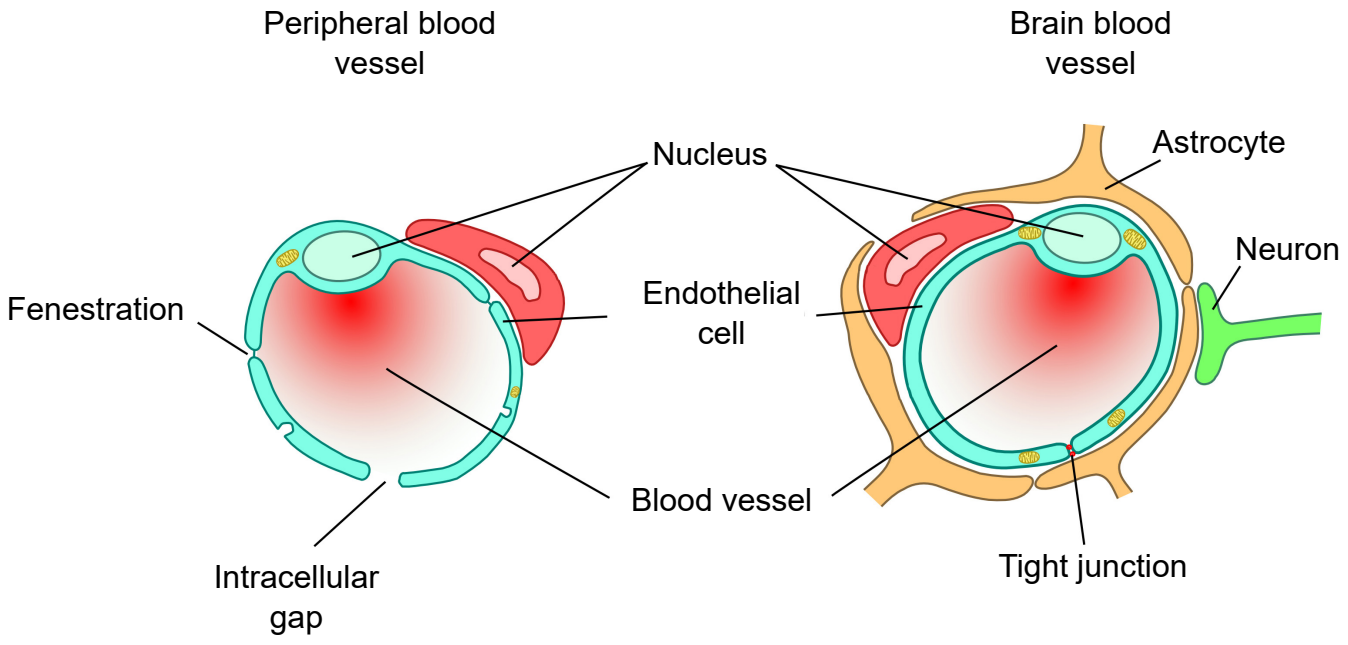
Figure 2.18 Diagram illustrating the two main types of stroke and the effect on blood flow to the brain

Blood-brain barrier (BBB)

It is important for oxygen and nutrients to pass from the blood into the brain tissue. Small blood vessels outside of the brain, such as the capillaries in the fingertips, have very thin walls, sometimes the width of a single cell, and are therefore highly permeable to gases. These vessels can either contain tiny holes or large protein structures that physically transport substances across the blood vessel. On the other hand, it is also advantageous to separate toxins and foreign pathogens in the bloodstream from getting into brain tissue. The **blood-brain barrier (BBB)** is an anatomical adaptation that selectively transports substances necessary for normal biological function, while simultaneously excluding potentially harmful invaders from the brain. The BBB physically surrounds blood vessels in the brain. It is made up of endothelial cells and a type of glial cell called an **astrocyte**. The BBB is injured in all variety of medical disorders, ranging from stroke, epilepsy, and Alzheimer's disease, just to name a few. It is still unknown what role the disruption of the BBB plays in brain disorders.

The exclusive nature of the BBB can be a double-edged sword. It is difficult to deliver a drug into the brain from the blood stream if that drug is unable to pass through the BBB. For example, the current gold standard pharmaceutical treatment for Parkinson's disease is to increase the brain's levels of dopamine. However, dopamine does not pass through the BBB. To get around this, physicians give the BBB-permeable substance L-DOPA, which the brain is able to convert into dopamine. Many other therapeutic drugs do not cross the BBB, so researchers are developing methods using electromagnetic fields to temporarily weaken the barrier, or surround the drugs in nanoparticles, which are so small that the body cannot identify as foreign.

Figure 2.19 The blood-brain barrier separates blood vessels from the brain tissue. Peripheral blood vessels do not have such an anatomical adaptation.



The ventricles and cerebrospinal fluid (CSF)

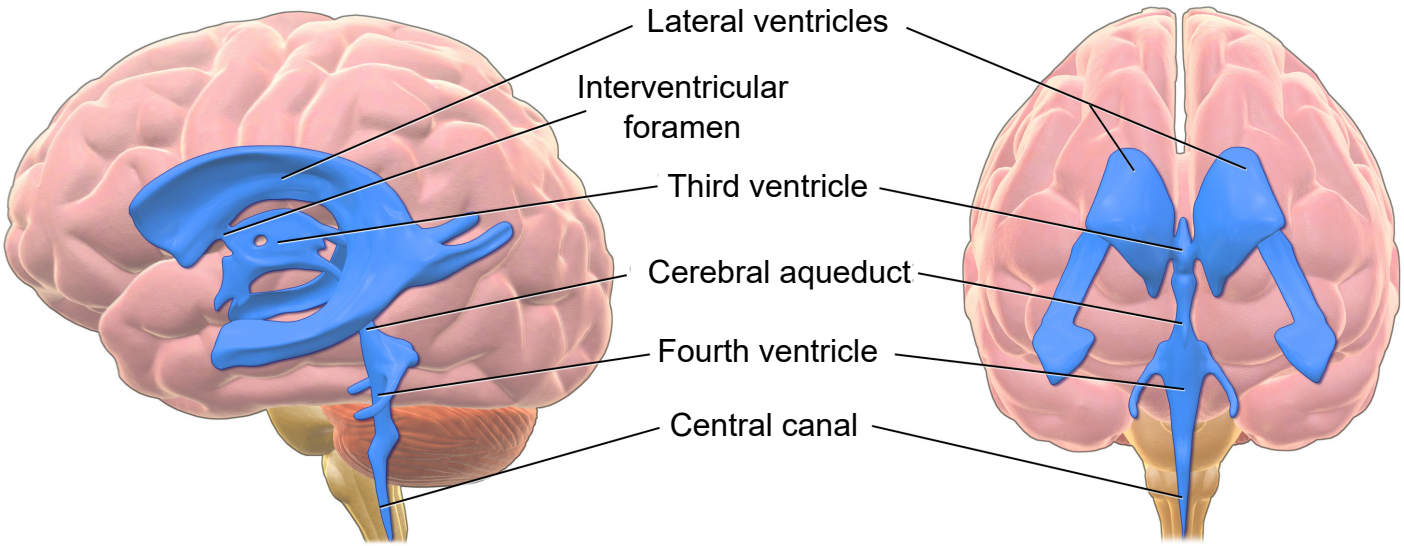
In a coronal section of the brain, it is very common to see large symmetrical “holes” near the medial aspect. These spaces are called **ventricles**.The human brain has a total of four ventricles in the brain. The two very large paired ventricles, one in each hemisphere, are the **lateral ventricles**. They are connected medially into the **third ventricle**, which extends to the posterior aspect of the brain. From here, an aqueduct that runs ventrally extends into the **fourth ventricle** before continuing into the **central canal**, a narrow space that runs all the way through the length of the spinal cord along the midline. The entire ventricular system is entirely connected.

The ventricles are filled with a liquid called **cerebrospinal fluid (CSF)**. CSF is basically a high salt water solution (sodium ions at 140 mM and chloride ions at 110 mM). Because of the high osmolarity of CSF, it is a very buoyant solution. Like a fully grown person who can float easily on the surface of the extremely salty Dead Sea, CSF allows the brain to remain “floating” inside the skull. Without CSF, the brain weighs almost 1.5 kg (~3 lbs). Cells and blood vessels at the ventral base of the brain would be crushed under the weight of the brain itself. But when the brain is surrounded by CSF, it weighs less than 50 grams, almost two orders of magnitude lighter!

CSF is also found in a tightly-regulated membranous sac called the **meninges** (more about it below) that encases the brain. In fact, more than 80% of the CSF in the body exists in this space outside the brain. This liquid serves as a form of “cushioning” that protects the brain from rapid he ad movements. If it weren’t for this physical protection, the inertia of head movement may cause your brain to smash against the inside of the rigid skull if you move your head too quickly. The CSF layer allows the head to withstand some sloshing of the brain, but a movement that is too abrupt can cause a traumatic brain injury, a condition that we will return to in chapter 15.

CSF can also function as a way to wash impurities out of the brain. The volume of CSF in

Figure 2.20 The ventricular system consists of several interconnected chambers filled with cerebrospinal fluid (CSF).



Clinical correlation: Hydrocephalus

A common condition affecting the brain of about 1 in 200 newborns and a small number of adults is hydrocephalus, historically called “water on the brain.” In patients with hydrocephalus, the volume of CSF increases, which elevates intracranial pressure, causing symptoms such as fever, stiff neck, headache, seizures, or altered mental status. In adults, the skull is rigid and unmoving. But in newborns with hydrocephalus, the plates of the skull are not completely fused together. Often, these children will have a bulging parts on the skull and an expansion of the forehead. Increased CSF volume can happen in a couple ways. The clearance of CSF may fail while production remains normal, or the entrance to the central canal in the spinal cord may be narrowed or blocked by a tumor, leading to an increase in the volume in the brain. A common treatment for hydrocephalus is to surgically implant a shunt, which is a hollow tube that runs from the ventricle down into the abdominal space, which allows for drainage thus decreasing intracranial pressure.

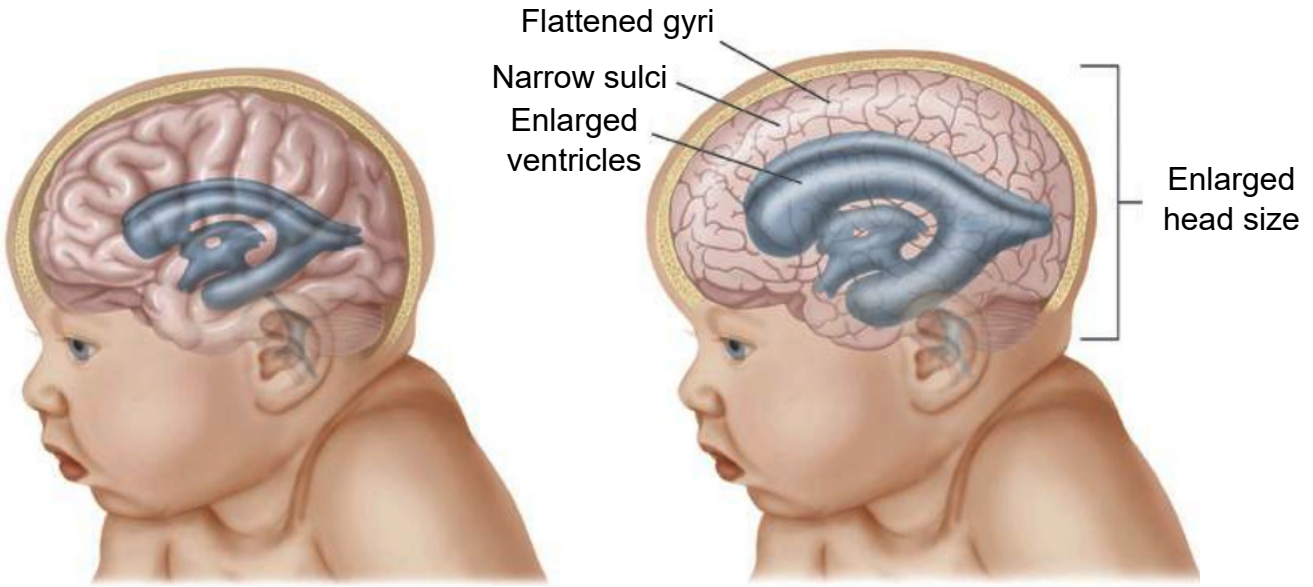


Figure 2.21 Hydrocephalus is one of the most common birth defects, affecting around 0.1% of births in the United States. It can also affect adults. Hydrocephalus can be deadly.

the typical human body is about 150 mLs, a little more than half a cup. Because there is frequent turnover of CSF, the material gets absorbed back into the body regularly. Each day, the body produces about half a liter of CSF, so the brain cycles through the entire volume a few times each day. Since the CSF is in close association with the neurons, cellular waste materials get dissolved into CSF, which can then be degraded and broken down outside of the brain.

Meninges

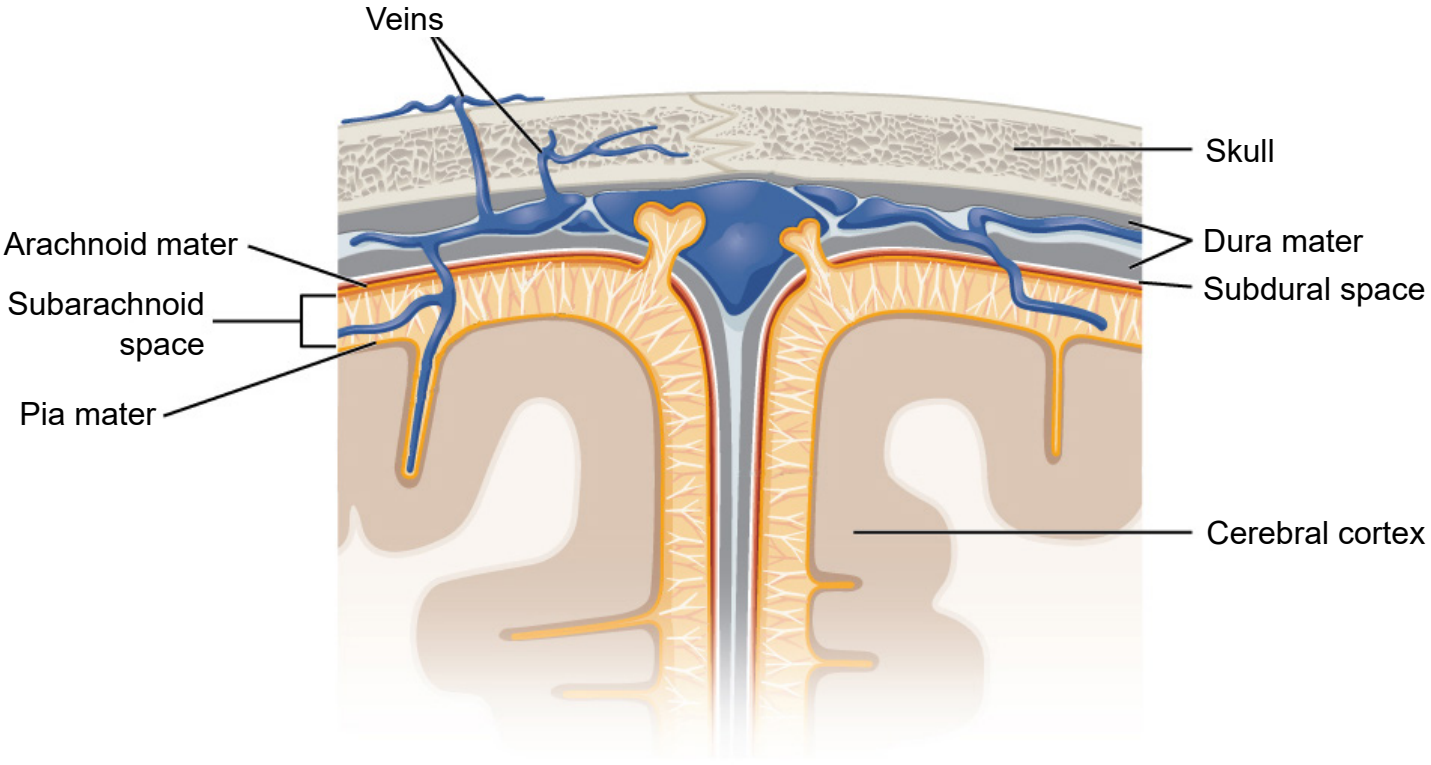
The brain is a squishy internal organ housed inside the skull. If there weren't some protective buffer separating the soft brain matter from the rigid bone, the jelly-like brain would be smashed up against the inside of the skull and get injured as the head moves around. The meninges are a series of protective membranes that minimize this kind of damage. They surround the brain and extend all the way down the spinal cord. Think of the meninges as an organic type of "bubble wrap" that encases a fragile nervous system.

There are three types of membranes that collectively make up the meninges. From the outside-most layer to the innermost, they are:

1. Dura mater. The dura is made of thick, fibrous material, and can get to be 0.8 mm thick in the adult body (if you took a piece of printer paper and fold it in half four times, that should give you an idea of how thick the dura mater is in the adult human). The dura mater is physically attached to the inside of the skull with highly resilient connections found at the sutures between the plates of the cranium. The name originates from Latin meaning *tough mother*.

2. Arachnoid mater. The arachnoid mater is the middle layer of the meninges. The fibers are very delicate and resemble a spider web, which is where the name comes from. Within this space, there are protrusions that allow for CSF to drain into sinuses, which allow for recycling of soluble substances. Most of the CSF in the brain exists underneath this layer in the *subarachnoid space.*

Figure 2.22 The meninges are a series of protective membranes that surround the CNS.



3. Pia mater. The pia mater is the third layer of the meninges. It is very fragile and in direct contact with the surface of the brain, and closely follows the sulci and gyri. The name means pious mother.

Clinical correlation: Meningitis

Inflammation of the meninges is a potentially deadly condition called meningitis. Exposure to infectious agents like viruses or bacteria such as Neisseria meningitidis that leaks from the blood into the meninges is a common cause of the inflammation. When the meninges are inflamed, the brain gets compressed from all sides, increasing intracranial pressure, producing many of the same symptoms seen in hydrocephalus: fever, stiff neck, headache, seizures, and altered mental status. The *N. meningitidis* bacteria and the viruses are highly transmissible in close contact, but vaccinations are highly effective at minimizing the infection rate. As with bacterial infections, broad-spectrum antibiotics are effective at treating the infection.

The Open Neuroscience Initiative is funded by a grant from the Vincentian Endowment Fund of DePaul University.

Cover: http://www.neuroanatomy.ca/

2.1 File:1201 Overview of Nervous System.jpg: OpenStaxderivative work: GnolizX [CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)] Changes made by Austin Lim.

2.2 Helixitta [CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0)] Changes made by Austin Lim

2.3 Looie496, Originally drawn by François Leuret and Louis Pierre Gratiolet [Public domain]

2.4 BruceBlaus. When using this image in external sources it can be cited as: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)] Changes made by Austin Lim

2.5 neuroanatomy.ca

2.6 [[::User:Nrets|Nrets]] [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0/)] Changes made by Austin Lim

2.7 Mysid, arrows were added by Was a bee [Public domain] Changes made by Austin Lim
https://commons.wikimedia.org/wiki/File:PSM_V46_D168_Mesial_view_of_the_human_brain.jpg changes made by Austin Lim

2.8 Originally from the collection of Jack and Beverly Wilgus, and now in the Warren Anatomical Museum, Harvard Medical School. [CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0)], John M. Harlow, M.D. [Public domain]

2.9 Internet Archive Book Images [No restrictions]

2.10 vsion and JmarchnThis SVG image was created by Medium69.Cette image SVG a été créée par Medium69.Please credit this : William Crochot [Public domain]. Changes made by Austin Lim.

2.11 Internet Archive Book Images [No restrictions]

2.12 OpenStax [CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)]. Changes made by Austin Lim.

2.13 BruceBlaus [CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0)]. Changes made by Austin lim.

2.14 Rhcastilhos [CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0)] . changes made by Austin Lim.

2.15 OpenStax College [CC BY 3.0 (https://creativecommons.org/licenses/by/3.0)], OpenStax College [CC BY 3.0 (https://creativecommons.org/licenses/by/3.0)], changes made by Austin Lim

2.17 Rhcastilhos [Public domain], Ceccomaster [CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0)], changes made by Austin Lim

2.18 ElinorHunt [CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0)], changes made by Austin Lim

2.19 Kuebi = Armin Kübelbeck [CC BY 3.0 (https://creativecommons.org/licenses/by/3.0)], changes made by Austin Lim

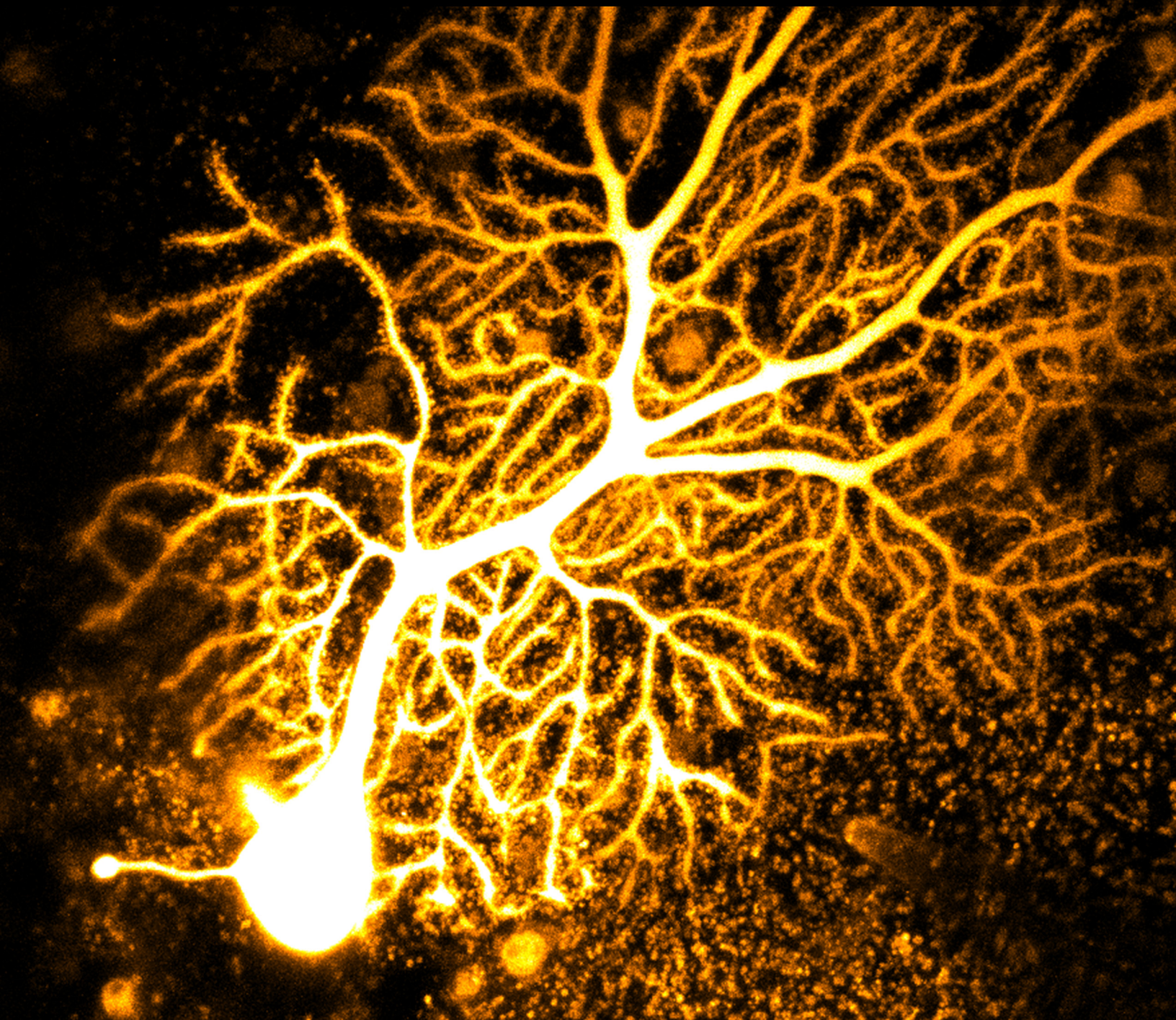
2.20 BruceBlaus [CC BY 3.0 (https://creativecommons.org/licenses/by/3.0)], changes made by Austin Lim

2.21 https://upload.wikimedia.org/wikipedia/commons/9/95/Hydrocephalus_CDC.png, changes made by Austin Lim

2.22 https://upload.wikimedia.org/wikipedia/commons/4/4c/1316_Meningeal_LayersN.jpg , changes made by Austin Lim

Chapter 3:

Cellular Anatomy of the Nervous System



Austin Lim, PhD (DePaul University)
Editor: Katie Willis, PhD (University of Oklahoma)

For the majority of human history, the only way we were able to study the structure of the brain was with crude, butcher-like methods. Wait for a person to die, saw a giant hole in the top of the skull, take the brain out, and slice it into pieces to see if there was some correlation between the way the brain looks and the way they died. With these methods, only major changes in gross anatomy could be observed, such as that resulting from severe birth defects or trauma.

Brain analysis methods were enhanced by the scientific adoption of light microscopy. Naturalists in the mid 1600s such as Antonie van Leeuwenhoek, Jan Swammerdam, and Robert Hooke began looking at biological substances up close, and the brain proved to be a complex and interesting sample of tissue.

A major advancement in the study of neuronal morphology came about in the late 1800s. The Italian anatomist and biologist Camillo Golgi identified a shortcoming with the cellular analysis techniques of the time: structures in the CNS were impossible to distinguish from one another. The cells in the brain were so densely packed together, that it became difficult to identify which cellular material belonged to which cell. Golgi came up with a new technique using a silver compound that caused the silver to precipitate inside the cell membranes. However, not every cell took up the silver. Instead, only a small fraction of neurons, maybe 1% or even less, were completely stained in black, which stood out remarkably well against the light yellow background of the surrounding tissue. This reaction, initially called the "black reaction," is now known as a "**Golgi stain**." (Despite being more than a hundred years old, we currently don't know the mechanism by which the silver stain is taken up into the neurons, or what determines why certain cells take the stain and others don't.) Because of the great contrast between cell and background, every single part of the neuron was completely filled, allowing Golgi to do drawings of the morphology of this nervous tissue. Based on his staining results, Golgi supported the idea that

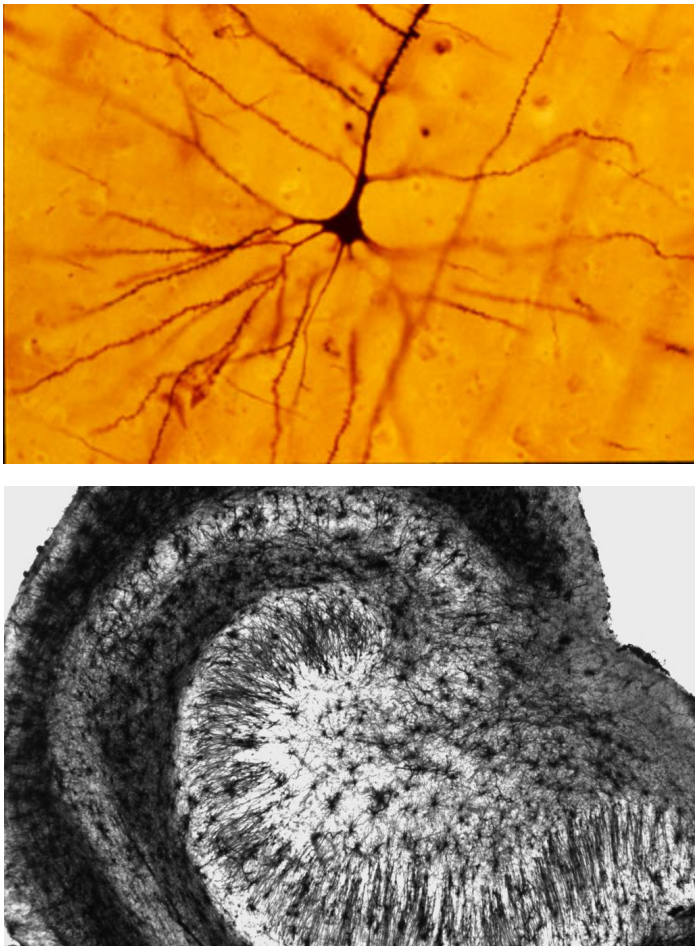


Figure 3.1 A microscope image of a single Golgi stained neuron (top) and a slice of Golgi stained brain (bottom).

the parts of the nervous system are all one very large, physically connected network. This idea was known as the **reticular theory**.

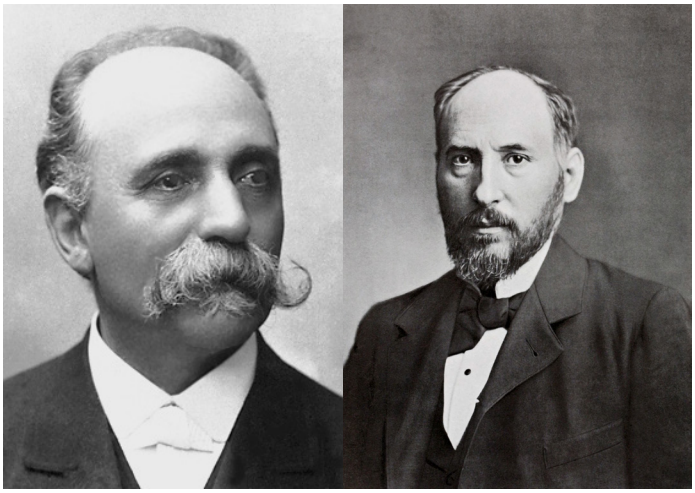
About 10 years later, the Spanish neuroanatomist Santiago Ramon y Cajal repeated some of Golgi's staining experiments with other sections of nervous tissue. Looking at similar darkly-filled neurons, Cajal arrived at a different conclusion: the nervous system is not a giant net, but rather a series of individual units that are separated from one another physically. This idea came to be known as the **neuron doctrine**.

Both Golgi and Cajal were awarded the shared Nobel Prize in Physiology or Medicine in 1906 for their accomplishments in helping to understand "the structure of the nervous system." Even though Cajal's neuron doctrine was adopted widely by scientists, the elucidation of this organization would not have been made possible without Golgi's development of the silver stain. The sharing of this prestigious award was ironic because of the many disagreements between the two scientists. Cajal commented on their relationship, saying:

"What a cruel irony of fate of pair, like Siamese twins united by the shoulders, scientific adversaries of such contrasting character!"

Cajal's neuron doctrine was eventually given more support with the aid of modern techniques, like electron microscopy, that are capable of physically seeing the distance between two neurons. The neuron doctrine represents our current understanding of how the nervous system is organized, and this chapter is focused on describing the anatomical features of the nervous system at the cellular level.

Figure 3.2 Camillo Golgi (left) and Santiago Ramon y Cajal (right) were both awarded the joint Nobel Prize in Physiology or Medicine in 1906.

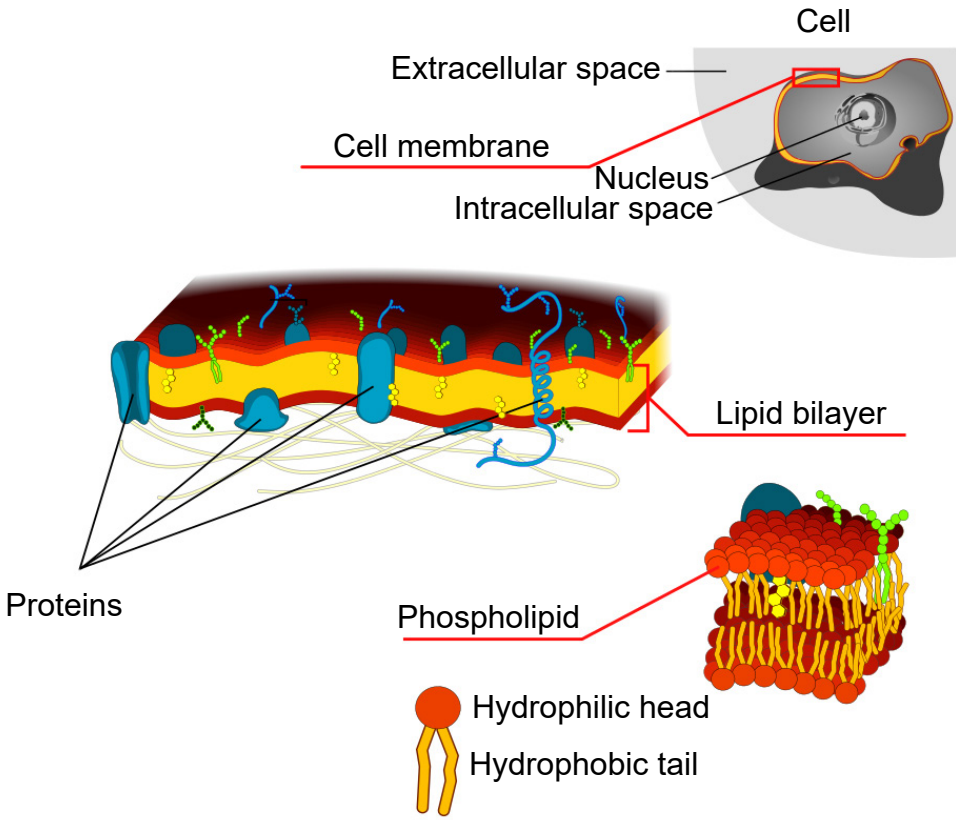


Chapter 3 outline

3.1 Characteristics of neurons

The main units of the nervous system are cells called **neurons**. Although neurons do have a variety of adaptations that make them unique from other types of cells in the body, they are still cells. Therefore, they contain all of the basic features of a typical mammalian cell. For example, they are made up of an aqueous cytoplasm bounded by a **cell membrane**. This cell membrane, also called a plasma membrane or lipid membrane, consists of a sheet of several individual molecules called phospholipids, which consist of two *hydrophobic* (water-fearing) tails and a *hydrophilic* (water-loving) end. These phospholipids arrange themselves into a bilayer with the hydrophilic tails touching each other and the hydrophilic sides facing the cytoplasm and the extracellular space, which are both mostly water. Because of the chemical properties of the cell membrane, it is very effective at keeping ions and charged molecules separated, while allowing small molecules like water and oxygen across the cell.

Figure 3.3 The cell membrane is made up of a lipid bilayer, which consists of organized phospholipids.



They also have all the organelles that you would see in other cell types, like a nucleus and mitochondria.

The number of neurons in the adult human brain, according to our current best estimate, is close to 86 billion. This number was calculated using a revolutionary technique, the isotropic fractionator or "brain soup" developed by Brazilian neuroanatomist Suzana Herculano-Houzel. To put this number in context, we

have about 37 trillion cells in the whole body, so neurons in the brain make up about 0.2% of all cells in the body.

Below are some unique characteristics that neurons have in common.

Neurons are electroactive

The inside of most cells have a negative electrical charge compared to the solution outside of the cell. This difference between the electrical charges is called the **membrane potential**, or **transmembrane potential**. When a neuron is at rest, the amplitude of this charge may be somewhere between -50 millivolts (mV) and -90 mV. Often times, we say that in general the cell rests at around -70 mV. Membrane potential is abbreviated as V_m.

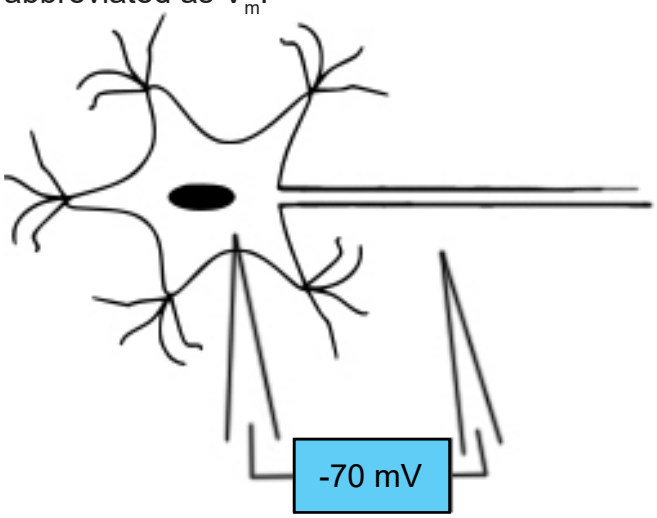


Figure 3.4 Compared to the extracellular space, the voltage of the neuron is generally negative.

Most neurons do not spend their entire lives at -70 mV. Instead, neurons have special proteins embedded in their cell membranes that allow for charged ions, such as sodium or chloride, to move into or out of the cell. When a net positive charge enters into the cell, the Vm becomes more positive. Likewise, when net negative charge enters the cell, the membrane potential becomes more negative. The membrane potential of a neuron can go from -70 mV to +45 mV and back to -70 mV in as quickly as two milliseconds!

Neurons are not the only electroactive cells in the body. Our lives depend on cells in the heart that are also capable of changing in potential, as they respond to rhythmic electrical impulses that drive our heartbeat. This is why being struck by lightning may cause your heart to desynchronize in activity, and why a shock from a defibrillator can electrically “kick-start” the heart back to a meaningful pattern.

Neurons are specialized for rapid communication

Many cells are capable of sending and receiving chemical signals across long distances and time scales. But neurons are able to communicate with a combination of electrical and chemical signals in a matter of milliseconds. Additionally, the shape of neurons and the organization of the neurons on a microscopic level make them effective for sending signals in a very specific direction. Many neurons have an incoming receiving end and an outgoing sending end. The placement of one neuron next to the correct partner is very important, and many chemical signaling systems are in place to ensure that the developing nervous system is properly wired together.

Neurons are “forever” cells

We are constantly replacing non neuronal cells. For example, the cells in our bones replace themselves frequently at a rate of about 10% each year. Our body makes new skin cells to replace the dying skin cells on the surface so that we have a “new” skin every month. The cells along the inside of our stomachs, exposed to very harsh acidic conditions, get replaced about every week. About 100 million new red blood cells are created every minute!

On the other hand, the mature nervous system generally does not undergo much **neurogenesis**, the creation of new neurons. The neurons that we have after development are the ones that we will keep until we die, and this permanence of neuronal count makes them different from almost every other cell of the body. However, the idea of adult neurogenesis is a topic of debate among neuroscientists, since some areas like the olfactory system and the hippocampus display new nerve cell production.

Figure 3.5 The handle of George Washington's axe breaks and is soon replaced. Later, the head breaks as well. Is the axe still George Washington's axe? This paradox is similar to the cells of our body, which are replaced frequently - with neurons being the exception.

...But, neurons can change

Even though new neurons are not created in most areas of the brain, neurons still have the capability to change in their structure and function. Some of these changes, such as physical changes to the structures of the input sites of the neurons, are believed to last for a lifetime. We use the word **plasticity** to describe the ability for the brain to alter its morphology (derived from Greek plastikos, meaning "capable of being shaped or molded" - think of plastic surgery, where a person changes their physical appearance; see chapter 13, Leaning and Memory).

Also, neurons do have the capacity to repair themselves to some extent. Neurons of the PNS may get injured or completely destroyed as a result of trauma to the body. Afterwards, those injured neurons can regrow to connect once again with their original partner. This regrowth seems to depend on a few chemical signals that the body produces, such as nerve growth factor and brain derived neurotrophic factor. However, this process is often very slow, and does not always successfully restore the nervous system to the way it was pre-injury.

3.2 Cellular anatomy of neurons

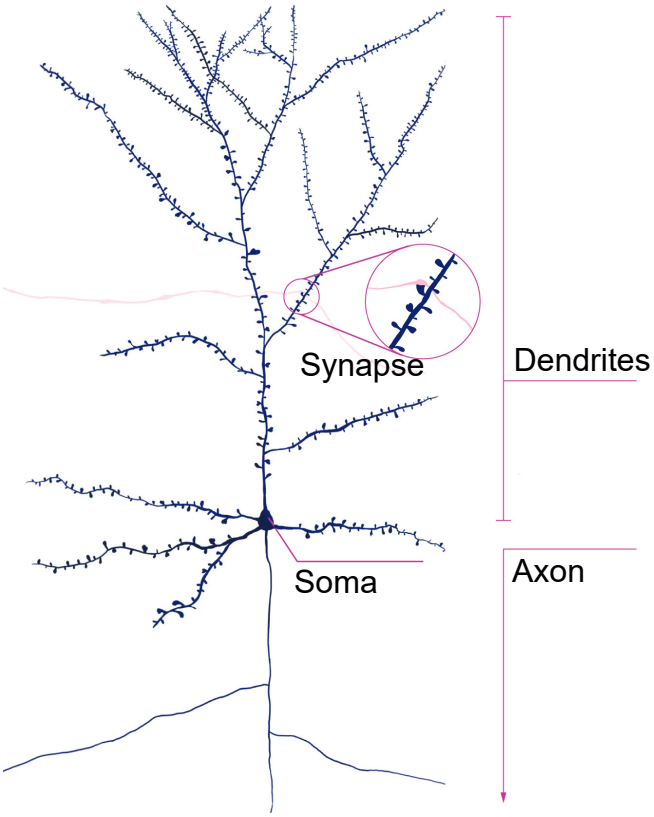


Figure 3.6 The basic anatomy of a neuron.

The main function of neurons is to use changes in electrical properties in order to communicate with connected cells. This communication usually moves in one direction, and we will use this pathway as an outline for discussing the anatomical structures of the neurons.

Dendrites

Information first enters the neuron through the **dendrites**, the branch-like extensions that protrude from the cell body. The word dendrite originates from the ancient Greek word “dendro-”, meaning “tree” (think of rhododendron). Dendrites look like the branches of a tree: they reach outward away from the center of the cell body, generally getting thinner the farther away you look.

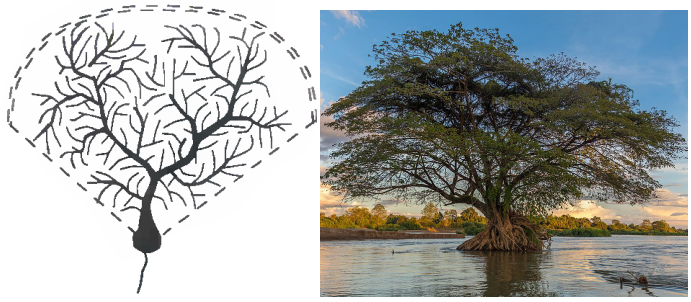


Figure 3.7 The dendrites are the highly branched input sites of the neuron. They are named because of their resemblance to trees.

Along the dendrites of some neurons, it is possible to see tiny protrusions of cell membrane that stick out from the main dendrite. These bumps are called **spines**. Spines can roughly be classified based on their approximate shape; thin, mushroom, and stubby being some of the more easily recognizable forms. A spine may be about 100 nm in diameter, making it smaller than the wavelength of visible light, with a total volume of about 0.1 femtoliter - one ten-quadrillionth of a liter!

Chemical signals released by another cell are received by the dendritic spines, and so each spine may represent an input site of communication. Some cells, like a pyramidal neuron in the hippocampus, may have more than 30,000 spines, indicating that one cell may detect information from several incoming cells.

We believe that spines are one of the most important sites where the nervous system is able to change. For example, neurons change shape after exposure to various environmental conditions, such as stress or exposure to drugs. Tiny changes to the surface of the neuron at the level of dendritic spines is an example of plasticity. Dendritic plasticity is thought to underlie the reason that we can learn new facts or maintain

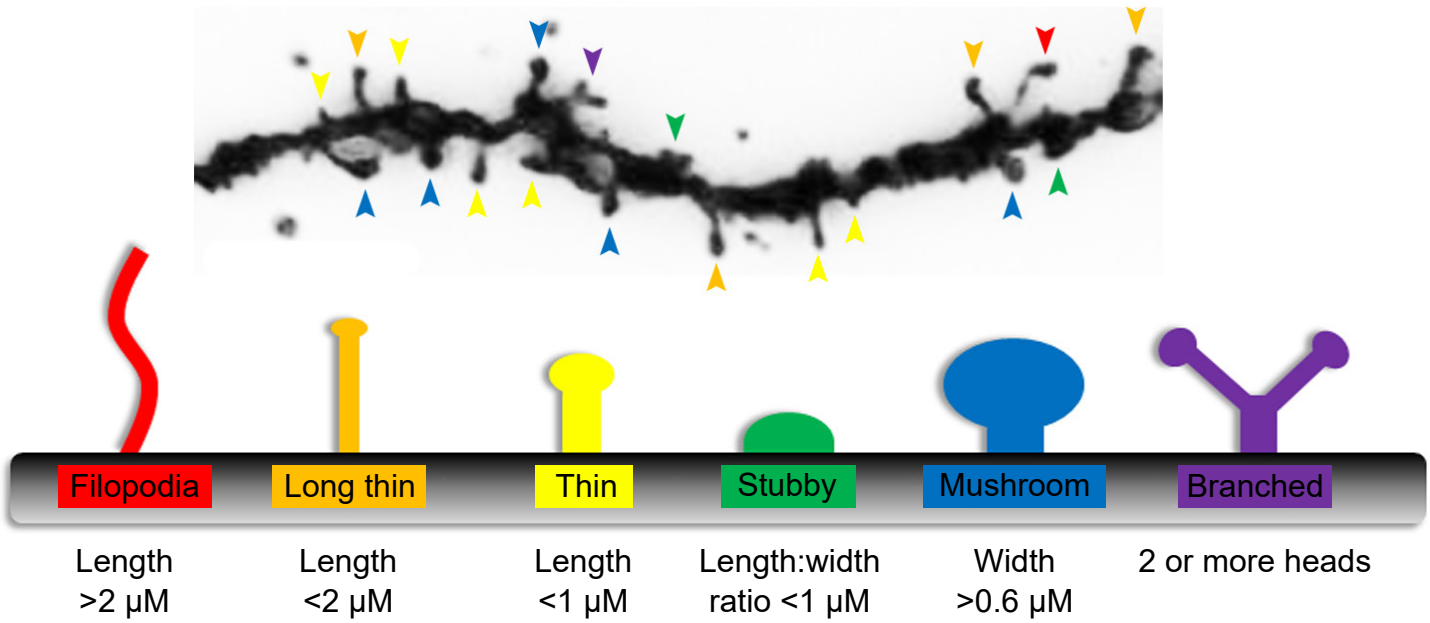


Figure 3.8 Tiny extensions of the cell membrane on dendrites are called spines. They can be classified based on their shapes and sizes.

memories about our childhood over long periods of time. Some set of tiny, submicroscopic changes to the morphology of dendritic spines may represent a single complex memory that you form.

A neuron does not need spines for receiving information or for plasticity to take place. Many cells lack spines, but are still capable of permanently changing. The input site may be anywhere along the dendrite, or even at the cell body, the "center" of the neuron.

Cell body (soma)

Information that arrives through the many dendrites of a neuron eventually filters into the **cell body**, or the **soma**, of the neuron. Somata (plural of soma) vary in size across different types of neurons, the largest somata belonging to the Betz cells of the motor cortex with an area upwards of 100 microns in diameter: about the size of a single grain of salt. On the opposite end of the spectrum, granule cells of the cerebellum are so densely packed so that they make up more 70% of all neurons in the brain, maybe 4 microns in diameter.

The cell body contains many of the organelles that are essential for the production of proteins that the neuron needs. Many of these organelles are not unique to neurons, and can be found in other cell types. The most apparent organelle visible under high magnification (400x) is the **nucleus**, which houses DNA and other genetic material. From our understanding of the central dogma of molecular biology, this DNA is transcribed into a string of single-stranded genetic code called **messenger RNA (mRNA)**, which is exported out of the nucleus. This mRNA is then used as a guide for the synthesis of proteins.

The next step of protein creation depends on organelles that are adjacent the nucleus. Physically continuous with the membrane that surrounds the nucleus is a folded membranous organelle called the **endoplasmic reticulum** (ER). Attached to the ER are several **ribosomes**, which are the molecular machines that read the mRNA and translate that code into proteins. The **Golgi apparatus** are layers of folded plasma membranes that function in transport. They

are found near the nucleus, although small protrusions of these organelles may reach into the other parts of the neuron.

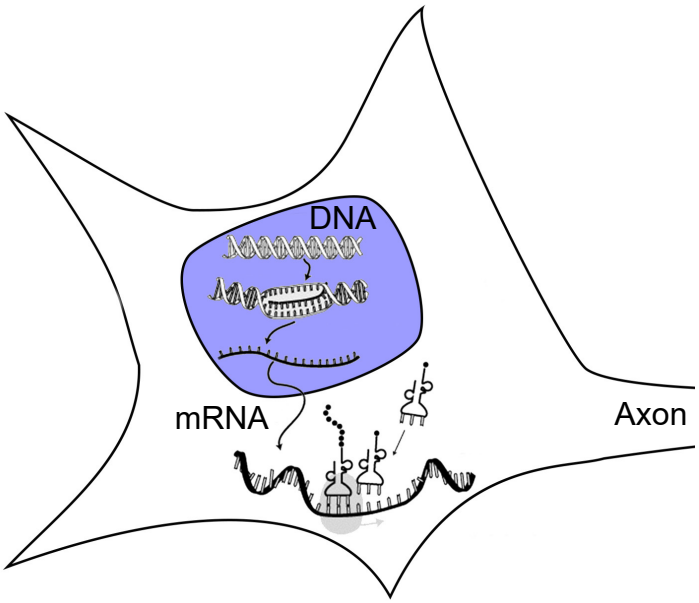


Figure 3.9 Like non-neuronal cells, DNA is housed in the nucleus (purple). DNA is transcribed into mRNA and exported out of the nucleus, where it can be translated into protein.

Axon

The **axon** is the main output extension of the neuron. While neurons only have a single axon extending from the cell body, this axon can branch several times after exiting the soma. In branching, an axon from a single neuron is able to communicate with many other neurons at the same time. Axons are usually thinner than dendrites, some being only a micron in diameter. Several axons can bundle and travel together; these are **nerves**. Axons can be very long; the longest axon in the human body is part of the sciatic nerve that runs from the posterior end of the spinal cord down the leg to control the muscles of the big toe.

The first section of the axon, the junction between the cell body and the beginning of the projection, is a patch of axonal membrane with very unique characteristics. This section is

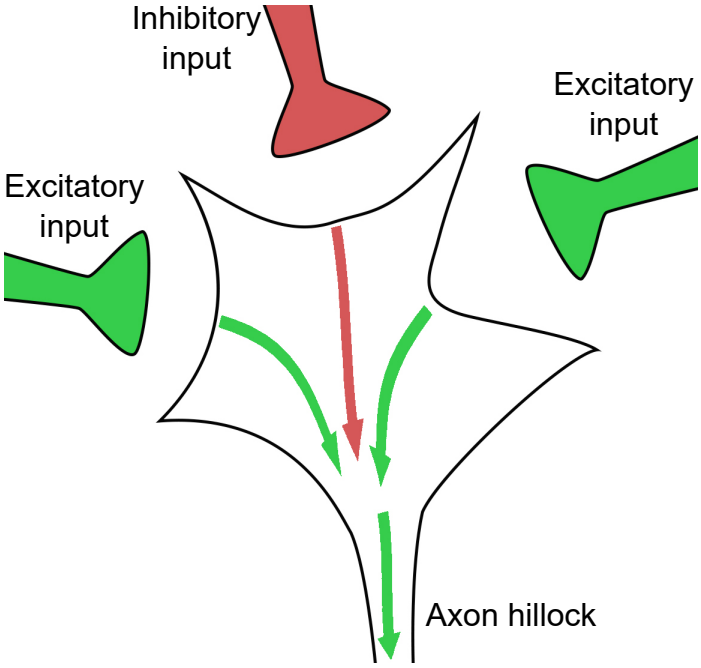


Figure 3.10 The axon hillock acts as an integration center that sums all the excitatory and inhibitory inputs, and decides on a binary output called the **action potential**.

called the **axon hillock**. With respect to signal transduction, the axon hillock works as the integrative center of the neuron. It is responsible for deciding whether to pass a signal onto the next cell. The cell membrane of the hillock performs a complex set of "cellular arithmetic" that weighs all of the incoming signals: excitatory, inhibitory, and modulatory signals. After all of the calculations have been performed, the membrane either sends a signal or not.

At the end of each branch of the axon is a small swelling, which is called the **axon terminal** or **terminal bouton**. The terminal is the part of the neuron that is specialized for the production and release of the neurotransmitters that are used for communication between neurons. One subarea of specific interest in the axon terminal is a small patch of membrane called the **active zone**. Embedded in the cell membrane at the active zone is a variety of proteins that are important for the process of neurotransmitter release.

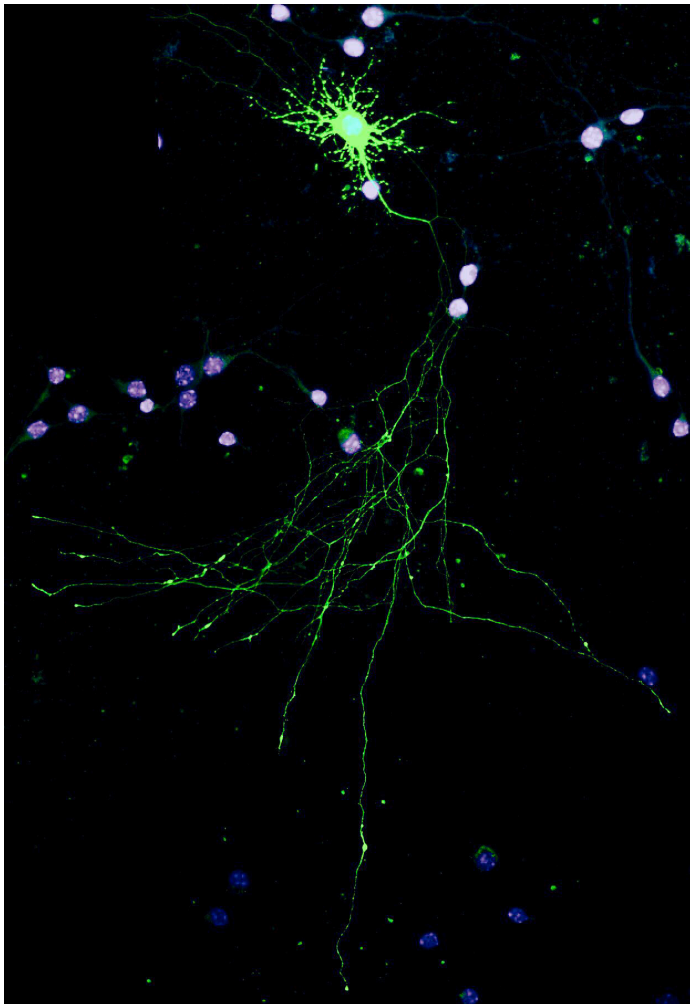


Figure 3.11 Neurons (green) extend a single axon from the soma, which can branch extensively.

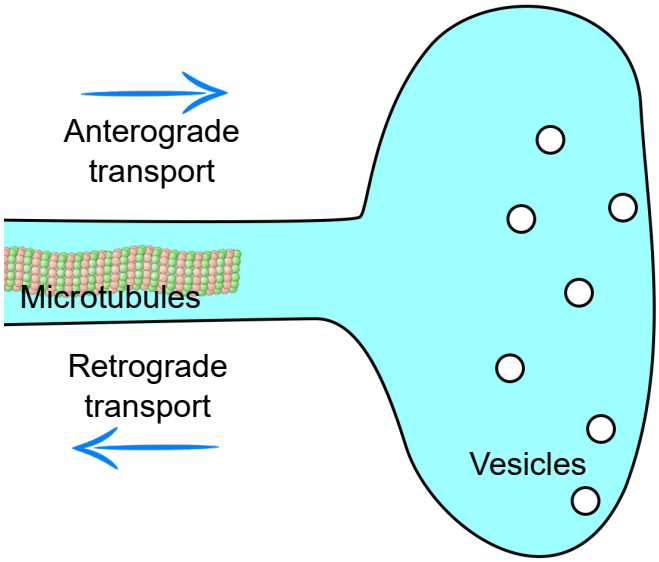


Figure 3.12 Diagram of an axon terminal, showing microtubules and vesicles.

Inside the axon

Most of the proteins synthesized in neurons are created in the cell body very close to the nucleus, where the mRNA that is exported from the nucleus is able to easily interact with the rough endoplasmic reticulum and the ribosomes. But, some of these proteins are needed far away from the cell soma, at the axon terminal, for example. Neurons need some system of transport system that can move newly created proteins to where they need to go. Inside the length of the axon runs an organelle called **microtubules**, which function like a molecular railway for proteins. The cells use motor-like proteins that can carry other proteins along the microtubules. When substances are transported away from the cell body, it is called anterograde transport, while the process of moving proteins towards the cell body is retrograde transport.

Another important organelle that is found along the length of the axon are **neurofilaments**. These organelles are made up of several different proteins that serve as a cellular “scaffolding” that helps keep the structure of the axon intact. Mature neurons can be very dense with neurofilaments, which increases the diameter of the axon.

Within the terminal are a number of **vesicles**, small spherical “packages” made of cell membrane that are coated in special proteins. Within these vesicles are the molecules that the neurons use for chemical communication. When the action potential travels down the axon and reaches the axon terminal, the cell membrane changes in electric charge, and this causes vesicles to fuse with the inner membrane of the

neuron. During fusion, the vesicular membrane merges with the cell membrane at the axon terminal, causing the contents of the vesicle to be released outside the cell.

Outside the axon

Some neurons have a special modification surrounding their axon called a **myelin sheath**. Myelin is comprised of several tightly-wrapped layers of cell membrane that encompasses a short section of the axon. Myelin might be wrapped as many as 250 or 300 times around a single section of axon. Myelin does not extend fully enclose the entire length of the axon from soma to terminal, but rather surrounds short sections at a time. The spaces of exposed axon between each section of myelin are called **nodes of Ranvier**. On average, these nodes are about 1 micron long.

Myelin serves a few functional purposes. Myelin increases the speed by which an electrical signal is transmitted. Some of the most heavily myelinated axons are able to send signals up to 120 meters per second (almost 270 miles per hour, faster than a Formula 1 racer.) Myelin also increases the effective thickness of the cell membrane along the axon. In doing so, myelin acts as an insulator that causes signals to more reliably be passed down the axon.

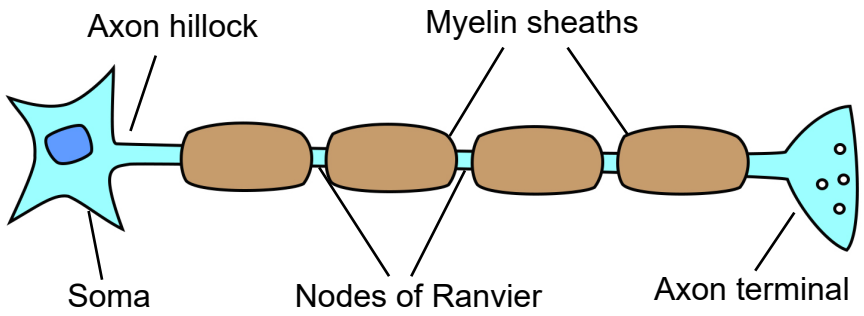


Figure 3.13 Myelin sheaths are made up of layers of cell membranes that surround small sections of axons. The nodes of Ranvier are the exposed sections of axons between the myelin.

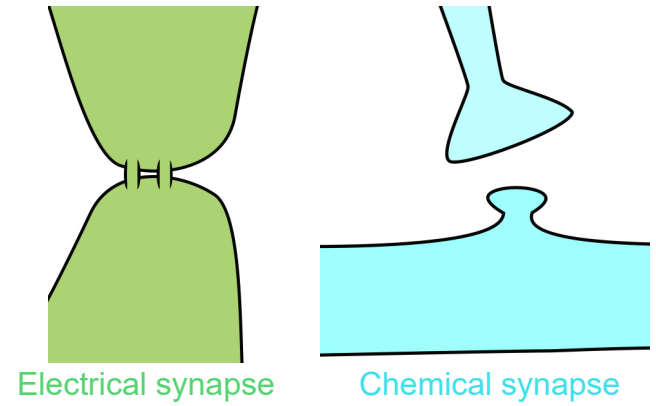


Figure 3.14 Electrical synapses physically share cytoplasm (left) while chemical synapses use neurotransmitters to communicate (right).

Synapse

The synapse is the physical distance that separates two neurons. In agreement with Cajal's neuron doctrine, the nervous system is not made up of a single cell with a giant, shared cytoplasm, but rather a series of neurons in close proximity, separated by a gap of extracellular space.

This distance between two cells can vary depending on the nature of the synapse. An electrical synapse may be less than 5 nanometers apart. Cells connected by electrical synapses share cytoplasm but have two separate cell membranes. On the other hand, a chemical synapse is a larger distance, about 15 - 40 nm across. Adjacent neurons connected by chemical synapses do not share cytoplasm.

Clinical connection: Multiple sclerosis

Multiple sclerosis (MS) is a disease that results from destruction of myelin in the CNS. When myelin is damaged, signals do not reliably propagate from the brain to the body, from the body to the brain, or between areas of the brain. MS damages myelin at both descending and ascending neurons, so a person with MS might experience muscle weakness, poor balance, and muscle spasms (efferent motor neurons) as well as numbness and pain (afferent motor neurons). It can also affect neuronal signaling in the brain, causing symptoms such as cognitive impairment, vision loss, and changes in affect leading to depression.

Multiple sclerosis is a common neurological disorder that affects some 2.5 million people worldwide. MS typically presents in adulthood, usually between 20-50 years old. Unfortunately, there is currently no known cure for MS. Therapy is focused on either slowing the progression of the disease, helping patients recover after an attack, or decreasing the severity of the symptoms. While MS has the potential to be debilitating and painful, the disease itself is not necessarily lethal, and MS only decreases life expectancy by 5-10 years on average. People with MS are just as likely to die from natural causes as much as a neurotypical person.

One of the leading theories about the cause of MS is a faulty immune system. A normally functioning immune system identifies and destroys foreign pathogens. Sometimes, this system makes a mistake, and the immune response recognizes normal parts of the body as being a "foreign object", causing the body to destroy itself. This is called an *autoimmune disorder*. In the case of MS, it is believed that the immune system identifies and targets myelin for destruction, leading to demyelination. Although immunosuppressants may slow the progression of MS, these drugs increase the risk of a person developing an illness that a healthy immune system would be able to prevent.

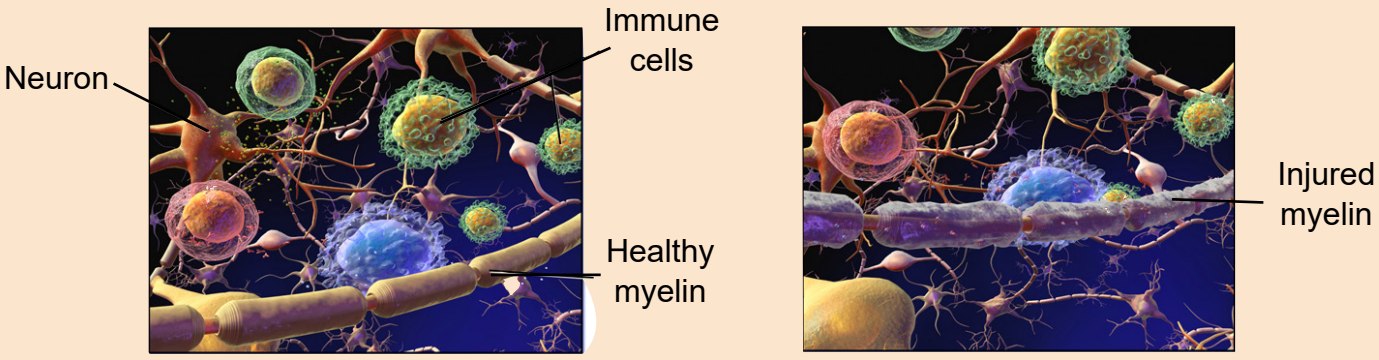


Figure 3.15 Multiple sclerosis is a disease resulting from the destruction of myelin, which hinders the ability for neurons to communicate properly.

Classifications of neurons

Neurons have a wide variety of shapes and can be divided roughly into three different classes depending on their morphology. In order of increasing morphological complexity, they are:

1. Unipolar cell

Unipolar cells have a single cellular extension coming off the soma that acts as both the receiving and the sending end. Unipolar cells are very common among invertebrates; humans do not have unipolar neurons.

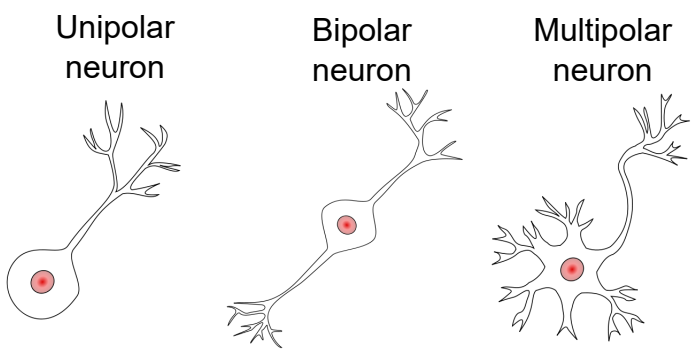
2. Bipolar cell

A bipolar cell, as can be implied from the name, has a single dendrite and a single axon. They are not very common, but can be found in human sensory systems, such as in the eye or in the signaling pathway connecting the ear to the brain.

3. Multipolar cell

Multipolar neurons have several dendrites and one single axon. These cells are the most common among all the neurons in the human nervous system. Most drawings of neurons you will see in this book are multipolar cells.

Figure 3.16 Neurons can be classified based on their morphology.



In addition to classifying neurons based on their morphology, neurons can also be divided based on their functions.

1. Sensory

Sensory neurons are the afferent neurons that are responsible for obtaining information about either the outside world or the internal environment and passing that information towards the CNS. Sensory neurons can detect a variety of stimuli, ranging from photons of light (visual system) or chemicals floating in the air (olfactory system), to carbon dioxide levels in the blood (chemoreceptors) or stretching of the muscles (spindle receptors). Since they are highly specialized for detecting different types of stimuli, sensory neurons have a large variety of shapes and structures.

2. Motor

Motor neurons, or motoneurons, carry signals from the CNS to the body. There are two main types of motor neurons. Somatic motor neurons control skeletal muscle movement, such as flexing or extending the muscles of the arm. They release neurotransmitter directly onto muscles. Autonomic motor neurons, on the other hand, release neurotransmitter onto a clump of neurons outside of the CNS called the autonomic ganglion, which then signal to the smooth muscle, cardiac muscles, or glands.

3. Interneuron

Interneurons exist as a relay between the sensory or motor neurons and the CNS, or between each other. They represent a very broad class of neurons, and make up an important part of reflex circuits, such as the knee-jerk response.

The above classifications are only a rough guideline for separating neurons. As with most biologists, neuroscientists enjoy classifying cells based on their properties. It is almost always preferable to use the most specific name possible in identifying neurons. Many cells have been identified based on the way they look (chandelier cells), some named on the neurotransmitter released (cholinergic interneurons), some named by where they are found (cerebellar granule cell), and yet others named based on the person who discovered them (Purkinje cells).

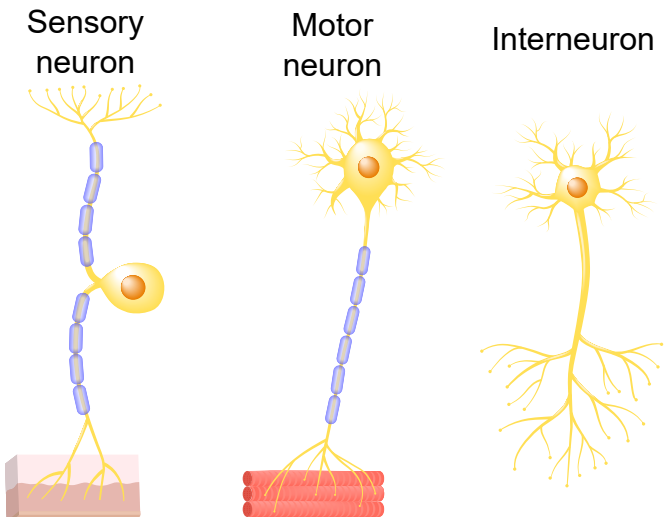


Figure 3.18 Neurons can also be classified based on their function.

Visualizing the synapse

The original microscopes that van Leeuwenhoek used to visualize cells and microorganisms relied on the transmission of light to see objects up close. For us to be able to see objects, light needs to bounce off the intended target. Light particles travel in a wave, and if the object we intend to visualize is smaller than the wave, then the wave would pass right over the object. The shortest wavelength of visible light is around 380 nanometers, so light microscopy is limited to visualizing objects in the micrometer or larger range - a 2,000x magnification with the best possible lenses.

In the early 1950s, a group of scientists developed a technique called **electron microscopy (EM)**. In EM, a beam of electrons is aimed at a sample in a vacuum, and the reflection of those electrons can be collected and detected with a computer. Since electrons have a much shorter wavelength than visible light, you are able to resolve objects as small as 50 picometers - more than a 10,000,000x magnification!

Using EM, you are able to get ultrastructural resolution of the synapse. In two adjacent neurons, EM allows you to distinguish the boundaries of both neurons, and to clearly see the synapse in between. You can also see the vesicles contained in the axon terminal, sometimes in the middle of fusion.

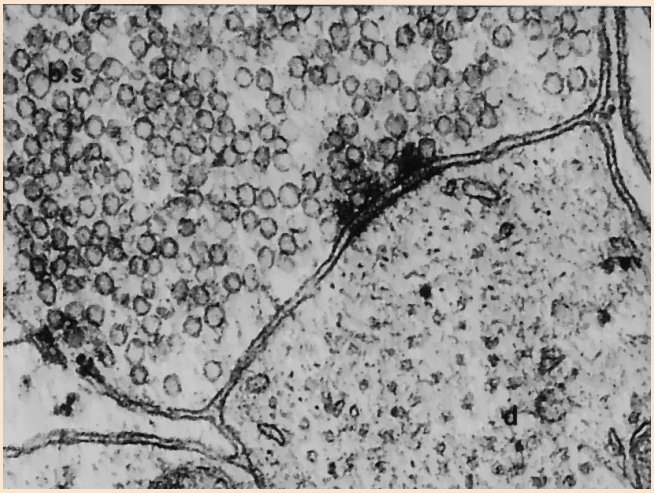


Figure 3.17 Electron microscope image showing a synapse at ultrastructural resolution (less than 100 nm).

3.3 Cellular functions of glia

Although most of neuroscience is concerned with understanding the functions of neurons, there are other cells in the nervous system that are just as interesting. These cells are grouped together under the umbrella classification of **glia**. Historically, when these non-neuronal cells were visualized under the microscope, the histologists and anatomists had no idea about their function. They were seen all around the neurons, so the assumption was that these cells were structural elements, a sort of living glue that held the nervous together. Today, we know that these glia serve all variety of functions; unfortunately the misnomer "glia," derived from the Latin word for glue, is still used to describe these non-neuronal components of the nervous system. We estimate that the brain has roughly an equal number of glial cells and neurons—86 billion of each.

There are many different classes of glia, but we will focus on five types.

1. Astrocytes

Astrocytes are named for their characteristic star-shaped morphology. Astrocytes have a dense expression of the protein **glial fibrillary acidic protein (GFAP)**, and this protein is often used as a marker for differentiating astrocytes from other cell populations.

One of the main functions of astrocytes in the brain is to help maintain the blood-brain barrier. At the end of the extensions of the astrocyte are protrusions called "endfeet." These endfeet are often wrapped around the endothelial cells that surround the blood vessels. The endfeet release important biological compounds that allow the endothelial cells to remain healthy as they

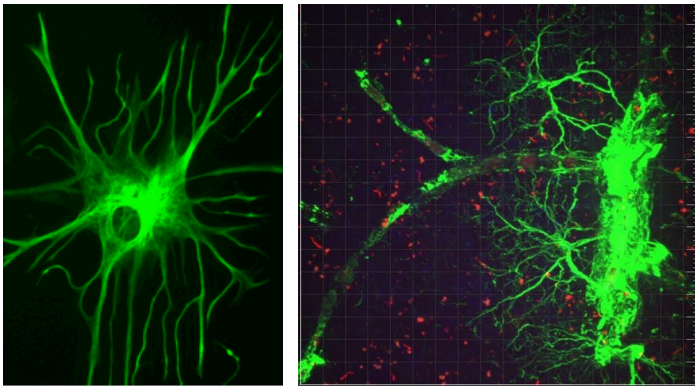


Figure 3.19 Astrocyte labeled by staining the protein GFAP (left). Astrocytic endfeet are closely associated with blood vessels as part of the blood-brain barrier (right).

function in maintaining the blood-brain barrier.

Astrocytes are also very closely associated with synapses. Astrocytes have a very dense expression of proteins on their cell surface that can transport molecules of the neurotransmitter glutamate, for example, inside the astrocyte. By acting as a glutamate "sponge," astrocytes are able to decrease the strength of a glutamate signal. Through a similar uptake mechanism, astrocytes can also affect the extracellular concentration of ions such as potassium, which then has an influence on cellular excitability. Because of these interactions between astrocytes and neurotransmission at the synapse, we use the phrase **tripartite synapse** to refer to the three components of a synapse: The presynaptic neuron, the postsynaptic neuron, and the astrocyte.

Astrocytes also synthesize and produce a variety of **trophic factors**, which are helper molecular signals that serve several different functions. For one, trophic factors signal to neurons that the neuron should continue to live,

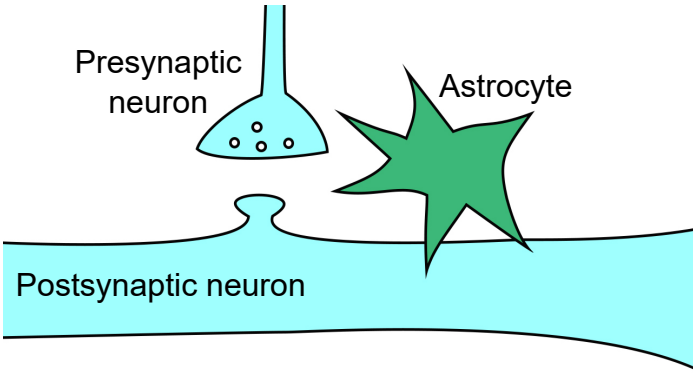


Figure 3.20 Astrocytes can modify communication between two neurons and are therefore part of the tripartite synapse.

or that specific synapses should be maintained. They help guide the neurons as they reach out, forming synapses where appropriate.

2. Oligodendrocytes

Oligodendrocytes only exist in the CNS. Their name is derived from their morphology. Oligo- refers to “a small number” (think oligarchy, a government ruled by a few people), and dendro- refers to “tree” (like a rhododendron). Each oligodendrocyte has a few branches that reach away from the cell body.

The main function of the oligodendrocytes is to add a layer of myelin around the axons of nearby CNS neurons. A single oligodendrocyte is able to myelinate up to 50 segments of axons. As cells that produce myelin, they are responsible for increasing the conduction speed of nearby neurons as they send signals. When the oligodendrocytes begin myelinating, they are able to produce almost 3 times their weight in membrane per day. By maturity, the oligodendrocyte supports cell membrane that is 100 times the weight of the cell. Due to the need to support such large amounts of myelin, it’s estimated that they have the highest metabolic rate of any type of cell in the brain.

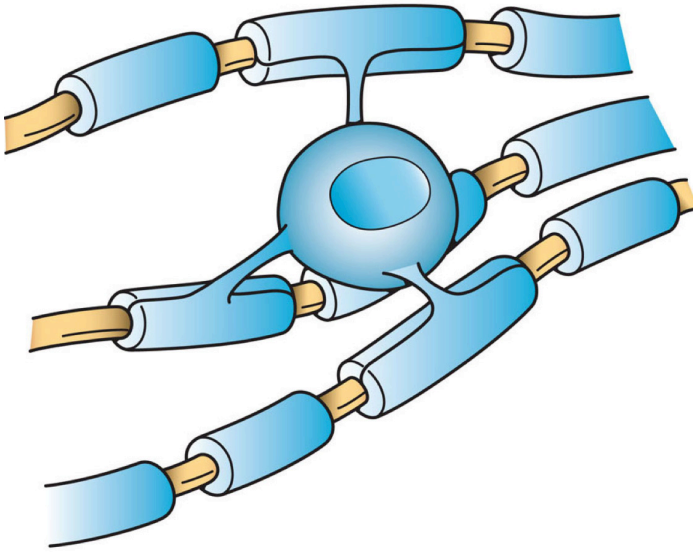


Figure 3.21 Oligodendrocytes are capable of myelinating multiple axon segments at once.

3. Schwann cells

Schwann cells are named after the German biologist who first described these glial cells. He found that there are cells that are wrapped around the axons of nerve cells that project towards muscles - myelin. These Schwann cells can only be found in the PNS.

The main action of Schwann cells is to provide a section of myelin sheath for PNS neurons, and in this way, they function similarly to the oligodendrocytes. Schwann cells produce only a single section of myelin, compared to oligodendrocytes, which myelinate multiple sections.

Schwann cells also function in the regeneration of injured axons. When nerves in the PNS are damaged after trauma, Schwann cells rapidly mobilize to the site of injury. Here, the remaining loose myelin hinders the regeneration process, and the Schwann cells destroy the extra cellular membranes. Schwann cells also produce signaling molecules that guide the injured axons to the correct targets, which helps the axon regrow.

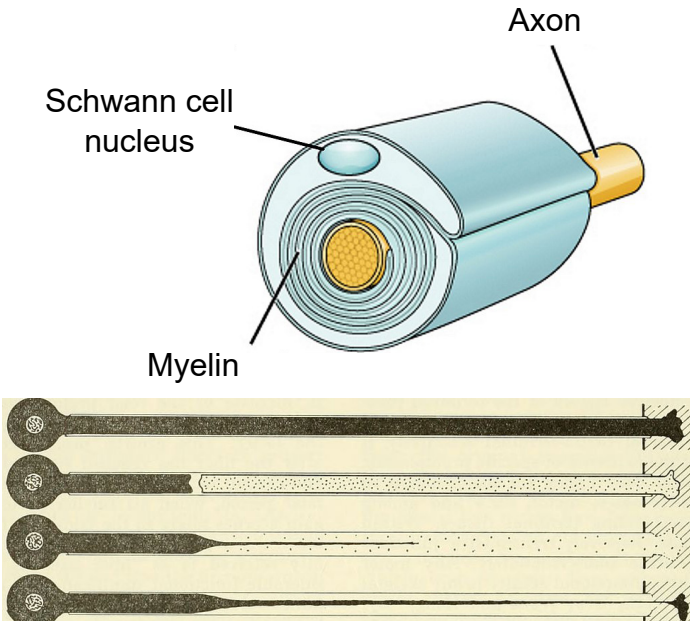


Figure 3.22 Schwann cells myelinate a single segment of an axon in the PNS (top). They also contribute to regeneration of injured nerves (bottom).

4. Microglia

Microglia are a bit different from the other glial cell populations. For one, microglia are more immune cells rather than neural. They act as cellular scavengers that travel throughout the brain and spinal cord. It is estimated that microglia make up 10-15% of all cells in the brain.

As immune cells, microglia identify and destroy clumps of proteins, dead / dying cells, or foreign pathogens that enter into the brain. After an injury to the CNS, like a traumatic blow to the head, microglia rapidly react to the area of the insult. The marker Iba1 is often used to identify when microglia are reacting to an injury.

5. Ependymal cells

Along the inside of the ventricles are a lining of glia called the ependymal cells. These ependymal cells are columnar with small finger-like extensions called cilia that extend into the ventricles and into the central canal that runs down the inside of the spinal cord.

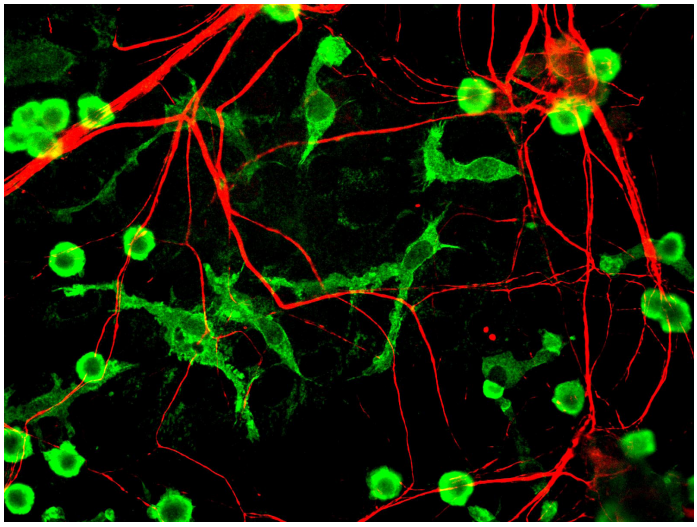


Figure 3.23 Microglia (green) are much smaller than neurons (red). They are neural immune cells that exhibit morphological changes when performing cellular cleaning.

Ependymal cells produce CSF. In total, the body can make about half a liter of CSF each day (a little more than two cups.) The ependymal cells are part of a structure called the **choroid plexus**, the network of blood vessels and cells that form a boundary between the blood and the CSF.

Figure 3.24 Ependymal cells line the inside of the ventricles and produce cerebrospinal fluid.

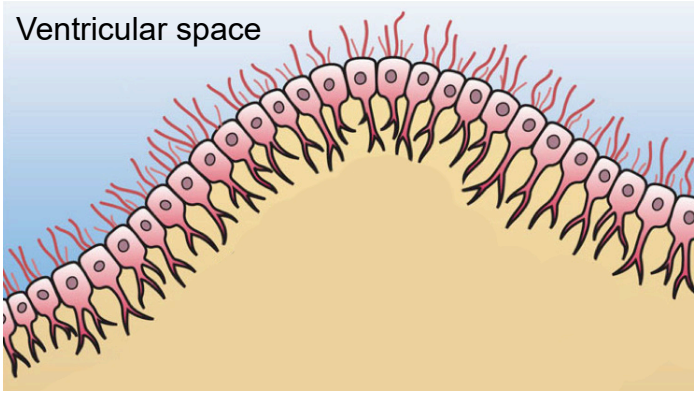
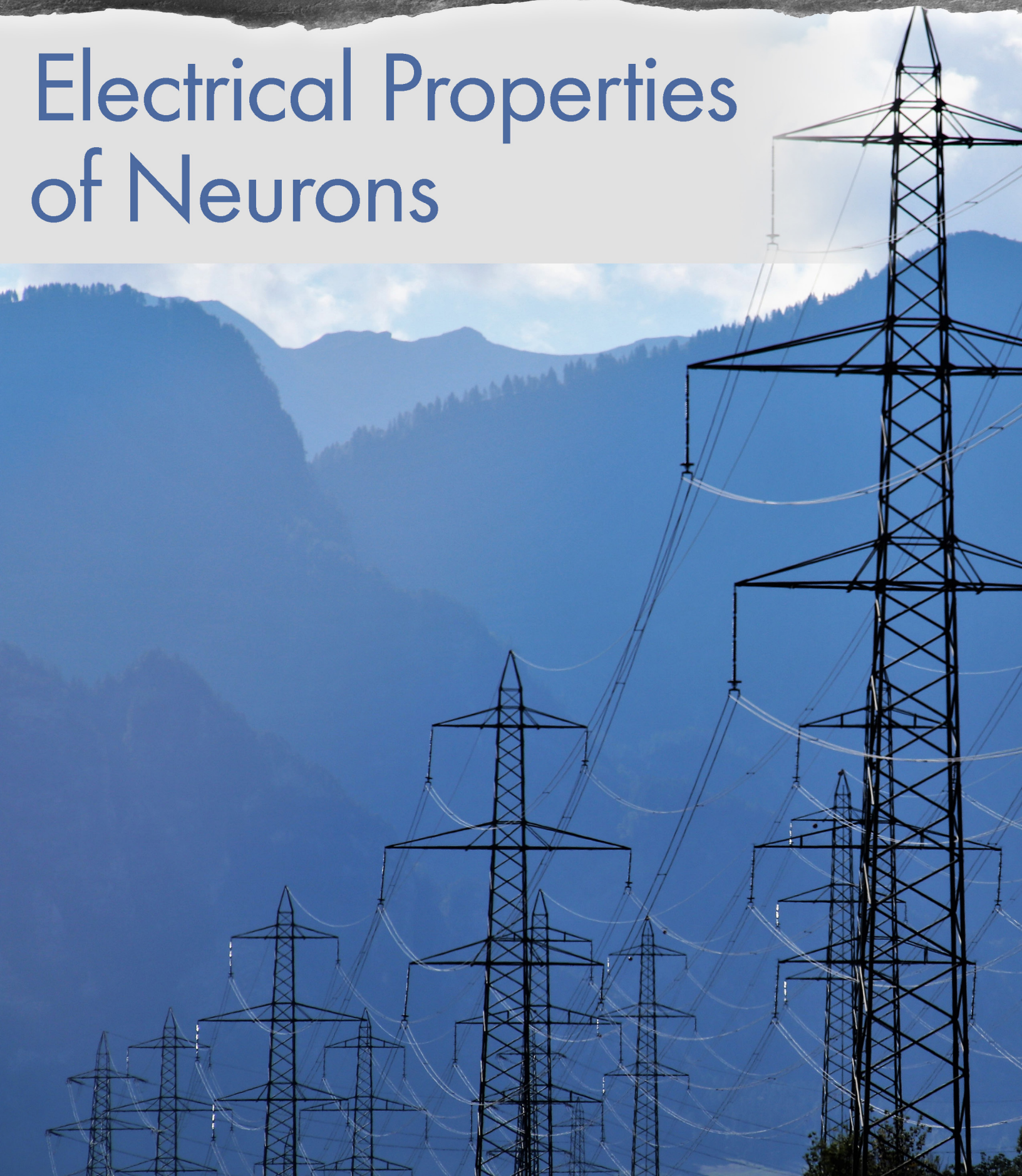


Image credits
Cover: Dana H. Simmons, Ph.D. dana-simmons.com
3.1 https://upload.wikimedia.org/wikipedia/commons/6/6d/GolgiStainedPyramidalCell.jpg
https://upload.wikimedia.org/wikipedia/commons/1/1a/Human_hippocampus_2.5x.jpg
3.2 https://upload.wikimedia.org/wikipedia/commons/5/57/Camillo_Golgi_nobel.jpg
https://upload.wikimedia.org/wikipedia/commons/3/30/Cajal-Restored.jpg
3.3 https://upload.wikimedia.org/wikipedia/commons/5/5f/Cell_membrane_detailed_diagram_simplified_ca.svg modified by Austin Lim
3.4 https://svgsilh.com/9e9e9e/image/2040692.html modified by Austin Lim
3.5 https://pixabay.com/photos/axe-old-lumberjack-blade-1008981/
3.6 https://upload.wikimedia.org/wikipedia/commons/0/0a/Anatomy_of_a_Neuron_with_Synapse.png modified by Austin Lim
3.7 https://upload.wikimedia.org/wikipedia/commons/7/74/Dendritic_arborization_patterns_2.png
https://upload.wikimedia.org/wikipedia/commons/9/90/Submerged_Albizia_Saman_in_the_Mekong_at_sunset_%28close-up_view%29.jpg
3.8 https://upload.wikimedia.org/wikipedia/commons/6/6e/Geometric_characteristics_of_dendritic_spines.png modified by Austin Lim
3.9 https://upload.wikimedia.org/wikipedia/commons/f/fb/MRNA-interaction.png modified by Austin Lim
3.11 Image courtesy of Connor Maltby
3.12 https://upload.wikimedia.org/wikipedia/commons/8/86/201704_microtubule.svg modified by Austin Lim
3.15 https://upload.wikimedia.org/wikipedia/commons/2/2a/Multiple_Sclerosis.png modified by Austin Lim
3.16 https://upload.wikimedia.org/wikipedia/commons/9/92/Neurons_uni_bi_multi_pseudouni.svg modified by Austin Lim
3.17 https://upload.wikimedia.org/wikipedia/commons/6/6c/Synapse_neuro-neuronale.png
3.19 https://upload.wikimedia.org/wikipedia/commons/5/56/Human_astrocyte.png
https://upload.wikimedia.org/wikipedia/commons/3/34/Astrocytes_Wrapping_around_Blood_Vessel.png
3.21 https://upload.wikimedia.org/wikipedia/commons/2/2e/Oligodendrocyte_illustration.png
3.22 https://upload.wikimedia.org/wikipedia/commons/5/52/CeluladeSchwann.jpg
https://upload.wikimedia.org/wikipedia/commons/0/03/Analysis_of_development_%281955%29_%2817549733973%29.jpg modified by Austin Lim
3.23 https://upload.wikimedia.org/wikipedia/commons/0/0b/Microglia_and_neurons.jpg
3.24 https://upload.wikimedia.org/wikipedia/commons/c/ca/Glial_Cell_Types.png modifed by Austin Lim

The Open Neuroscience Initiative is funded by a grant from the Vincentian Endowment Fund of DePaul University.

Chapter 4:

Electrical Properties of Neurons



Austin Lim, PhD (DePaul University)
Editor: Dana Simmons, PhD (University of Chicago)

One of the interesting properties of neurons is that their cell membranes are **electroactive**, meaning that they are sensitive to electrical charge. Neurons are capable of changing the electrical potential of their membranes, which makes them a very dynamic population of cells. These changes, which can happen on the scale of milliseconds, provide a rapid method for one part of a neuron to send signals over long distances – hence, the analogy of electrical wires running overhead. In this chapter, we will go into depth describing the molecular and cellular components that contribute to the electrical properties of a neuron, both at rest and during an action potential.

Chapter 4 outline

4.1 Ion channels

The cell membrane that separates the inside of the cell from the outside is a very effective boundary. It is described as being s**electively permeable**, which means that some molecules are able to travel across the membrane easily, other molecules have an intermediate ability to cross, and other molecules are completely incapable of passing. Generally, gases and molecules of water are able to pass through the cell membrane easily. Large molecules like glucose, and charged molecules like ions or amino acids, are unable to pass across the membrane.

Most cells of the body, including neurons, have specialized **transmembrane proteins** embedded in the cell membrane. These transmembrane proteins are huge protein complexes that span the entirety of the membrane, with an outer side and an inner side. In the middle of the protein is a **pore**, which is essentially a "tunnel" that allows molecules and ions to pass across the cell membrane. These proteins are called **ion channels**. These channels are passive since they do not use any cellular energy to move ions. Rather, they simply provide easy passage for ions. It may be useful to think of an ion channel as a "cellular door."

One important feature of ion channels is their ability to distinguish ions based on their chemical properties. For example, some

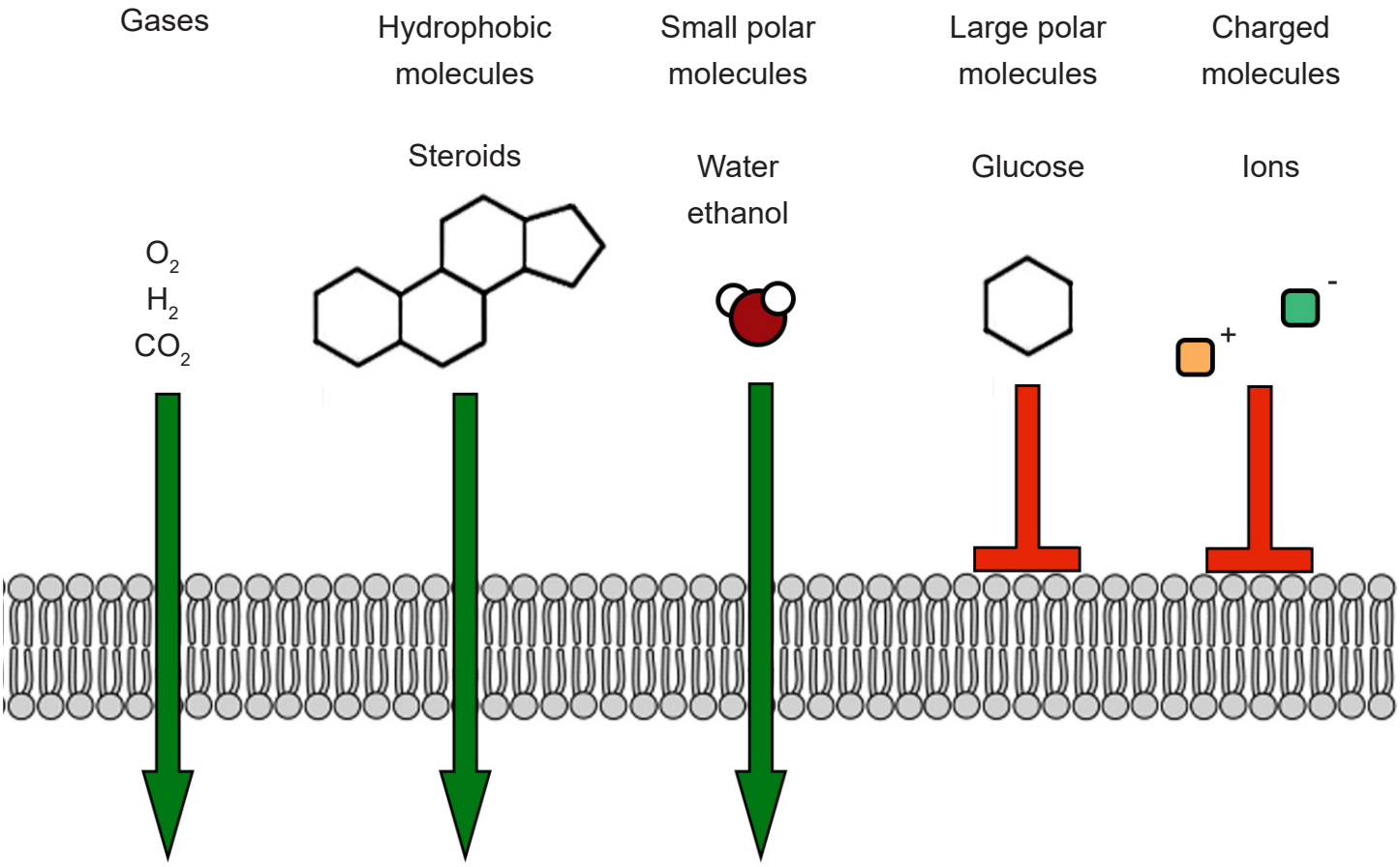


Figure 4.1 Selective permeability of molecules across the cell membrane.

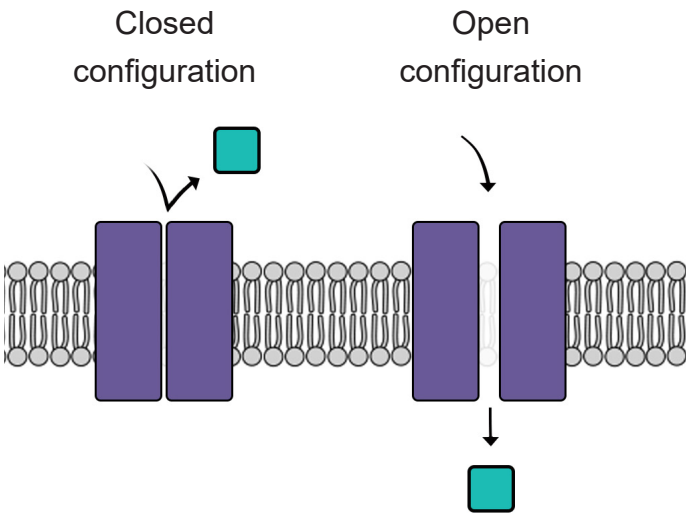


Figure 4.2 Ion channels (purple) are transmembrane proteins that can be open, allowing charged ions to cross the cell membrane through a molecular "tunnel" called the pore.

channels are selective for Na^+, while preventing the passage of all other ions. Each ion channel has special molecular characteristics that allow certain types of ions to pass through the pore while excluding other ions. Some features that allow for distinction between ions include:

1. Pore size. The molecular shape of an ion channel can exclude larger diameter ions if it is small enough, restricting the ability for the large ion to enter the pore – in the same way that a ping pong ball cannot pass through a narrow diameter of a straw.

2. Electrical charge. Ion channels can also prevent certain ions from passing through based on the electrical charge. The inside of the pore is lined with charged amino acids, and their charge allows only certain types of ions to pass.

For example, sodium channels have several negatively-charged glutamate residues inside the pore. Because of this, negatively-charged ions like chloride will be repelled by the negative charges, but the positively-charged sodium ions will be attracted to the inside.

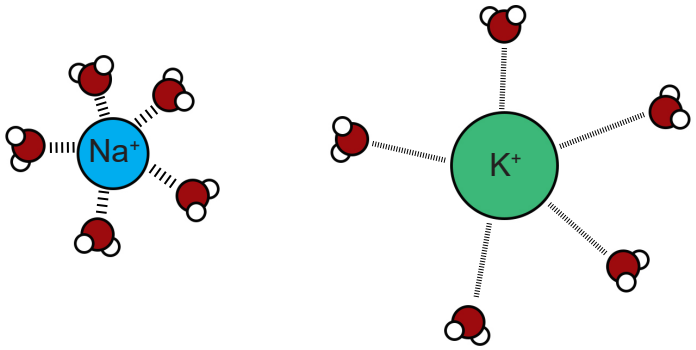


Figure 4.3 Ions dissolved in water are surrounded by a shell of water molecules. The potassium ion is larger, and the water molecules are farther apart, making the attraction weaker than in the sodium ion.

3. Hydration shell. Selectivity based on charge and small size is a simple task, but how can it be possible for an ion channel to allow larger ions to pass while excluding smaller ions? This is possible because it is energetically favorable to allow an ion that has been stabilized sufficiently by the properties of the amino acid residues lining the pore.

This process is related to the observation that a shell of water molecules surrounds each charged ion in solution. To pass through the pore, water molecules must be separated from the ion. The potassium ion has diffuse hydration shell, and so it is easier to separate water from a potassium ion compared to a sodium ion. The passage of potassium becomes more energetically favorable compared to passing a sodium ion, causing these pores to be selective.

Like doors, different types of ion channels open and close under different conditions. We can categorize ion channels into four major classes based on their opening and closing conditions.

1. **Leak channels** are persistently open. You can think of these leak channels as revolving doors that are never locked. Neurons usually have several leak channels, such as a leak channel for potassium (K^+) and a leak channel for chloride (Cl^-) ions. These two ions are at equilibrium, flowing back and forth across the cell membrane, and they help contribute to the properties of the neuron at rest.

2. **Voltage-gated ion channels** are sensitive to the electrical potential of the surrounding membrane. Their ability to open and close depends on the electrical charge of the membrane. Many of them, such as voltage-gated sodium (Na^+) channels, remain closed at negative potentials (resting condition) and only open at positive potentials (action potential).

3. **Ligand-gated ion channels** open in response to binding certain molecules (ligands), such as neurotransmitters. These types of ion channels are also called **ionotropic receptors**, and these will be described in much more depth in chapter 5.

4. The fourth class of ion channels is a catch-all category that includes a wide variety of channels that are used by the sensory systems. They open and close in response to unique stimuli depending on what they are able to sense. For example, some open and close depending when they are moved physically, such as a distortion or stretch. We have these in our ears (hair cells, chapter 8.1), in our skin (Pacinian corpuscles, chapter 8.3) and in our muscles (Golgi tendon organ, chapter 10). Photoreceptors in our eyes have ion channels that close in response to being hit by photons of light, and this activity is necessary for us to be able to see in both brightly and dimly lit environments.

4.2 The electrochemical gradient

Without ion channels, ions do not cross the cell membrane and therefore stay either outside or inside the cell. But when an ion channel is present and open, ions can move across the cell membrane, depending on the specific conditions inside and outside the neuron. There are two different forces that act on ions once an ion channel is open: the **electrical gradient** and the **chemical gradient**.

1. Electrical gradient

The electrical gradient refers to the electrical forces acting on charged molecules, "pulling" opposite charges together while also "pushing" like charges away from one another - just like the polarity of magnets.

At rest, the interior of the cell has a negative charge, about -70 mV. Because of these negative charges inside the cell, the electrical gradient causes positively charged ions to be attracted to the inside of the cell. At the same time, negatively charged ions will be repelled out of the cell.

The voltage of the neuron (V_m) can deviate significantly from -70 mV, reaching potentials as high as +40 mV. When the V_m changes, the force exerted by the electrical gradient changes as well. If the cell now has a positive potential instead of a negative potential, movement of charged ions will change as well. In this case, the electrical gradient now attracts negatively charged ions into the cell while pushing positively charged ions out of the cell.

2. Chemical gradient

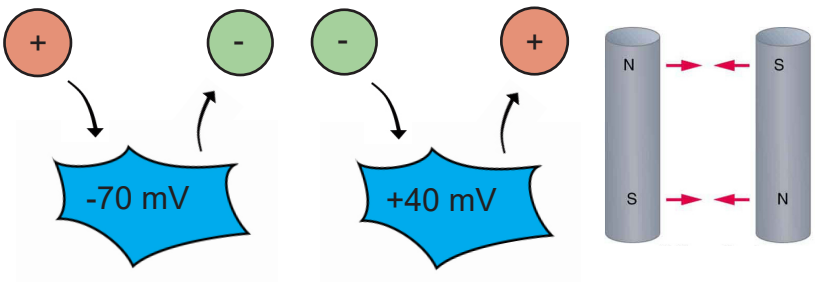
The chemical gradient refers to the natural process by which a high concentration of a substance, given enough time, will eventually diffuse to a lower concentration and settle evenly over the space.

Figure 4.5 The chemical gradient describes the same forces acting on ions as diffusion, where ions move from areas of high concentration to low concentration.

You can think of the chemical gradient as similar to the way people behave when they exit a full elevator. People don't want to be so close to each other inside a cramped space. Once the doors open, people scatter quickly, going to where they can have more personal space. They move from an area of high concentration to low concentration. Molecules behave the same way.

Unlike the electrical gradient, the chemical gradient is unaffected by biological movement of ions. In other words, the distribution of ions across the membrane does not change

Figure 4.4 The electrical gradient describes the forces acting on ions as they follow the rules of magnetism in physics - "opposites attract, and similar charges repel."



significantly. The number of ions that flow across the cell membrane is so minute that the relative distribution of ions across the cell membrane essentially does not change.

These two forces are collectively called the electrochemical gradient. In order to accurately predict the forces acting on a given ion, you need to know 1) the charge of the ion and 2) the relative concentrations of ions across the membrane.

Take a look at the following two examples :

1. Sodium (Na^+) movement across the membrane

Consider the positively-charged ion, sodium (Na^+). Because it has a positive charge, it is electrically attracted to the inside of the cell, which has a negative charge at rest. Therefore, when a sodium-permeable ion channel opens and Na^+ is now able to pass across the cell membrane, the electrical gradient causes the sodium to move into the cell.

We also know that the concentration of sodium outside the cell (written as $[Na^+]_o$, where the brackets indicate concentration and the "o" refers to "outside") is high compared to the concentration of sodium inside the cell (written as $[Na^+]_i$) by roughly an order of magnitude. Therefore, when that sodium-permeable ion channel opens, Na^+ will move from the area of high concentration to low concentration, causing Na^+ to move into the cell.

Given that both the electrical and chemical gradients are in favor of moving Na^+ into the cell, opening a sodium channel causes a **net flow** of Na^+ ions into the cell.

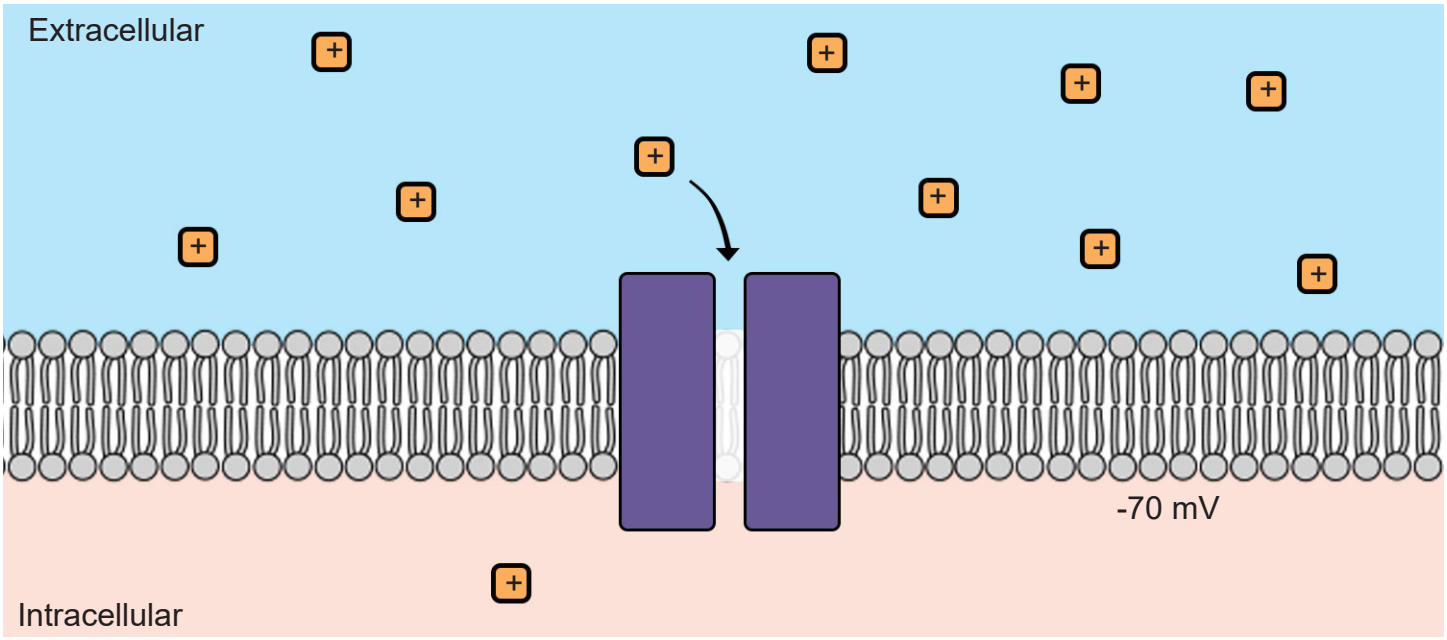


Figure 4.6 When a sodium channel opens (purple), sodium ions (orange) are acted on by two forces. They have a positive charge, which is attracted to the negatively-charged inside of the cell (electrical gradient). They are at a higher concentration outside the cell, which causes them to move to the inside of the cell with the lower concentration (chemical gradient).

2. Potassium (K^+) movement across the membrane

Just like with sodium, we can use our knowledge of potassium's charge and the distribution of K^+ ions across the membrane in order to predict the forces acting on K^+ ions when a potassium channel is open.

We know that the inside of the cell is negative and K^+ has a positive charge, so the electrical gradient causes K^+ to be attracted into the cell. We also know that the inside of the cell has a relatively high concentration of K^+ ions compared to the outside of the cell; therefore the chemical gradient pushes K^+ out of the cell.

Since the ion concentrations across the membrane do not change significantly, but the membrane voltage can, you could imagine there exists some V_m where the flow of ions moving in are exactly equal to the flow of ions moving out. In this case, the two forces acting on the ion oppose one another perfectly, resulting in a **dynamic equilibrium** (This process is called "dynamic" because there is a constant movement of ions going in and going out, and yet it is at an "equilibrium" because there is no net movement of charge: the charge moving in is balanced by the charge moving out.) This exact value of the Vm is called the **equilibrium potential** for the ion, which is also abbreviated $\mathbf{E_x}$ where x is the ion of interest. The equilibrium potential differs for each ion, and is described in Table 4.1.

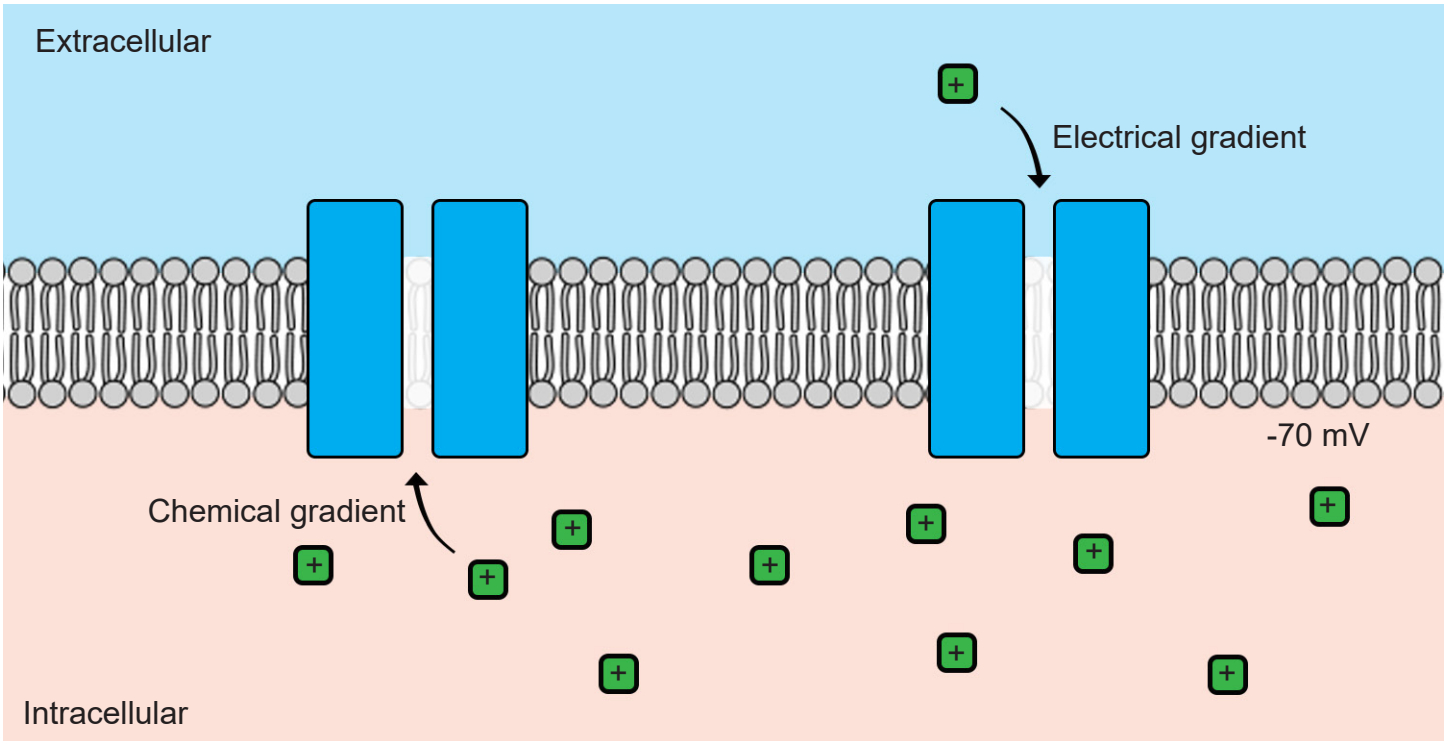


Figure 4.7 When a potassium channel opens (blue), potassium ions (green) are acted on by two opposing forces. They have a positive charge, which is attracted to the negatively-charged inside of the cell (electrical gradient). At the same time, they are at a higher concentration inside the cell, which causes them to want to move outside of the cell (chemical gradient). This process is at equilibium.

Ion	Abbreviation	Concentration outside $[X]_o$	Concentration inside $[X]_i$
Sodium	Na^+	140 mM	15 mM
Potassium	K^+	5 mM	150 mM
Chloride	Cl^-	120 mM	10 mM
Calcium	Ca^{2+}	1 mM	100 nM = 0.1 μM
Magnesium	Mg^{2+}	2 mM	0.5 mM
Miscellaneous anions	A^-	20 mM	100 mM

Table 4.8 Charged ions are unequally distributed across the cell membrane. Typical concentrations of ions in a mammalian neuron.

4.3 Calculating E_x and the Nernst equation

Even though the **equilibrium potential** (E_x) for each ion can be determined experimentally, it is also possible to calculate E_x using a mathematical equation called the **Nernst equation**. By understanding the Nernst equation, you can predict the direction that ions will move when an ion channel opens given various conditions. The Nernst equation is as follows:

$$E_x = \frac{RT}{zF} \ln\left(\frac{[x]_o}{[x]_i}\right)$$

Equation 4.9 The Nernst equation calculates the reversal potential for a given ion *x*.

E_x is the equilibrium potential. As a potential, it has units in volts, usually measured in millivolts. When a cell's membrane potential is exactly at this E_x, the ion is at equilibrium: ion movement into the cell is matched by ion movement out of the cell. E_x is sometimes also called the **reversal potential**, because once the cell membrane crosses from this potential, the net movement of the ions reverses directions. This value is also called the **Nernst potential** in honor of Walther Nernst, the Nobel prize-winning German chemist who first developed the formula to describe electrochemical reactions.

R is the ideal gas constant. It has a value of 8.314 J / K * mol. The R term is included in the equation because it is a way to convert between the number of molecules and the energy that these molecules exert.

T is the temperature in Kelvin (to convert from Celsius into Kelvin, add 273 to the Celsius value. C + 273 = K). Usually, room temperature is around 293 K, and biological temperatures are closer to 310 K. This term is included in the equation since all biological processes are temperature dependent, especially those at the level of molecular machinery.

z is the electrical charge of the ion. For sodium and potassium, z is +1, for chloride or monovalent anions z is -1, and for divalent cations like calcium and magnesium, z is +2. This term is essential since the Nernst potential is used to predict the direction of ion movement based on the charge of the ion. This value represents the influence exerted by the electrical gradient.

F is the Faraday constant which has a value of 96,485 Coulombs / mol.

$[x]_o$ and $[x]_i$ represent the concentration of ion x outside the cell and inside the cell, respectively. Their units are generally in mM, but these units cancel out in the equation. These values represent the forces acting on the ions by the chemical gradient.

The Nernst equation has many complex constants and terms, so neuroscientists often use the **back-of-the-envelope equation** as a shortcut for quickly calculating the equilibrium potential. This shortcut equation condenses down the R and F constants, turns the natural log into a base 10 logarithm, and assumes the calculations are done at physiological temperature. The shortcut formula can be written as:

$$E_x \cong \frac{61}{z} \log\left(\frac{[x]_o}{[x]_i}\right)$$

Equation 4.10 The "back-of-the-envelope" equation is a shortcut to estimate the reversal potential for an ion x.

The Nernst equation is able to calculate the reversal potential for individual ions assuming that the appropriate ion channels are open. However, in neurons, not all ion channels are opened or closed at the same time. In a neuron at rest, usually Na^+ does not enter into the cells since sodium channels are closed. K^+ ions , on the other hand, are often moving through leak channels all the time. Chloride channels are generally open as well, but they pass less current at rest than potassium channels do.

A formula called the **Goldman-Hodgkin-Katz equation (GHK equation)** combines the Nernst potentials of three relevant ions (Na^+, K^+, and Cl^-) into a single equation that, when evaluated, gives us the value of the membrane potential V_m.

The GHK equation is written as:

$$V_m = \frac{RT}{F} \ln \frac{p_K[K^+]_o + p_{Na}[Na^+]_o + p_{Cl}[Cl^-]_i}{p_K[K^+]_i + p_{Na}[Na^+]_i + p_{Cl}[Cl^-]_o}$$

Equation 4.11 The Goldman-Hodgkin-Katz equation is used to calculate the membrane potential given permeability of ions and their concentrations across the cell membrane.

When you analyze the GHK equation, you will notice that it is essentially a combination of the Nernst equations for the equilibrium potentials for the three ions. The GHK equation also introduces a new term, the value **p** which stands for **permeability**: the ability for an ion to cross the membrane through ion channels. Permeability itself does not have a unit.

It is easier to think of permeability as the "weight" of each equilibrium potential. The higher the permeability for a given ion, the closer the V_m is to the E_x for that ion. For example, consider the value of the GHK equation for neurons at rest. Under these conditions, the permeability for K^+ (p_K) is 1, p_{Cl} is 0.55, and p_{Na} is 0.04. Therefore, the resting membrane potential V_m will be closest to a combination of E_K and E_{Cl}, since these two terms dominate the GHK equation. Just as a reminder, the resting membrane potential is around -70 mV, and E_K is around -80 mV while E_{Cl} is around -60 mV.

However, during an action potential (described in section 4.4), permeability for sodium increases significantly. As p_{Na} rises, the membrane potential will shift closer towards E_{Na}, which is +55 mV. The GHK equation provides a mathematical explanation for how movement of Na^+ across the membrane causes the cell to become more positive.

4.4 The action potential

The **action potential** is a short-lasting (1 or 2 milliseconds), temporary change in membrane potential that can travel down the length of the axon. Action potentials are the main method of communication that neurons use, and each action potential triggers the release of neurotransmitters at the axon terminal of a chemical synapse. Action potentials are all-or-nothing responses, meaning that only a large change in potential will be able to pass the signal forward, but a small, sub-threshold change in V_m will not. Sub-threshold changes in V_m are called **graded potentials**.

During an action potential, the membrane potential deviates from the resting membrane potential of around -70 mV (this value can range from -90 mV to -50 mV, depending on the type of neuron) to positive potentials. When Vm goes from negative to a more positive potential, this change is described as a **depolarization**. Likewise, when Vm changes to a value that is more negative, it is called a **hyperpolarization**. The action potential has a very characteristic shape: a rapid depolarization, followed by a prolonged hyperpolarization, before repolarizing and slowly returning to the resting membrane potential over the next few milliseconds.

Cells regularly receive many inputs at once. To trigger an action potential, the sum of the inputs must be sufficiently depolarizing to bring the Vm to a value that is more positive than the **action potential threshold**. For an average neuron, the action potential threshold is around -55 mV. Anything less than the action potential threshold will fail to send the signal forward.

The change in membrane potential seen in an action potential is driven by the movement of ions, predominantly sodium and potassium, across the cell membrane through voltage-gated ion channels. An action potential takes place in 5 steps, some of which overlap in duration.

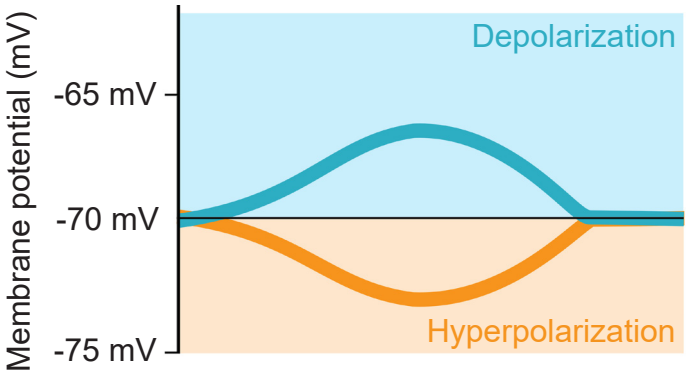


Figure 4.12 When the neuronal membrane potential becomes more positive above the resting membrane potential, we call it depolarization. When the membrane potential becomes more negative, we say it is hyperpolarized.

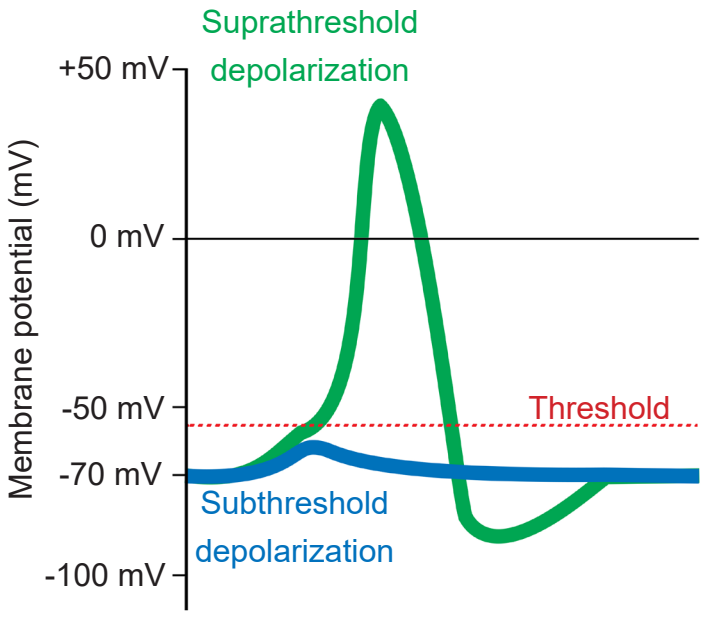


Figure 4.13 If the depolarization exceeds the action potential threshold, roughly -55 mV, the neuron will fire an action potential (green).

1. Depolarization from incoming neurons

Presynaptic neurons that release neurotransmitters onto dendrites cause a small amount of ion movement via postsynaptic ligand-gated ion channels. These small deviations in membrane voltage are called **postsynaptic potentials,** or **PSPs**. Release of excitatory neurotransmitters causes small depolarizations which are called **excitatory post synaptic potentials (EPSPs)**, while release of inhibitory neurotransmitters cause small hyperpolarizations, or **inhibitory post synaptic potentials (IPSPs)**. These PSPs are characterized by a fast but small rising phase where the V_m changes quickly followed by a slower decay phase where the V_m resets back to resting potential.

Usually, a single EPSP only causes a small amount of depolarization that isn't sufficient for bringing Vm above the threshold potential. Reaching the threshold often requires multiple EPSPs. Multiple EPSPs are added together in one of two ways: spatial summation and temporal summation. **Spatial summation** happens when two small EPSPs from two adjacent inputs are triggered. **Temporal summation** occurs when multiple EPSPs from the same input occur close together in time. Each individual EPSP has a normal decay period, and adding a second EPSP during that decay period causes the two to add together, which may now be large enough to bring Vm above the action potential threshold. When a large enough total depolarization of membrane potential reaches the soma, an action potential is initiated at the axon hillock.

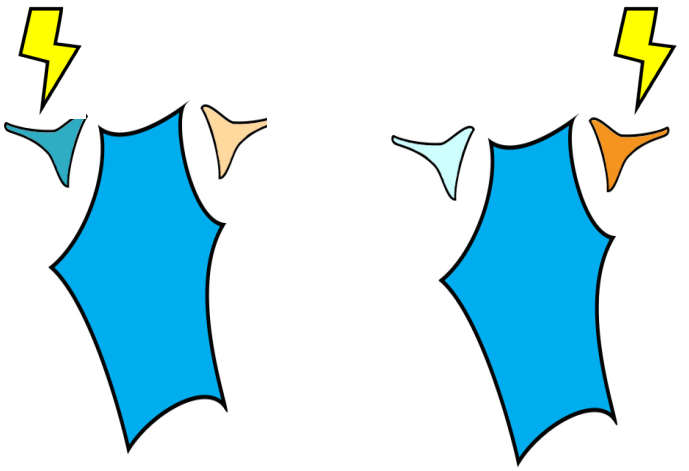


Figure 4.14 Activation of excitatory inputs (left) causes depolarization, while activation of inhibitory inputs (right) can cause hyperpolarization.

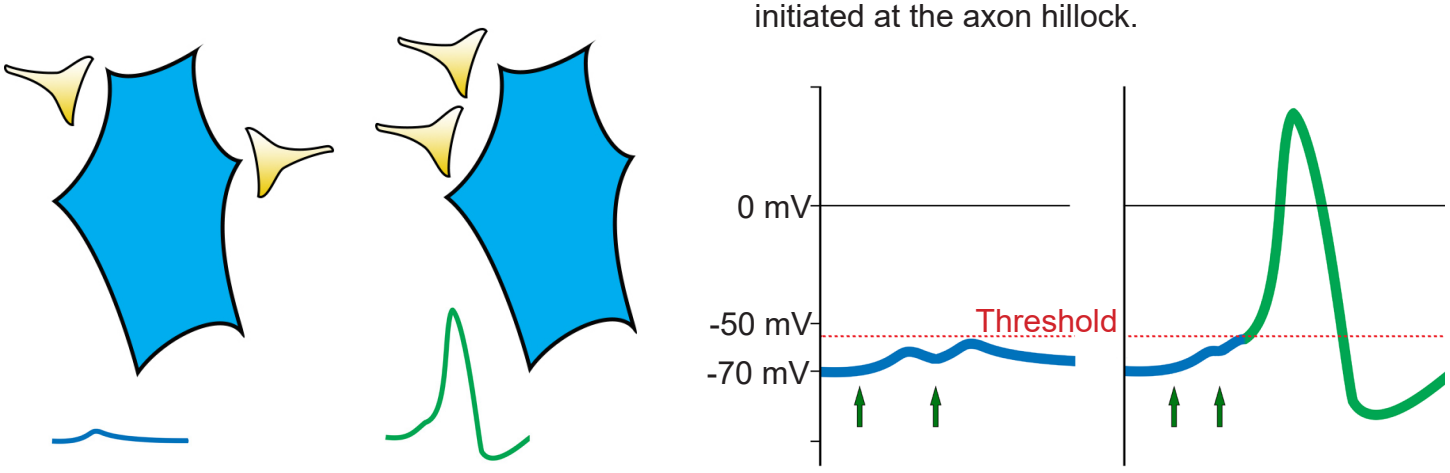


Figure 4.15 Spatial summation is the addition of signals from two inputs adjacent to each other (left). Temporal summation is the addition of two signals close together in time (right).

2. Opening of voltage-gated Na^+ channels

At rest, the voltage-gated ion channels are almost all closed. But as the V_m depolarizes, the channels are more likely to open. At this point, we use our understanding of the electrochemical gradient to predict the movement of Na^+ ions. The positively charged sodium ions are drawn to the negative potential on the inside of the cell, and the ions move from the area of high concentration (outside the cell) to an area of lower concentration (inside the cell). Both forces are in favor of Na^+ entering the cell when the voltage-gated sodium channels open, and this movement of positive charges into the cell causes further depolarization.

Na^+ movement across the cell membrane into the cell causes the V_m to depolarize to very positive potentials. The peak of the action potential can reach +40 mV.

3. Opening of voltage-gated K^+ channels

Embedded in the cell membranes are other voltage-gated channels that are selective for K^+ rather than Na^+. Like the voltage-gated Na^+ channels, these K^+ channels begin to open when the cell starts to depolarize, allowing K^+ ions to move across the cell membrane.

When the cell has depolarized to about +40 mV, we can use our knowledge of the electrochemical gradient to make predictions about what happens to K^+ ions. The positive

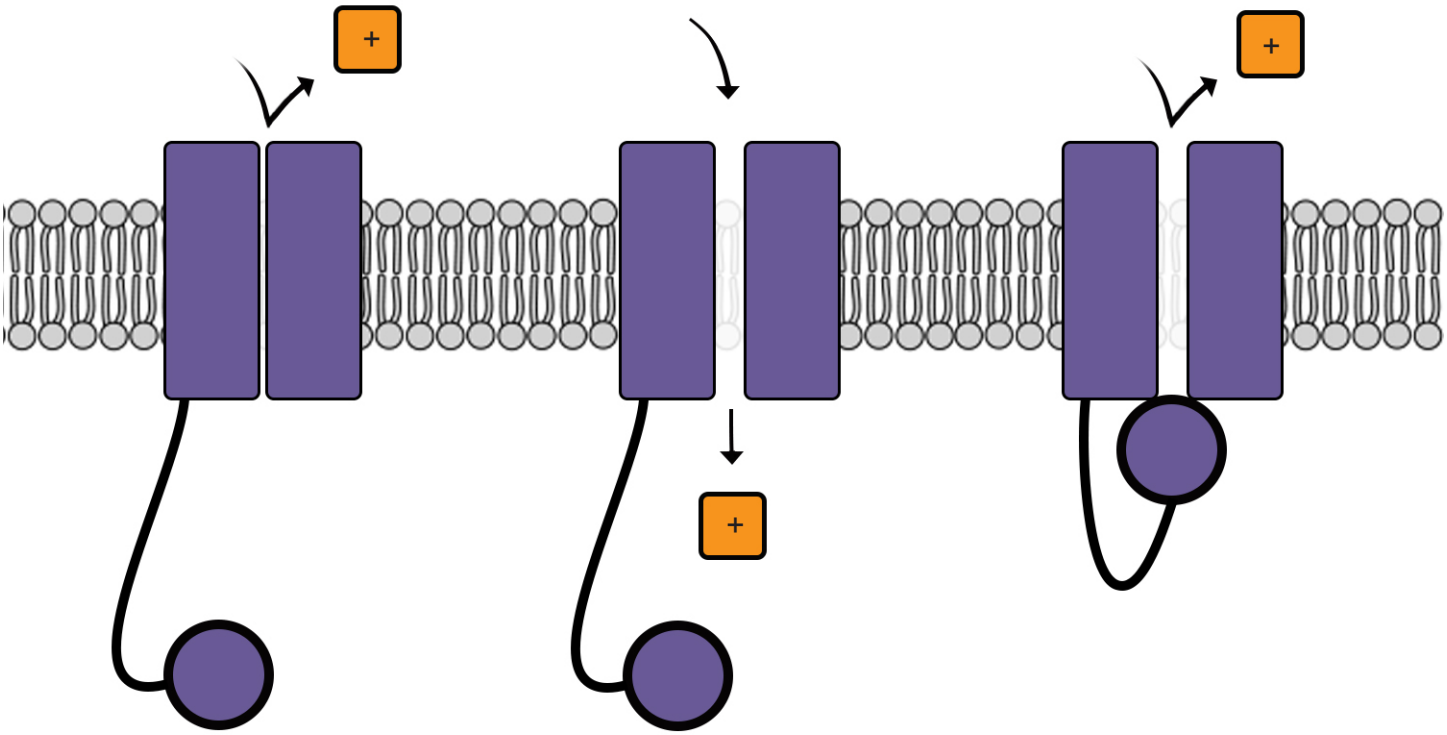


Figure 4.16 Voltage-gated Na^+ channels have a molecular structure that allows the channel to temporarily change structure in response to membrane voltage. A small depolarization to -50 mV causes the channels to open, while a large depolarization above -30 mV causes the channels to inactivate, blocking further sodium currents.

potential inside the neuron repels the K^+ ions (electrical gradient), and the relatively high concentration of K^+ inside the cell pushes K^+ out of the cell (chemical gradient). In total, both of these forces act on the K^+ ions to leave the neuron. Movement of positive charge carriers out of the cell makes the interior of the cell (and V_m) more negative. K^+ movement through the voltage-gated potassium channels causes the cells to become more negative than the resting membrane potential.

4. Inactivation of voltage-gated Na^+ channels

Voltage-gated Na^+ channels have a very complex molecular structure. In addition to having a pore through which ions can move across the cell membrane, they have an inactivation gate that can block the flow of sodium ions. When the cell membrane reaches a positive potential, the inactivation gate closes, thus preventing further movement of excitatory, depolarizing Na^+ ions.

These voltage-gated Na^+ channels undergo the inactivation process very rapidly, often times faster than a millisecond.

5. Deactivation of voltage-gated K^+ channels

In this last step of the action potential, the main current flow is an outward current as the positively-charged K^+ ions are driven out of the cell through voltage-gated K^+ channels by the electrochemical gradient. When positive charges leave the cell, the interior once again becomes more negative.

Unlike the voltage-gated Na^+ channels, the process of deactivation is much slower. On average, it may take these channels a few milliseconds to deactivate.

When these K^+ channels deactivate, the hyperpolarizing current stops, which causes the membrane potential to gradually return to the equilibrium of resting potential.

Clinical connection: Chronic pain

Nociception is a complex sensation that the nervous system detects rapidly. Pain is the perception that the body is experiencing some kind of injury or noxious stimulus, so pain triggers a reflex that causes the person to withdraw from the painful stimulus, which decreases the severity of the damage. Pain can also create memories that discourage people from future contact with pain-inducing situations.

When we experience a noxious stimulus, that information is sent towards the central nervous system by means of action potentials. And in most cases, the detection of pain is a healthy, protective sensation.

There are, however, some people with a dysregulation in their somatosensory systems that cause them to experience pain even in the absence of injurious stimuli, a condition called **allodynia**. Although we still don't understand what causes allodynia, it is known that some change in voltage-gated sodium channel properties causes an increase of excitability of pain-sensing neurons is increased.

In summary, you can also think of the shape of the action potential as consisting of three phases.

1. Depolarization. The upwards deflection (-70 mV to +40mV) of the action potential that lasts for half a millisecond. Vm becomes more positive because of Na^+ ions that enter the cell through the voltage-gated sodium channels. At

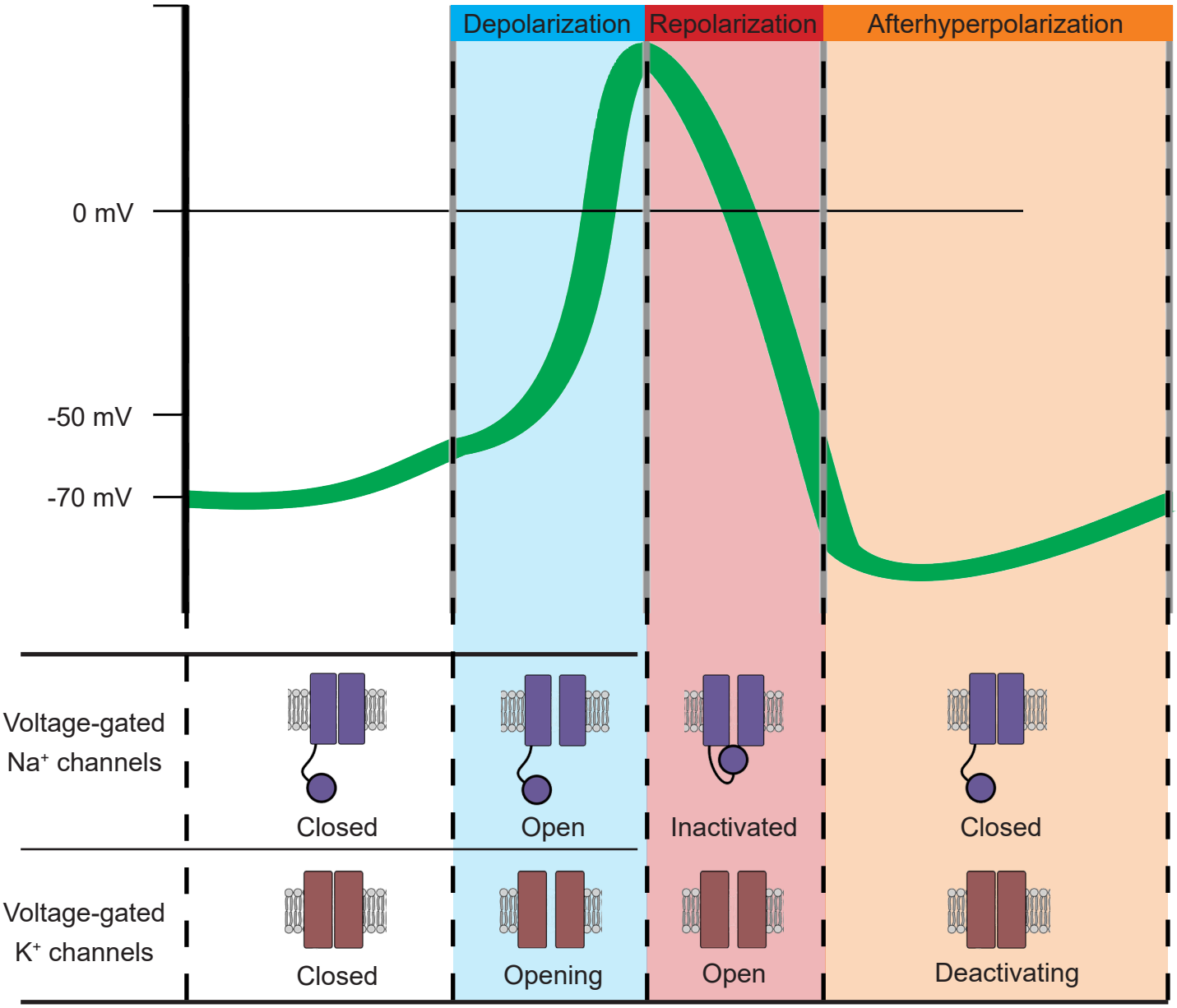


Figure 4.17 Voltage-gated ion channels open, inactivate, and deactivate at different times during the action potential.

this step, the voltage-gated potassium channels also start to open.

2. Repolarization. The rapid downwards deflection (+40 mV to -70 mV), also lasting for about half a millisecond. Here, the voltage-gated sodium channels have almost all inactivated, and positively-charged K^+ ions are being driven out of the cell through the voltage-gated potassium channels. As positive charges leave the cell, V_m becomes more negative.

3. Afterhyperpolarization. The slow return back to the resting membrane potential (-70 mV to -80 mV and back to -70 mV) that can last for a few milliseconds. This return back to the equilibrium of the resting state is due to the gradual deactivation of voltage-gated potassium channels.

Ions moving across the membrane change the membrane potential. The best way to think about the change in V_m in response to

Randomness in ion channel properties

Voltage-gated ion channels do not follow a precise "if-then" flowchart in determining when to open, inactivate, or deactivate. Rather, ion channels are very probabilistic in nature. For example, consider as a voltage-gated sodium channel changes from the closed state to the open state. A single channel like this might show probabilistic behavior as follows: at -70 mV, there is a 0.1% chance to open. At -30 mV, there is now a 50% chance to open. At +20 mV, there is a 99.9% chance to transition to the open configuration.

Likewise, these channels also transition from the open state to the inactive state on a probabilistic basis as well, with the greatest probability of inactivating at depolarized potentials.

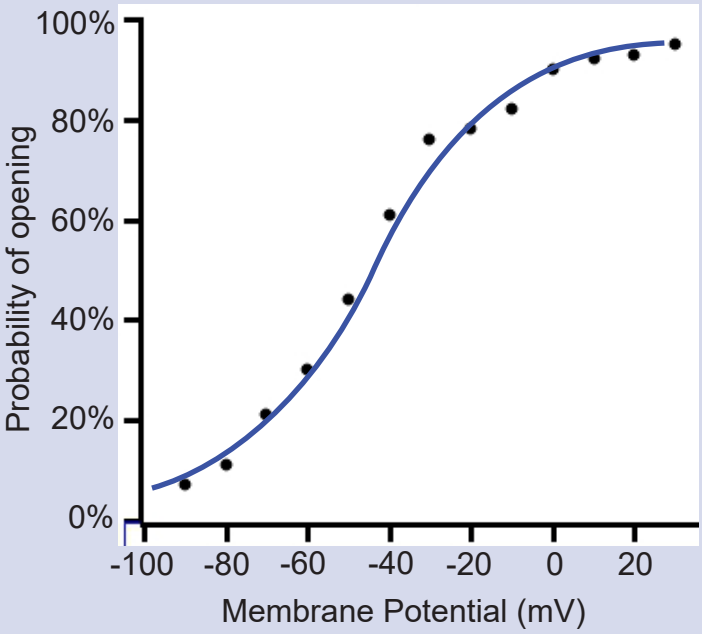


Figure 4.18 Opening of voltage-gated ion channels is probabilistic, with a greater likelihood of opening at depolarized potentials.

ion movement is to consider the equilibrium potential of each individual ion that is moving. For example, E_{Na} is +55 mV. When a voltage-gated sodium channel opens and Na^+ ions are able to move across the cell membrane, the membrane potential shifts towards E_{Na}, which is a depolarization. During the repolarization phase, when the voltage-gated K^+ channels open, the potential across the cell membrane now shifts towards the equilibrium potential for K^+, which is close to -80 mV.

The steps of the action potential can also be thought of in mathematical terms by examining the variables of the GHK equation. As permeability for a given ion increases, it shifts the balance of V_m more heavily towards the equilibrium potential of that specific ion. So, during the depolarization phase of the action potential when the voltage-gated sodium channels open, the permeability term for sodium (p_{Na}) increases to as high as 20, which causes Vm to move towards E_{Na} (+55 mV). And during the repolarization phase, voltage-gated potassium channels open, while the voltage-gated sodium channels are inactived. When this happens, p_K increases, p_{Na} drops to zero, and as a result, the V_m shifts towards the equilibrium potential for potassium (-80 mV).

In the moments following the beginning of an action potential, there is a time window where a second action potential cannot be fired. This time window, about half a millisecond in duration, is called the **absolute refractory period**. On a molecular level, the absolute refractory period exists because the voltage-gated sodium channels are inactivated, which prevents them from being able to pass any more inward excitatory current. The absolute refractory period may last for up to a millisecond.

A few milliseconds after the end of the absolute refractory period begins a phase called the **relative refractory period**. During this phase, it is more difficult to fire an action potential

compared to the resting condition. At the level of the receptors, the relative refractory period exists for two main reasons. First, only some of the voltage-gated sodium channels have “reset” from their inactive state. Second, many of the voltage-gated potassium channels are still in the open state, which allows the movement of K^+ ions, causing the neuron to rest at a more negative potential. Also, increased movement of K^+ ions hinders depolarization, as the GHK equation demonstrates, since potassium movement pulls membrane potential closer towards the equilibrium potential for K^+. This relative refractory period lasts as long as the afterhyperpolarization, and therefore exists on a gradient: the sooner after the repolarization phase of the first action potential, the more difficult it will be to reach threshold for a second action potential.

4.5 Movement of the action potential

Up to this point, we have only considered a single piece of cell membrane and how it responds during an action potential. For an action potential to travel from the axon hillock to the axon terminal, this temporary change in V_m must physically move down the length of the axon. The ability for the action potential to travel is made possible because the Na^+ ions that enter through voltage-gated sodium channels are not restricted to the cytoplasmic volume beneath the ion channels. These Na^+ ions behave just like any other substances that are in an area of high concentration: they move towards areas of low concentration. As Na^+ rushes into the cell, they follow the chemical gradient and quickly move toward areas of low concentration, which is the next section of membrane. When they do so, the next patch of the membrane becomes depolarized. This depolarization causes the next set of channels to open, which results in even more sodium influx. From here, the process repeats down the axon. The action potential is a chain reaction that starts at the axon hillock and doesn't stop until the axon terminal.

An action potential is like a wave. It only moves in one direction, and there are two main reasons why:

1. First, sodium ions move down their concentration gradient. The Na^+ ions that enter through a voltage-gated sodium channel are less likely to want to move in the reverse direction, since this area currently has a relatively high Na^+

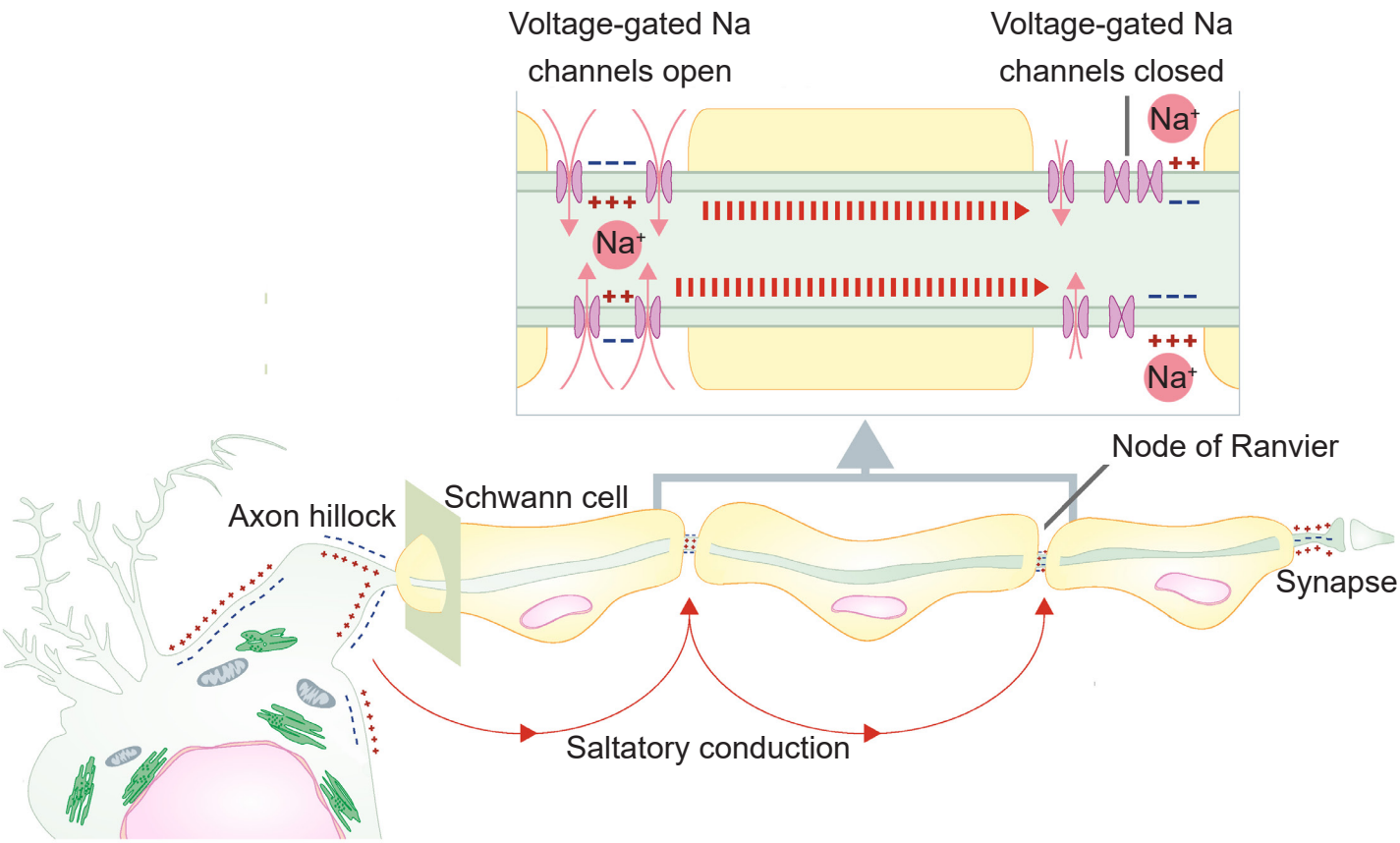


Figure 4.19 Movement of an action potential down a myelinated axon relies on saltatory conduction.

Analgesia and motor signaling

A trip to the dental office sometimes requires an injection of an analgesic like lidocaine before the dentist performs any otherwise-painful procedures. On a molecular scale, lidocaine is an inhibitor of voltage-gated sodium channels. By blocking these channels, lidocaine prevents action potentials from traveling up the afferent pathway, preventing incoming sensory inputs from the teeth and gums from reaching the central nervous system. When these signals are blocked, we don't feel any drilling.

Figure 4.20 Pufferfish and other species likely evolved a symbiotic relationship with bacteria that produce the deadly TTX to prevent predation.

Just as incoming signals can be inhibited through voltage-gated sodium channel inhibitors, so can outbound signals – sometimes with deadly consequences. Through evolution, nature came across a variety of voltage-gated sodium channel inhibitors that function as poisons to prevent predation. Because the efferent pathway is used to signal components of the motor system, blocking downward signals can lead to paralysis. One family of fishes, the pufferfish, may have a very high concentration of a deadly voltage-gated sodium channel inhibitor called **tetrodotoxin**, or **TTX**. This toxin is produced by a bacteria that live symbiotically in pufferfish organs. Ingestion of TTX prevents neurons from communicating. One efferent signaling pathway, the **phrenic nerve**, signals the diaphragm to move up and down. Without proper medical treatment, people exposed to doses as small as 1 milligram of TTX can die of respiratory failure in hours.

Despite these tremendous risks, pufferfish is a delicacy in Japan. The dish, called fugu, can cause numbness of the lips and even a mild intoxication.

concentration. Instead, they will move towards the area with a low concentration of ions, which is the forward direction.

2. Secondly, the previous patch of membrane is in the absolute refractory period, which makes it impossible for an action potential to travel backwards. During this time, many voltage-gated sodium channels are still inactivated, and most voltage-gated potassium channels are open.

The role of myelin

Previously, in discussing the cellular anatomy of the neurons, we described myelin, layers of lipid that are sometimes thickly wrapped around the axon of the neuron. Myelin increases **conduction velocity**, the speed at which an action potential travels down the length of the axon. Myelin works by physically blocking leak potassium channels in the cell membrane. By covering these channels, positive charges

become unable to exit the cell, which causes the signal to move more rapidly.

A myelinated axon still requires some influx of sodium ions for the signal to travel, however. This influx occurs at the **nodes of Ranvier**, the unmyelinated segments of the axon. These areas are very dense with voltage-gated sodium channels. Because the changes in electrical charge are detected at intervals, we sometimes say that the signal is passed by **saltatory conduction** (*saltare* is the Italian word for "jump").

In this chapter, we described the molecular and cellular underpinnings of the action potential, the main method of signaling that the nervous system uses for communication. In the following chapter, we will explain how neurons talk to one another using those electrical signals in combination with chemical signals called neurotransmitters.

The Open Neuroscience Initiative is funded by a grant from the Vincentian Endowment Fund of DePaul University.

Title page: Image by pasja1000 from Pixabay, Image by TeroVesalainen from Pixabay
4.4. OpenStax / CC BY (https://creativecommons.org/licenses/by/4.0) https://commons.wikimedia.org/wiki/File:Openstax_college-physics_22.5_magnet-pole-attraction.jpg
4.5 (diffusion fig BruceBlaus [CC BY 3.0 (https://creativecommons.org/licenses/by/3.0)]
4.18 Correa AM, Bezanilla F. Gating of the squid sodium channel at positive potentials: II. Single channels reveal two open states. Biophys J. 1994 Jun;66(6):1864-78. doi: 10.1016/S0006-3495(94)80980-4. PMID: 8075324; PMCID: PMC1275912.
4.19 https://commons.wikimedia.org/wiki/File:Propagation_of_action_potential_along_myelinated_nerve_fiber_en.svg modified by Austin Lim
4.20 https://upload.wikimedia.org/wikipedia/commons/4/4e/Pufferfish_%28Butete%29.jpg

Chapter 5:

Signaling Between Neurons

Austin Lim, PhD (DePaul University)
Editor: Monica Javidnia, PhD (University of Rochester)

Previously, we described the electrical properties of a single neuron. A lone neuron can send action potentials as a means of communication, but cells become much more interesting when they have partners to talk to. The nervous system of the worm C. elegans is only 300 neurons, and yet it is complex enough to engage in moderately intricate behaviors like responding to repellant or attractant odors, social feeding, and long-term learning. The human brain, with its 86 billion neurons, can engage in these behaviors and so many more - only because of communication between the different neurons in the brain.

In this chapter, we will focus on the molecular-level features of communication between neurons, starting from the anatomical differences between synapses.

A lone neuron can send action potentials as a means of communication, but cells become much more interesting when they have partners to talk to. The nervous system of the worm C. elegans is only 300 neurons, and yet it is complex enough to engage in moderately intricate behaviors like responding to repellant or attractant odors, social feeding, and long-term learning. The human brain, with its 86 billion neurons, can engage in these behaviors and so many more - only because of communication between the different neurons in the brain.

In this chapter, we will focus on the molecular-level features of communication between neurons, starting from the anatomical differences between synapses.

Chapter 5 outline

5.1 Electrical vs. chemical synapses

The synapse is not a part of the structure of a neuron, but rather the site of close proximity between two communicating neurons. There are two main types of synapses: electrical and chemical.

Electrical synapses

Electrical synapses are the simpler of the two types of synapses. In an electrical synapse, the main driver of communication between two neurons is a change in potential, and the carrier of charge is almost always an

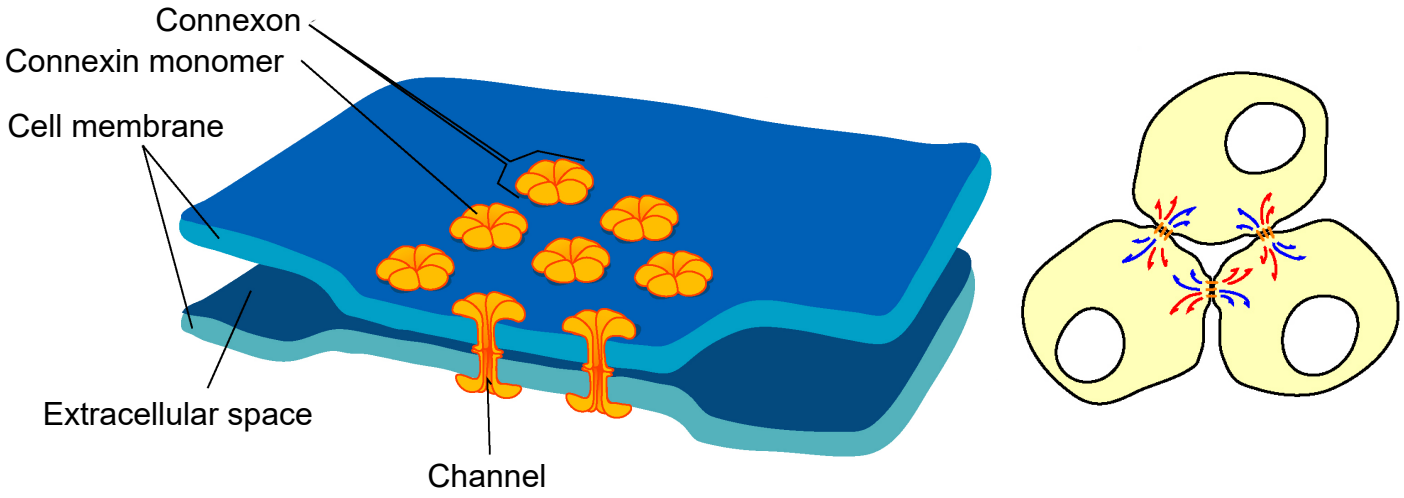


Figure 5.1 An electrical synapse exists between two closely-connected neurons. Cytoplasm passes between the two neurons through a protein complex called a connexon.

ion. The electrical synapse simply means that the cells share their cytoplasm with an adjacent cell. Electrical synapses are what Camillo Golgi imagined when he proposed the reticular theory of nervous system organization. Oftentimes an entire network of many hundreds of neurons is connected by these synapses.

Imagine two neurons that are connected by an electrical synapse. First of all, both of them are complete cells on their own. Each one contains a complete plasma membrane surrounding the neuron, a nucleus, and all the individual organelles needed to carry out that cell's basic life processes. Electrical synapses share the cytoplasm between the two connected cells, so ions, ATP, and larger signaling molecules and proteins are able to move between the two cells. For this to happen, there exists a specialized physical channel between the two that allows for the passage of cytoplasm called a **connexon** or **hemichannel**. Each hemichannel itself is made up of six transmembrane proteins called **connexins** (you can remember the difference between the channel and the individual protein because proteins often end with the letters -in). When two connexons contact each other, they interact closely with each other and form the **gap junction**, which is the structure that connects neurons electrically.

Neurons that are connected by electrical synapses are remarkably close to each other. The synaptic gap between the two electrically connected neurons is about 3.5 nanometers. Logically, the neurons must be close since the hemichannels are like a physical "bridge" between the two cells.

Figure 5.2 Electrical synapses are similar to a skybridge that physically connects the cytoplasm between two neurons.

Electrical synapses are capable of passing information bidirectionally. This means that a signal does not always move sequentially from the presynaptic cell to the postsynaptic cell. Rather, ions and signaling molecules are free to move through the connexons in either direction. Also, each cell within an electrically-coupled network can receive inputs at any of the cells, making it able to detect several signals at once - the same way a huge satellite dish can detect more signals than a small dish.

Electrical synapses likely evolved because of evolutionary pressures that selected for speed. These synapses can pass signals as fast as electrical charges can move through an electrolyte-rich fluid like cytoplasm, which is almost instantaneous. Therefore, an escape reflex that is made up of communication across electrical synapses is advantageous for animals that need to escape predators. For example, crayfish exhibit a reflexive abdominal flexion response when exposed to threatening stimuli, causing the animal's body to dart away from a threat within a fraction of a millisecond. Comparing across the phylogenetic tree, electrical synapses are often found in less complex organisms, including arthropods such as insects and crustaceans, where such reflexes are likely more critical for survival.

Another advantage of electrical synapses is that they can form a large network of interconnected neurons with synchronized activity. For example, neuroendocrine cells in the hypothalamus are connected by electrical synapses. When the "go" signal arrives, all the cells depolarize at once, which can result in the massive release of hormones into the bloodstream. A network can also cause sudden, powerful inhibition. Like an angry mob of people chanting, a network of electrical synapses connecting inhibitory interneurons allows the network to send an immediate "shut-down" signal under specific circumstances.

Clinical connection

Charcot-Marie-Tooth (CMT) disease

Charcot-Marie-Tooth (CMT) disease is a rare genetic disorder that damages parts of the peripheral nervous system including the motor nerves, resulting in muscle weakness and difficulty with walking, and the sensory nerves, causing some to experience abnormal sensations such as tingling or pain in their extremities. These symptoms are characteristic of signal transduction failure resulting from deficits in myelin. One type of connexin protein, called Cx32, is heavily expressed in Schwann cells, the glia that produce myelin in the PNS. Mutations in the gene that codes for Cx32 are associated with the X-linked form of CMT disease, and knocking-out the gene in experimental mice cause the mice to express similar symptoms as human CMT.

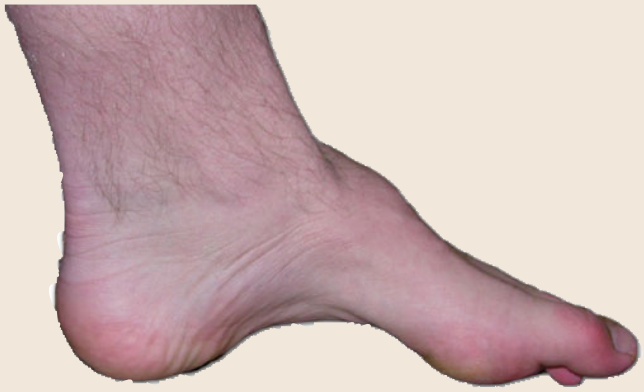


Figure 5.3 People with CMT disease often have abnormally shaped feet.

Chemical synapses

At a chemical synapse, a signaling molecule is released by the presynaptic cell to influence the postsynaptic cell. These signaling

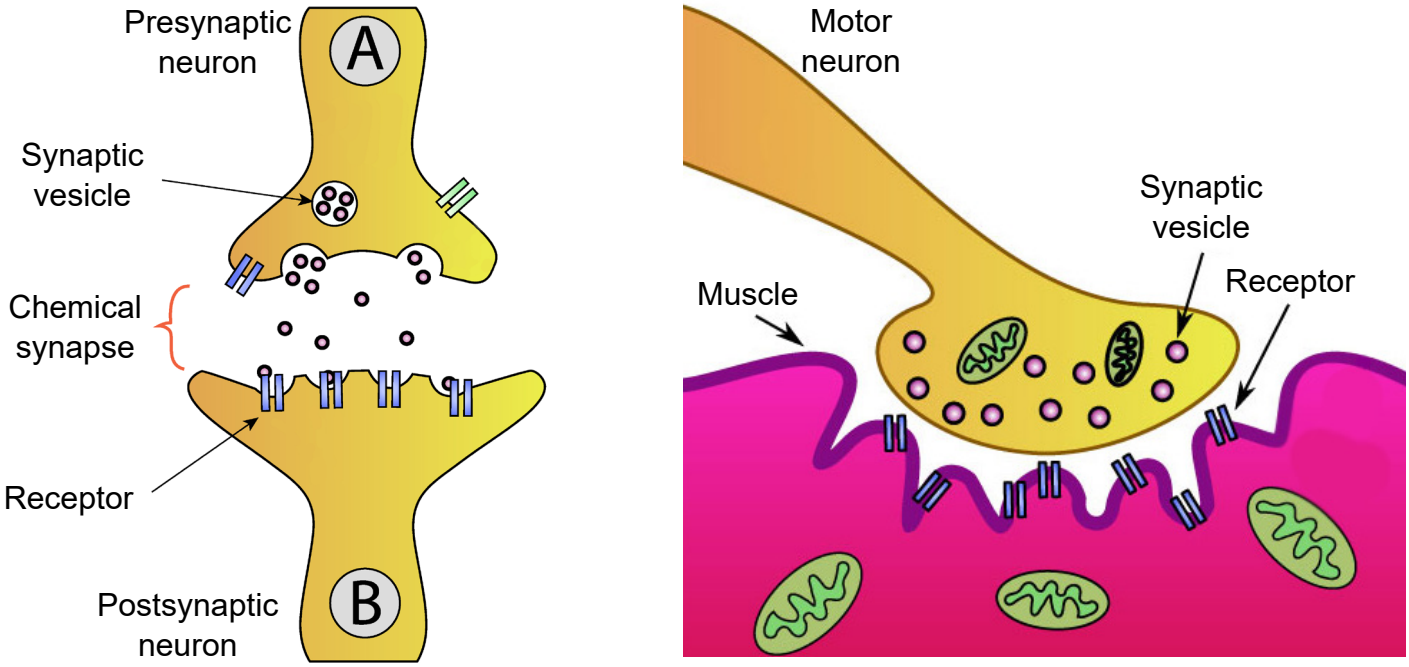


Figure 5.4 Chemical synapses are the site of close interaction between two neurons (left) or a motor neuron and a muscle fiber, which is called the neuromuscular junction (right).

molecules, generally called neurotransmitters, are synthesized or stored by neurons. After being released, these neurotransmitters diffuse randomly across the synapse, where they are able to affect nearby neurons once the chemical binds to its corresponding receptor.

Since chemical synapses do not rely on a direct physical protein "tunnel" to connect the two neurons, the distance between the two cells can be much larger. On average, a chemical synapse is a distance of about 20-40 nanometers, roughly a thousand times smaller than the diameter of a human hair.

A chemical synapse can pass a variety of signals, depending on the neurotransmitter and the receptor. For example, some signals are directly excitatory and allow positively charged cations to enter the neuron causing depolarization. Other signals are hyperpolarizing, and therefore inhibitory. And yet other signals are much more complex, inducing changes in protein expression that can modify cellular excitability over the course of minutes or hours.

Because of the complexity of the signals that chemical synapses can convey, evolutionary development through time has allowed for a tremendous variety of responses. Chemical synapses allow for fine-tuning of neural networks, giving these nervous systems a larger range of possibilities. The nervous systems of "higher" organisms like humans tend to have several chemical synapses since these signals are likely necessary for complex behaviors and cognition.

Many chemical synapses exist between the axon of one neuron and the dendrite of another neuron. One specific type of chemical synapse refers to the space between a motor neuron and muscle tissue, and this is called the **neuromuscular junction**, or **NMJ**. When the chemical signaling molecule acetylcholine (ACh) is released by the presynaptic motor neuron, it is detected by receptors that are expressed on the muscle. The release of ACh causes contraction of the muscle.

5.2 Properties of vesicles

Types of vesicles

Molecules of neurotransmitters are often stored in synaptic vesicles before being released. Synaptic vesicles are tiny spheres of lipids just like the cell membrane. These vesicles can be roughly characterized into one of two classes:

1. **Small vesicles**. These vesicles have a diameter of 40 nanometers and a volume of about 30 cubic microns. Given the size of neurotransmitters, we can estimate at somewhere on the order of thousands to tens of thousands of molecules of neurotransmitter can be stored in each vesicle. Small vesicles store most of the neurotransmitters we most often think of, including glutamate, GABA, dopamine, and norepinephrine, for example. Small vesicles are almost always exclusive found in the axon terminals.

2. **Large dense-core vesicles**. These vesicles are much larger than small vesicles, with a range of diameter from 100 to 250 nanometers. They store peptides such as dynorphin or enkephalin, which have chemical structures much larger than the other neurotransmitters. Since these peptides are packaged into their vesicles near the nucleus, large dense-core vesicles can be found in the cell bodies and all along the axons in addition to the axon terminal.

Loading of vesicles

Vesicles need to be filled with molecules of neurotransmitter before release into the synapse. In small vesicles, filling is only made possible through the action of giant transmembrane proteins called vesicular transporters. These are protein complexes that span the vesicular membranes, with one side facing the intracellular space and other facing the inside of the vesicle. Their main function is to take molecules of neurotransmitter from the intracellular space of the axon terminal and pump them into vesicles.

Many of the vesicular transporters are named based on the neurotransmitters that they are capable of recognizing and transporting. Some have a single substrate, such as **vesicular GABA transporters (VGAT)** which move GABA, **vesicular glutamate transporters (VGluT)** which move glutamate, and **vesicular acetylcholine transporter (VAChT)** which moves acetylcholine into vesicles. Others recognize a broad class

Electron microscopy

Electron microscopy is a technique that allows for the imaging of subcellular structures that are on the order of nanometers, such as synapses, small vesicles, and large dense core vesicles.

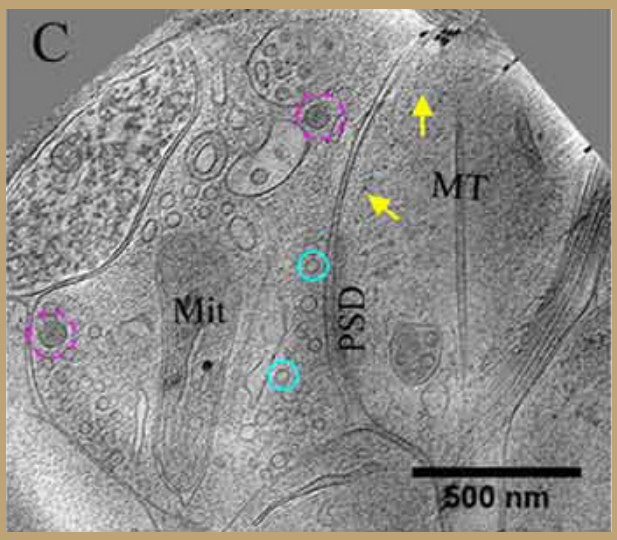


Fig 5.5 Electron microscope image showing small vesicles (cyan) and large dense core vesicles (magenta).

of neurotransmitters, such as the **vesicular monoamine transporters (VMATs)**, which are responsible for moving monoamines such as dopamine and serotonin into the vesicles.

Vesicular transporters are able to function because the interior of the vesicle is highly acidic compared to the interior of the cell. Vesicles have a high concentration of H+ ions (protons) because of the action of the transmembrane enzyme **vesicular ATP-ase**, or **v-ATP-ase.** These membrane-embedded proteins utilize the molecular energy contained in ATP to concentrate H+ ions in the intravesicular space. For each molecule of ATP used, one proton gets pumped into the vesicle.

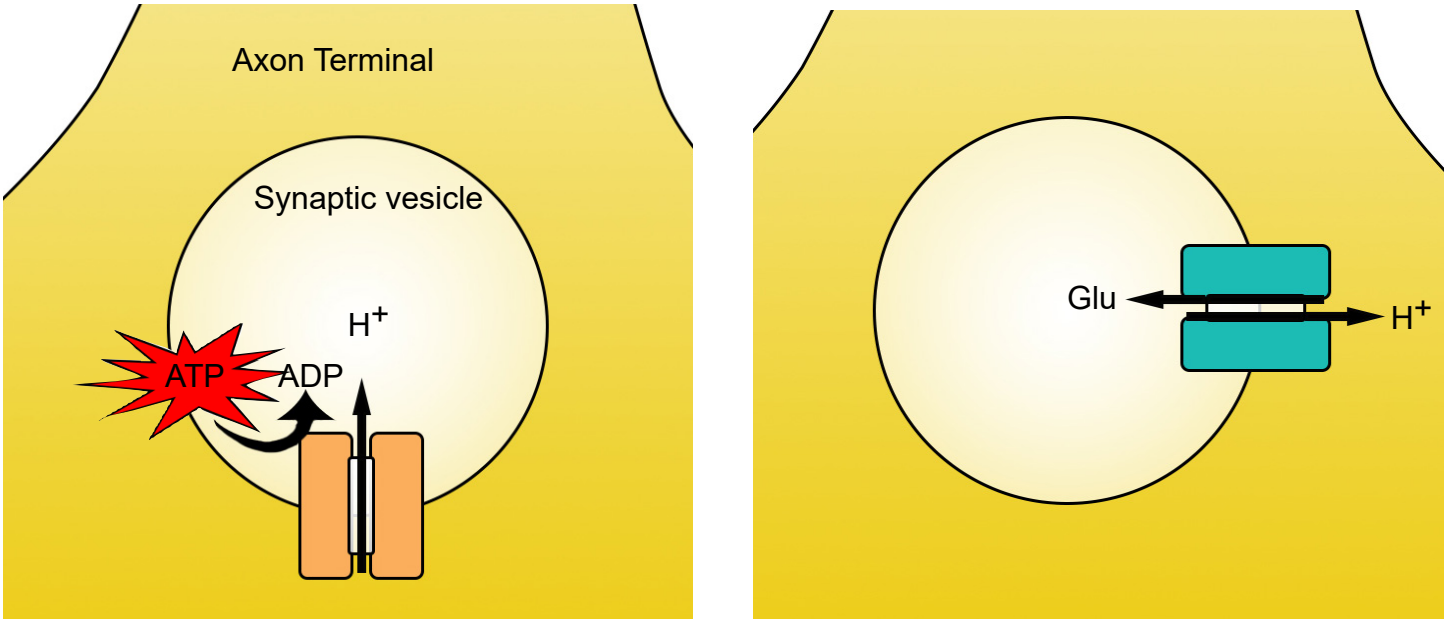
Vesicular transporters pump molecules of neurotransmitter against their concentration gradients, which is an energetically difficult task. To have enough energy, the vesicular transporters use the high intravesicular concentration of H+ to move molecules of neurotransmitter across the vesicular membrane. When a proton moves from an area of high concentration to low concentration, energy is generated. The vesicular transporters use this energy to push neurotransmitter in. Because H+ ions move opposite of the neurotransmitter molecules, vesicular transporters are called **antiporters**. Transporters have slightly different stoichiometries, as it requires two protons to move a single molecule of dopamine, while the energy from a single proton is sufficient to transport GABA or glutamate.

Location of vesicles

Synaptic vesicles can be found in one of three places at the axon terminal.

1. **Readily releasable pool (RRP)**. These vesicles are located close to the cell membrane at the axon terminal. In fact, many of them are already "docked," meaning that their coat proteins are already interacting closely with the proteins on the inside of the cell membrane. When the depolarizing charge of an action potential reaches

Fig 5.6 Synaptic vesicles in the axon terminal get filled by the action of two different vesicular transporter proteins. The v-ATP-ase uses energy to pump H+ into the vesicle against it's concentration gradient (left). Then, a vesicular transporter such as vGluT use the movement of H+ down it's concentration gradient to increase intravesicular concentration of neurotransmitter (right).

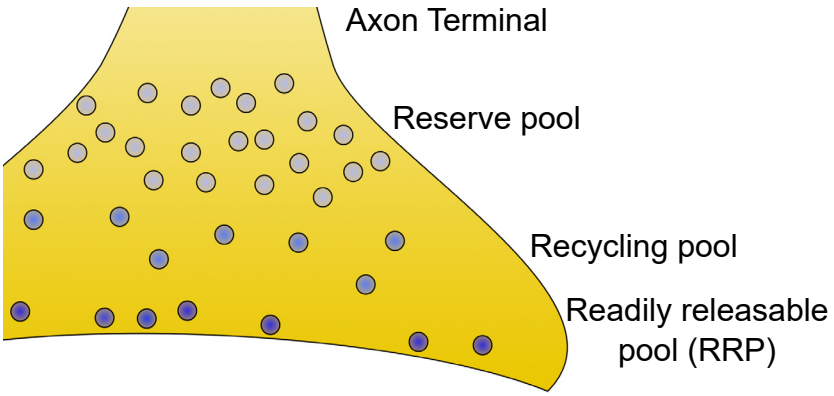


the terminal, these vesicles at the RRP are the first ones that fuse with the cell membrane and release their contents.

2. **Recycling pool.** These vesicles are the ones that have been depleted due to release. They are currently in the process of being refilled or reloaded with neurotransmitter. They are farther from the cell membrane, and the protein machinery is not primed for release, so it requires a more intense stimulation to release the contents of these vesicles.

3. **Reserve pool.** These vesicles are the farthest from the surface of the cell membrane, and most vesicles are held in this reserve pool. For these neurons to be released, very intense stimulation is required. Reserve pool vesicles may not even be recruited for release under physiological conditions.

Fig 5.7 Axon terminals contain hundreds of vesicles roughly divided into three categories.



Release of vesicles

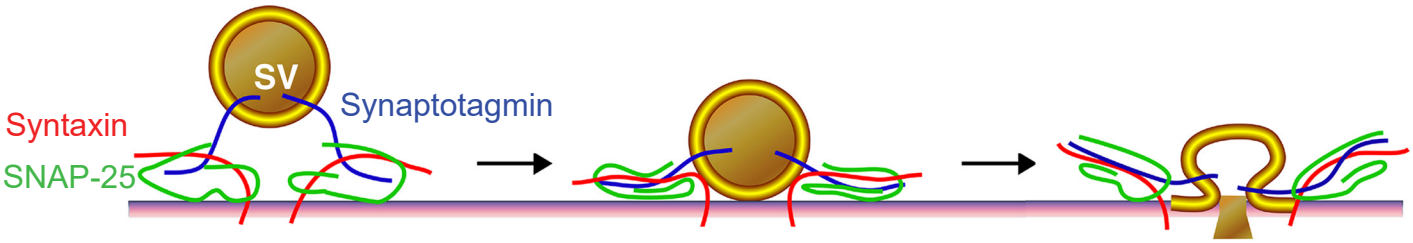
At a chemical synapse, the process of neurotransmitter release is very tightly regulated. If there were no mechanisms to control the release of chemicals at the synapse, nerve cells would deplete their entire stock of neurotransmitter. The signals that trigger muscle contraction at the NMJ would cause constant muscle tension. All sorts of brain signals would be active, and over-excitation would cause toxicity. Needless to say, regulated control of neurotransmitter release is a normal and essential part of nervous system function.

Regulation of release depends on several proteins that are important parts of the process. These proteins are often embedded within cell membranes of the vesicles or the neuronal membrane.

1. **V-SNAREs** are the proteins expressed on vesicles (v for vesicle). **Synaptobrevin** and **synaptotagmin** are two specific v-SNARE proteins that are involved during synaptic release.

2. **T-SNAREs** are proteins expressed on the cytoplasmic side of the axon terminal. The inside of the cytoplasm is the “target” for the vesicle (The t in t-SNARE). **Syntaxin** and synaptosomal nerve-associated protein 25, or **SNAP-25** for short, are t-SNAREs that function during vesicular fusion.

Fig 5.8 v-SNARE proteins interact with t-SNARE proteins to allow for vesicular fusion and release of neurotransmitter.



Clinical connection: Botulism

Botulism is a deadly condition that results from exposure to the spores produced by the bacteria Clostridium botulinum. Toxic spores can be found in the soil, contaminated foods, or water. The toxin itself is one of the most potent agents known to man - exposure to concentrations as low as 2ng/kg is lethal. The most common symptoms include muscle weakness or paralysis, especially the muscles of the face or the limbs. For about 5% of people who develop botulism, death results from paralysis due to respiratory failure.

Botulinum toxin is known to selectively cleave the proteins that comprise the SNARE complex. There are a few specific types of botulinum toxin with slightly different intracellular targets, but the result is the same on the molecular level: prevention of vesicular fusion eliminating neurotransmitter signals.

Despite being one of the deadliest toxins so far identified, millions of people pay to have a preparation of toxin called "Botox" injected into their face. For most, the injection of botox is a cosmetic procedure that can reduce the appearance of wrinkles by paralyzing the muscles. Botulinum toxin is also used medically for conditions resulting from excessive neurotransmitter release, such as muscle spasms, excessive sweating, or migraine.

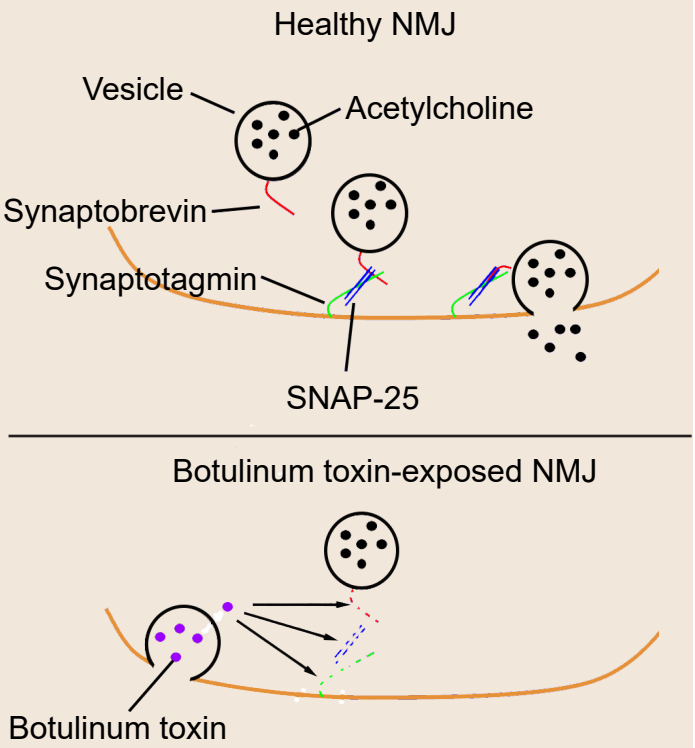


Fig 5.9 Botulinum toxin selectively cleaves vesicular fusion proteins, preventing acetylcholine from being released at the NMJ.

Fusing of vesicles

The last step of neurotransmitter release is the fusing of the cell membrane. In order to release their chemical contents into the synapse, vesicles need to fuse with the cell membrane. As the vesicular membrane merges with the interior of the neuronal membrane, the contents of the vesicle become exposed to the extracellular space. Only then are the neurotransmitters capable of activating receptors.

One of the key proteins required for vesicular fusion is the vesicle-embedded V-SNARE **synaptotagmin**. This protein is capable of detecting elevated levels of Ca^{2+} in the axon terminal. As it turns out, an elevation of Ca^{2+} in the intracellular space is the "go ahead" signal that causes neurotransmitter release.

The concentration of intracellular calcium, generally in the range of 100 nM, is much lower than the concentration outside the cell. Embedded in the cell membrane of the axon terminals are transmembrane proteins called "**voltage-gated calcium channels**" or **VGCCs**. As their name suggests, they function very similarly to the voltage-gated sodium channels described in previous chapters: they are large protein

complexes that normally remain closed, but when the surrounding neuronal membrane becomes depolarized, they physically change conformation and open up, allowing ions to move across the cell membrane. These VGCCs selectively pass only Ca^{2+} ions. The electrochemical gradient causes these Ca^{2+} ions to enter into the cell.

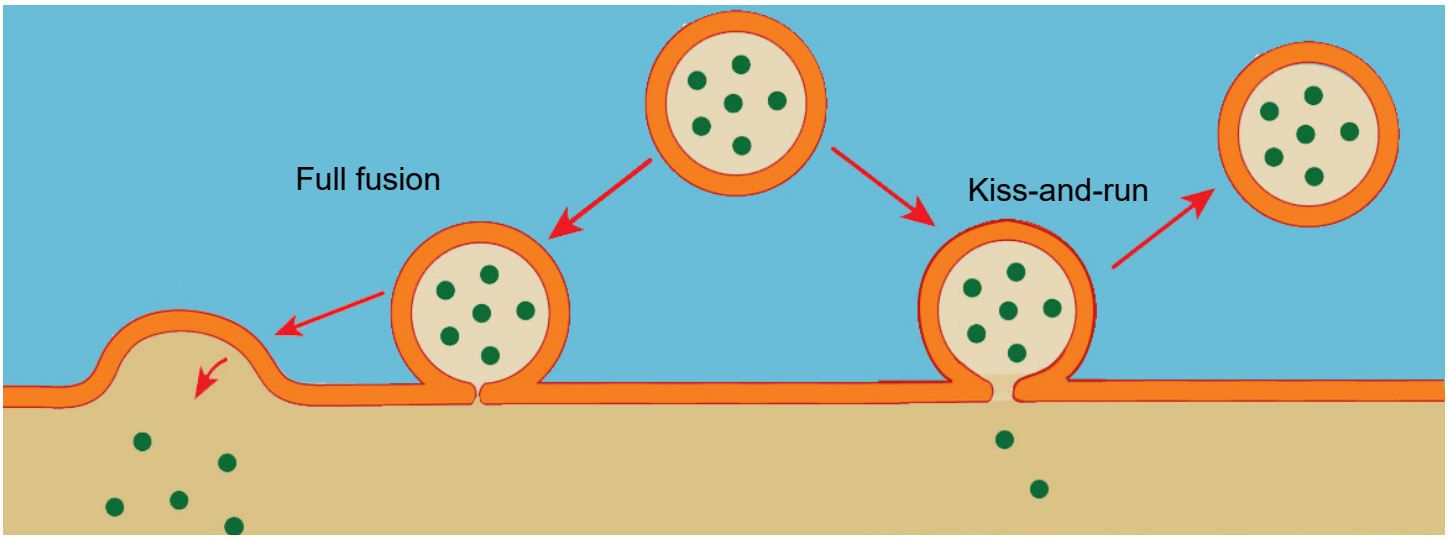
As the change in membrane potential travels down the length of the axon (the action potential), it causes a depolarization at the terminal, triggering calcium entry via the VGCCs. Ca^{2+} at the terminal binds with synaptotagmin. The v-SNAREs and the t-SNAREs interact with one another in the presence of Ca^{2+}, forming a molecular structure called a **SNARE complex**. The SNARE complex looks a lot like two twist ties that are wound tightly together. As they twist tighter together, it causes the vesicle membrane to approach the inside of the cell membrane, which results in vesicular fusion.

Vesicles are capable of fusing in at least two different ways.

1. Full fusion. A vesicle that undergoes full fusion experiences total exocytosis. The vesicular membrane becomes completely integrated with the cellular membrane, and all of the neurotransmitter spills into the synapse.

2. Kiss-and-run. This method of neurotransmitter release is incomplete fusion. The vesicle only partly connects with the interior surface of the cell membrane, and only a limited number of neurotransmitter molecules are able to enter the synapse via diffusion.

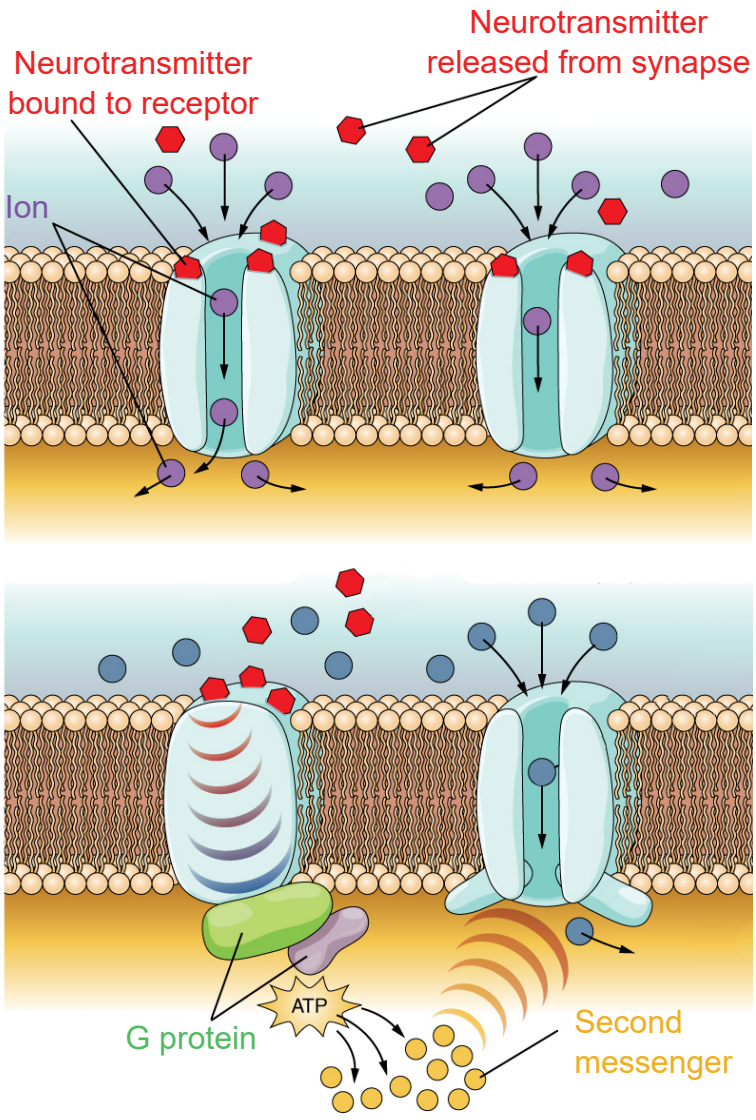
Fig 5.10 Synaptic vesicles either fuse completely (left) or partially in kiss-and-run (right).



5.3 Receptors

Receptors are proteins that are capable of sending a signal to change the function or activity of a neuron. Most receptors that function in neurotransmission are large transmembrane proteins. On the extracellular surface is a specific series of amino acid residues called the **active site**. The active site, also called the **orthosteric site**, is shaped to allow molecules of neurotransmitter to bind to the receptor.

Fig 5.11 Ionotropic receptors (top) allow ion movement after receptor binding, while metabotropic receptors (bottom) trigger second messengers to induce signaling.



Receptors are classified into one of two main categories.

1. Ionotropic receptors. Physically, ionotropic receptors are transmembrane proteins with a large-diameter pore through which ions can pass. These channels only open when a molecule of chemical binds to the active site on the extracellular side of the protein. These chemicals are also called **ligands**, and so ionotropic receptors are also called **ligand-gated ion channels**. Once a neurotransmitter activates the ionotropic receptor, ions will move based on the electrochemical gradient for that ion. As a result of ion movement, the cell's membrane potential will change. For example, nicotinic acetylcholine receptors are ligand-gated sodium channels, so when these receptors are activated by a ligand like acetylcholine, sodium ions enter into the cell causing depolarization and excitation.

Ionotropic receptors are able to induce a change in membrane potential very rapidly, on the scale of milliseconds.

Due to the nature of the amino acid residues that make up the pore of ionotropic receptors, they can be very selective for certain ions. For example, negatively charged residues lining the inside of the pore repel negatively charged Cl- ions while allowing positively charged cations to pass through the channel.

2. Metabotropic receptors. These receptor complexes cause the cell to change its metabolism in a way that leads to either excitation or inhibition. Ions do not pass through these receptors. Instead, metabotropic receptors use the actions of **G proteins**, proteins which induce changes in

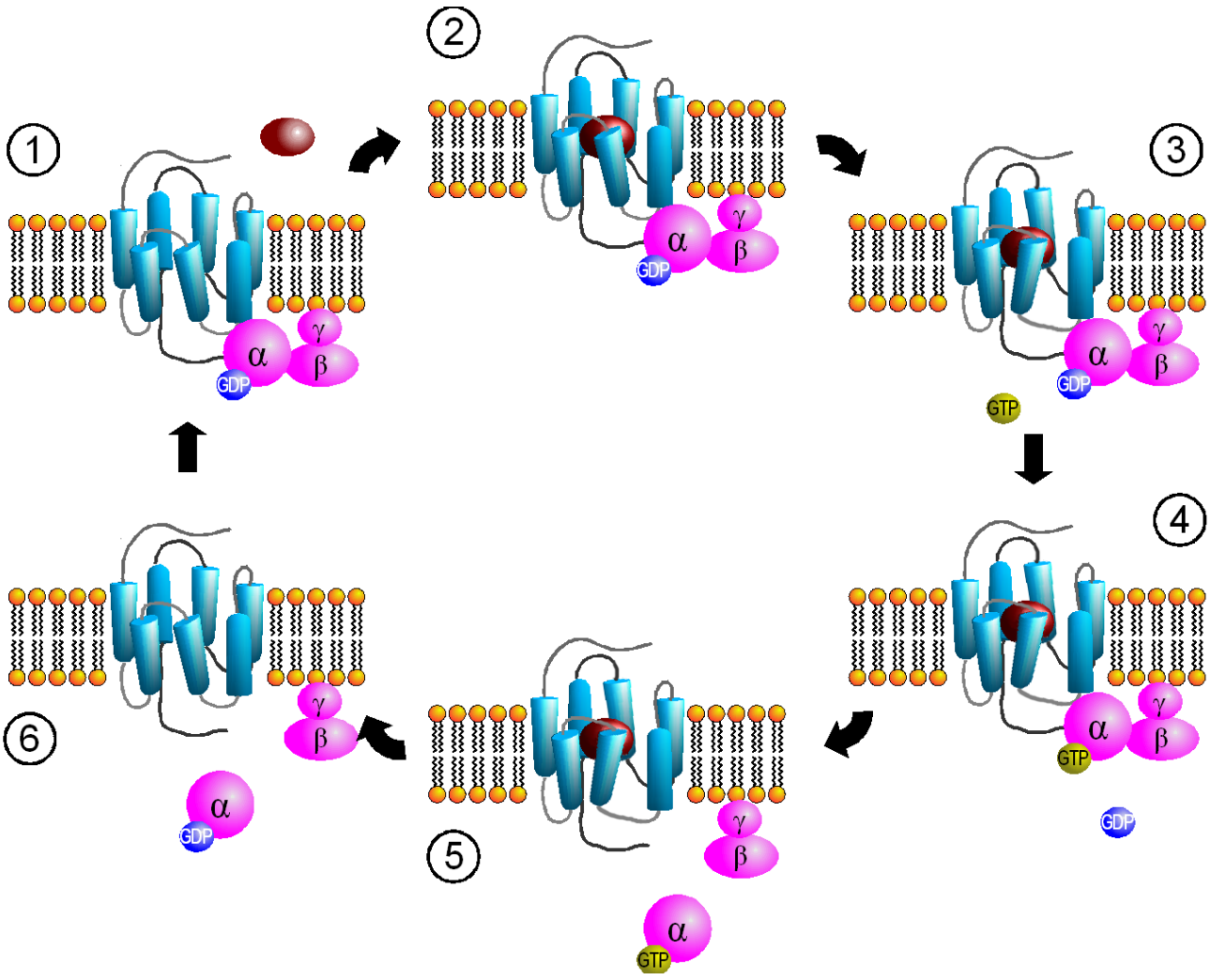
neuronal excitability through the action of second messenger signaling molecules.

Physically, metabotropic receptors are transmembrane proteins that contain 7 alpha-helix motifs that pass through the cell membrane. The N-terminus of the protein is extracellular, and the protein "weaves" back and forth across the cell membrane, resulting in a protein with three extracellular loops and three intracellular loops. Because of this shape, these receptors are also called **seven-transmembrane receptors**, or **7-TM receptors**.

Another name for these receptors is **"G protein-coupled receptors"**, or **GPCRs**. Metabotropic receptors are physically linked to proteins called G proteins, which exist on the inner surface of the cell membrane. Functionally speaking, these G proteins are capable of binding to molecules of **guanosine triphosphate (GTP)** or **guanosine diphosphate (GDP)**. Chemically similar to ATP, GTP can function as a source of energy. G proteins themselves exhibit catalytic activity of GTP. This means that they are capable of breaking down GTP into the less-energetic GDP. When GTP is bound to the GPCR, the receptor is active. When this molecule is hydrolyzed into GDP, the receptor becomes inactive.

Some G proteins are heterotrimeric, meaning that they are made up of three different subunits, alpha, beta, and gamma. The GTP binding sites and catalytic sites are found on the alpha subunit, the largest of the three subunits.

Fig 5.12 GPCRs signal via activation of the G protein attached to the intracellular side of the receptor.

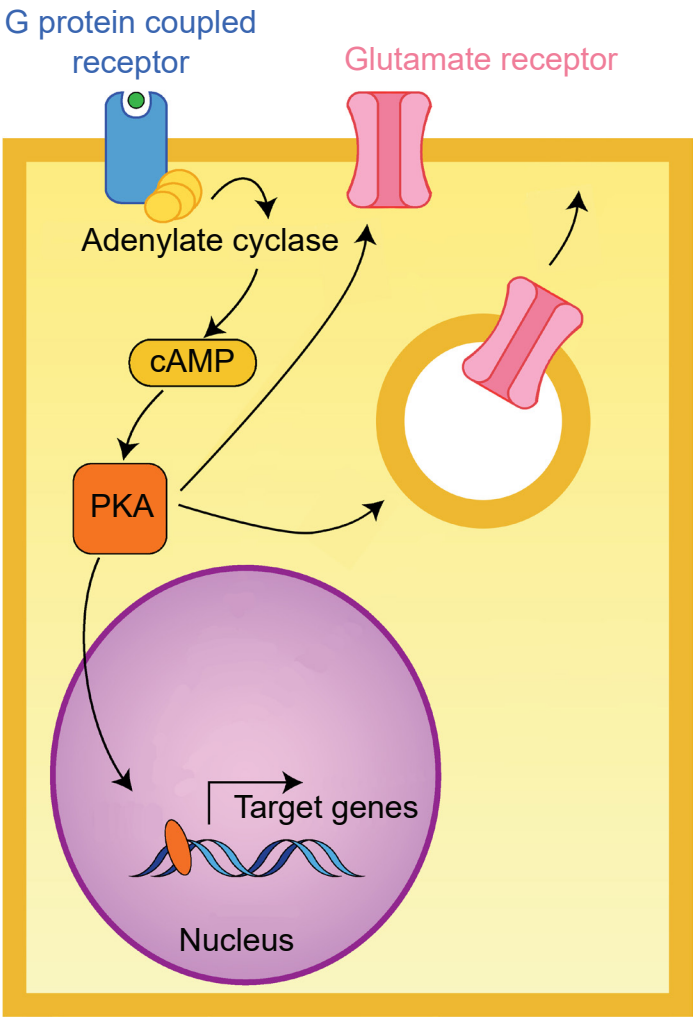


Usually, the alpha subunit becomes soluble after activation, while the beta / gamma complex stays embedded in the neuronal membrane. The G alpha subunit exists in different varieties.

$G_{\alpha s}$. When a neurotransmitter activates a GPCR coupled with the $G_{\alpha s}$ protein, the $G_{\alpha s}$ protein is excitatory (The "s" stands for stimulatory).

Binding of a ligand to the active site of Galphas-coupled GPCRs results in increased activity of the enzyme **adenylate cyclase (AC)**. AC itself is an enzyme that creates a second messenger, a molecule called **cyclic AMP (cAMP)**. Elevated levels of cAMP activate an enzyme called **protein kinase A (PKA)**.

Fig 5.13 GPCRs that are coupled with $G_{\alpha s}$ are excitatory through adenylate cyclase signaling.



PKA is a kinase, an enzyme that is phosphorylates other proteins. The addition of a phosphate group onto a protein changes its properties dramatically. PKA phosphorylates protein targets that increase cell excitation. For example, one target of PKA activity is the intracellular side of certain glutamate receptors. Phosphorylation causes these receptors to stay open longer than normal when they are activated by a molecule of glutamate. This means that a single molecule of glutamate causes more excitation (passes more depolarizing current into the cell) in the presence of increased PKA activity.

Targets of PKA activity also include the intracellular store of glutamate receptors. When phosphorylated, these receptors are trafficked to the neuronal membrane. An increase of receptors at the postsynaptic side leads to increased excitatory neurotransmission over a period of minutes and hours, representing one form of plasticity.

An even longer-term action of PKA is its ability to change the transcription of various genes, which can trigger the synthesis of proteins. Some genes downstream of PKA activity include the structural protein actin, important for morphological changes in neuronal structure, or ion channels which change neuronal responses to neurotransmitter release.

$G_{\alpha i}$. A GPCR that is coupled with $G_{\alpha i}$ causes a decrease in excitability. In many ways, $G_{\alpha i}$ proteins serve the opposite function as $G_{\alpha s}$ proteins - the "i" stands for inhibitory.

Whereas activation of $G_{\alpha s}$ increases the action of AC, $G_{\alpha i}$ proteins decrease AC activity. Therefore, $G_{\alpha i}$ activation decreases the intracellular concentration of cAMP, in turn decreasing PKA activity. Given the function of PKA as a kinase that increases cellular excitation as

described above, decreased PKA activity inhibits cellular activity through multiple mechanisms, some of which include decreased current through glutamate receptors, decreased trafficking of glutamate receptors to the presynaptic neuronal membrane, and decreased transcription of certain genes.

$G_{\alpha q}$. Generally, $G_{\alpha q}$ is an excitatory G protein. $G_{\alpha q}$ uses a different signaling pathway compared to the PKA pathway that is downstream of $G_{\alpha s}$ or $G_{\alpha i}$. $G_{\alpha q}$ protein activation leads to activity of the **enzyme phospholipase C (PLC)**.

On a biomolecular level, PLC acts on the phospholipid membrane molecule phosphatidylinositol 4,5-bisphosphate (PIP2). PLC is a hydrolytic enzyme, and it breaks PIP2 into two separate second messenger molecules: the soluble inositol triphosphate (IP3) and the membrane-embedded diacylglycerol (DAG). One of the functions of IP3 is to liberate Ca2+ from intracellular stores, which elevates intracellular Ca2+ levels, depolarizing the cell and activating calcium-dependent processes, which are often excitatory. DAG activates protein kinase C (PKC), an enzyme with substrates that increase neurotransmitter release probability or decrease potassium channel conductance.

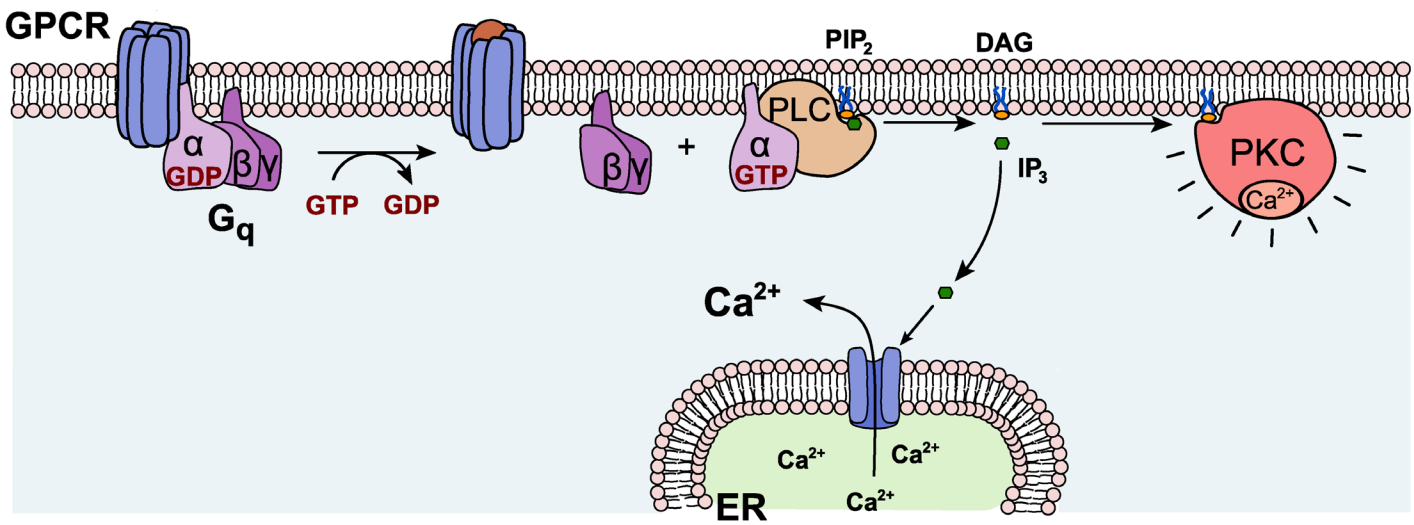
While the alpha subunits carry out a large part of GPCR-mediated changes in cellular excitation, the beta and gamma subunits also affect excitability. The beta and gamma subunits are bound together as a dimer, but they separate from the alpha subunit once the GPCR becomes activated. The beta-gamma complex can also function as a signaling molecule.

Compared to ionotropic receptors, metabotropic receptors affect neuron activity on a slower time scale, on the range of milliseconds to seconds, and possibly even longer depending on the downstream mechanisms that are activated.

Presynaptic receptors

In the discussion of receptors, it is common to think of receptors as being expressed at the dendrites, embedded within the postsynaptic membrane. However, not all receptors are located here. Some receptors are presynaptic, meaning they can be found at the axon terminal. Presynaptic receptors are often inhibitory and serve a self-regulatory function. Presynaptically-expressed receptors that respond to the same neurotransmitter that is released are called **autoreceptors**.

Fig 5.14 $G_{\alpha q}$ signals using PLC, which then produces two signaling molecules, IP3 and DAG.



5.4 Neurotransmitters

As described previously, neurotransmitters are the substances that are released at chemical synapses, and they are the signaling molecules that allow neurons to communicate with one another. To date, scientists have identified more than 100 neurotransmitters. Here, we will describe six classical neurotransmitters, their receptors, and their actions. Additionally, three atypical neurotransmitters will be introduced.

One important note to keep in mind as you think about neurotransmitters: the effect that a neurotransmitter has on a cell depends on the receptor. In other words, a neurotransmitter molecule can either excite or inhibit a neuron depending on the composition of receptors that are present. For example, glutamate is excitatory at most synapses in the nervous system. Glutamate exerts excitation by activating ionotropic glutamate receptors, which are ligand-gated cation channels. However, at one particular synapse in the eye, glutamate activates a metabotropic glutamate receptor that causes cellular inhibition.

Glutamate

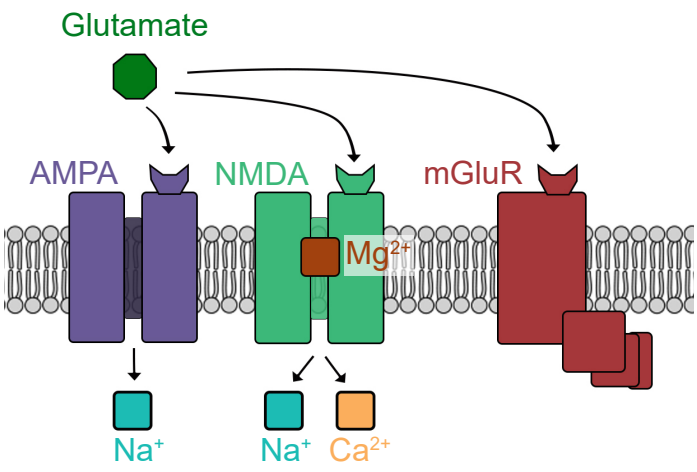
Glutamate (Glu) is the main excitatory neurotransmitter used by the nervous system. Glutamate is the same as the amino acid glutamic acid. There is more glutamate per volume of brain tissue than any other neurotransmitter. Glutamatergic neurons are identified by the presence of the **vesicular glutamate transporter (vGluT)**.

Glutamate can activate both ionotropic and metabotropic receptors. Ionotropic glutamate receptors are all ligand-gated cation channels, which makes them excitatory since they allow Na^+ to move into the cell. Ionotropic glutamate receptors are generally subdivided into three classes, named after exogenous chemicals that can activate the receptor. AMPA receptors are Na^+ channels, but some also allow Ca^{2+} entry. NMDA receptors allow both Na^+ and Ca^{2+} to pass across the membrane. When the cell is at rest, NMDA receptors also have a large magnesium ion in the pore that blocks ion movement through the channel. The third category of ionotropic glutamate receptors is called kainate receptors, which are similar to AMPA receptors.

The metabotropic glutamate receptors (mGluRs) signal using different G proteins. There are a total of 8 of these mGluRs, classified into three groups, called Group I, Group II, and Group III. Group I are excitatory GPCRs which signal via Gq, while Group II and Group III are inhibitory via the Gi signal transduction pathway.

One theory proposes that excess signaling by glutamate can lead to neuronal death, a phenomenon called **excitotoxicity**. Of the

Fig 5.15 Glutamate is the main excitatory neurotransmitter in the nervous system, acting at different categories of receptors, three of which are shown below.

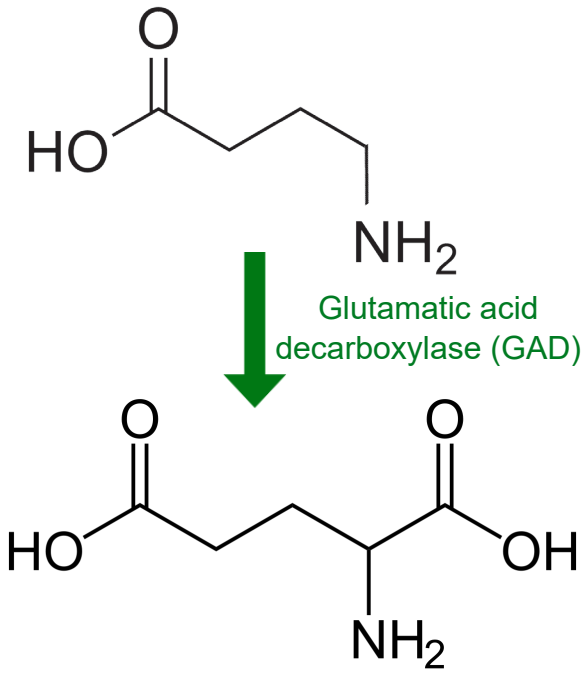


various glutamate receptors, the NMDA receptor is most strongly implicated in contributing to excitotoxicity, since uncontrolled elevated levels of calcium can be deadly for neurons. Excitotoxicity is observed in a variety of disease states ranging from neurodegenerative disorders such as Parkinson's disease, Alzheimer's disease, and multiple sclerosis, but also in injury such as concussion or stroke.

GABA + glycine

Gamma-aminobutyric acid (GABA) is the main inhibitory neurotransmitter in the brain. According to one estimate, about 25% of neurons in the brain are GABA-ergic. Chemically speaking, GABA is remarkably similar to glutamate. In fact, GABA is synthesized from glutamate in a single step by the enzyme **glutamic acid decarboxylase (GAD)**. GAD is often used as a biochemical marker for the presence of GABA-ergic neurons. Many interneurons use GABA as their chemical signaling molecule.

Fig 5.16 The inhibitory neurotransmitter GABA is synthesized from glutamate by the action of GAD.



The main action of GABA as an inhibitory neurotransmitter is to activate one of three main classes of receptors, called A, B, and C. $GABA_A$ receptors are ligand-gated chloride channels, so activating these ion channels causes Cl^- flux, which opposes the ability for the cell to reach action potential threshold. $GABA_B$ and $GABA_C$ receptors are both metabotropic receptors that inhibit neuronal activity through the action of the G_i protein.

A neurotransmitter that is similar to GABA is **glycine**. Another small amino acid, glycine is mostly used by neurons of the spinal cord and brain stem. Glycine is also inhibitory , and acts at glycine receptors, which are ligand-gated chloride channels.

Dopamine

Dopamine (DA) is a biogenic amine derived from the amino acid tyrosine through the action of several enzymes. One in particular, **tyrosine hydroxylase (TH)**, is the main marker that is used for identifying dopamine-producing neurons. Unlike glutamate or GABA, dopamine-producing neurons are not widely abundant in the brain. Instead, there are generally only a few patches of neurons that produce dopamine, most of which are found in the midbrain. Two areas include the **ventral tegmental area** and the **substantia nigra**.

There are five classes of dopamine receptors, named D1 through D5. All of them are metabotropic receptors. D1 and D5 are generally excitatory receptors, while D2, D3, and D4 are inhibitory receptors.

Since Roy Wise's theory proposed in the 1960s, DA has been known in pop culture as the "pleasure neurotransmitter" because of its involvement in the processing of reward and motivation. For example, if we use microdialysis

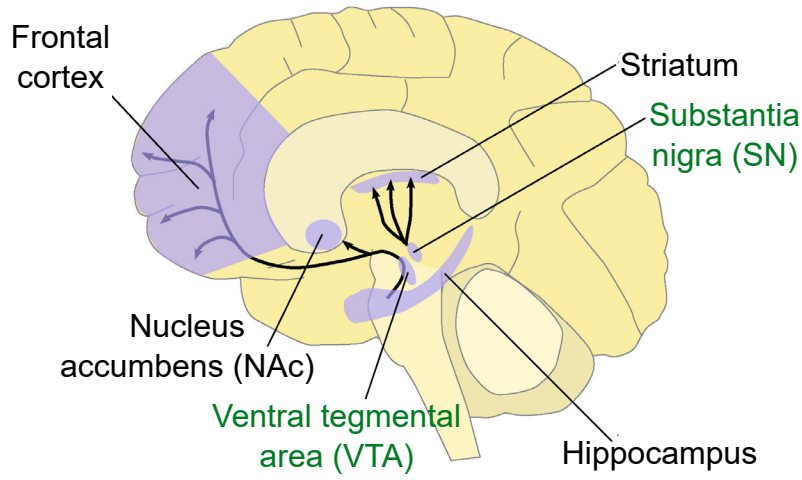
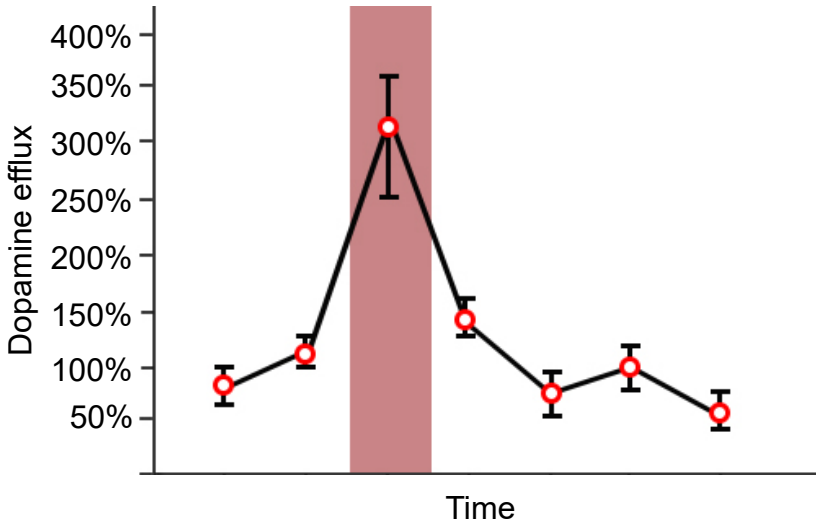


Fig 5.17 Brain dopamine is synthesized in two major midbrain nuclei, the VTA and SN, labeled in green.

(a technique to measure the concentration of chemicals) in the nucleus accumbens, dopamine levels spike in response to all sorts of pleasurable or rewarding stimuli: food, water, sex, sugar, and exposure to drugs of abuse. However, we now know that dopamine is much more complex than once believed. One theory suggests that dopamine elevation serves as a "learning signal" that causes us to pay attention to salient stimuli in the environment.

DA is also needed for normal motor control. When dopamine-producing neurons in the substantia nigra pars compacta (SNpc) degenerate, as in Parkinson's disease, a person develops the trademark symptoms: difficulties with motor control, resulting in a resting tremor, postural instability, and bradykinesia (slowness of movement). Reversing the dopamine deficiency by introducing an exogenous source of dopamine is our current gold standard of treatment for PD.

Fig 5.18 Levels of dopamine in the NAc rise when animals lever press for delicious foods, denoted in red.



Serotonin

Serotonin (5-HT) is a neurotransmitter that is derived from the dietary amino acid tryptophan. The enzyme **tryptophan hydroxylase** is the first step of serotonin biosynthesis and is often used as a marker to identify serotonergic neurons. As with dopamine, there are only a few areas of the brain that synthesize serotonin, the major one being the **Raphe nucleus** in the brain stem.

Receptors for serotonin have a wide variety of actions. We have identified seven major families of 5-HT receptors, which are designated by the number and subclasses which are designated by a letter. For example, the 5-HT2A receptor is metabotropic and excitatory via Gq signaling, while the 5-HT5 receptor is inhibitory via Gi signaling. Most of them are metabotropic receptors; only the 5-HT3 receptor is ionotropic.

Serotonin is heavily implicated in the regulation of mood and complex behavioral conditions. One of our most effective strategies for treating depression is the administration of a drug such as fluoxetine, which acts as a **selective-serotonin reuptake inhibitor (SSRI)**. Pharmacologically, fluoxetine increases

Clinical correlation: Parkinson's disease (PD) and L-DOPA-induced dyskinesia (LID)

Parkinson's disease is a debilitating neurodegenerative disorder that affects as many as 1% of all people aged 60 or older. Generally, PD is lethal within 16 years. By the time a patient presents to the clinic with motor dysfunction, they have already lost almost 60-80% of dopamine-producing neurons in this area!

For decades, clinicians have been using the biochemical precursor to dopamine, L-DOPA, to treat the symptoms. However, after chronic exposure to L-DOPA, the drug becomes less effective and has a shorter duration of therapeutic action. Worse still, frequent treatment can lead patients to develop **hyperkinesias**, an abnormal excess of movements. This iatrogenic disorder is called **L-DOPA induced dyskinesia (LID)**.

Biomedical engineers have developed a promising non-drug approach to treating PD called **deep brain stimulation**. A small stimulating device is surgically implanted into the subthalamic nucleus of the brain. When this brain area is stimulated, neural circuits are recruited which restores normal motor control.

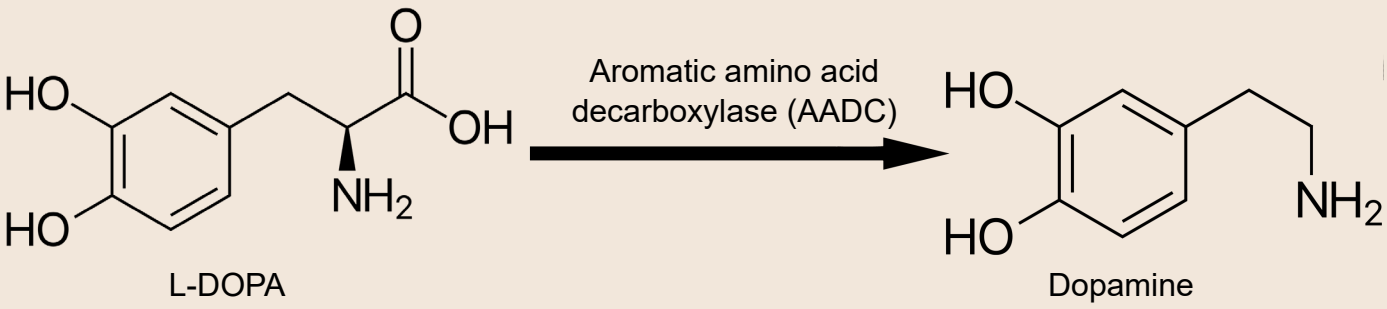


Fig 5.19 Patients with PD have characteristic changes in gait as a result of low dopamine (top). The current best pharmacological therapy is levodopa administration, which is the biochemical precursor to dopamine (bottom).

synaptic levels of serotonin by preventing reuptake, and for some people, this has a moderate ability to reverse depression. Serotonin signaling is also a target for drugs that treat anxiety, post-traumatic stress disorder, obsessive-compulsive disorder, schizophrenia, and more.

Acetylcholine

Acetylcholine (ACh) is a small molecule that is made by the enzyme **choline acetyltransferase (ChAT)**, which chemically bonds a molecule of acetyl-CoA with a molecule of choline. The presence of ChAT in a neuron is used as a biochemical marker for neurons that produce acetylcholine.

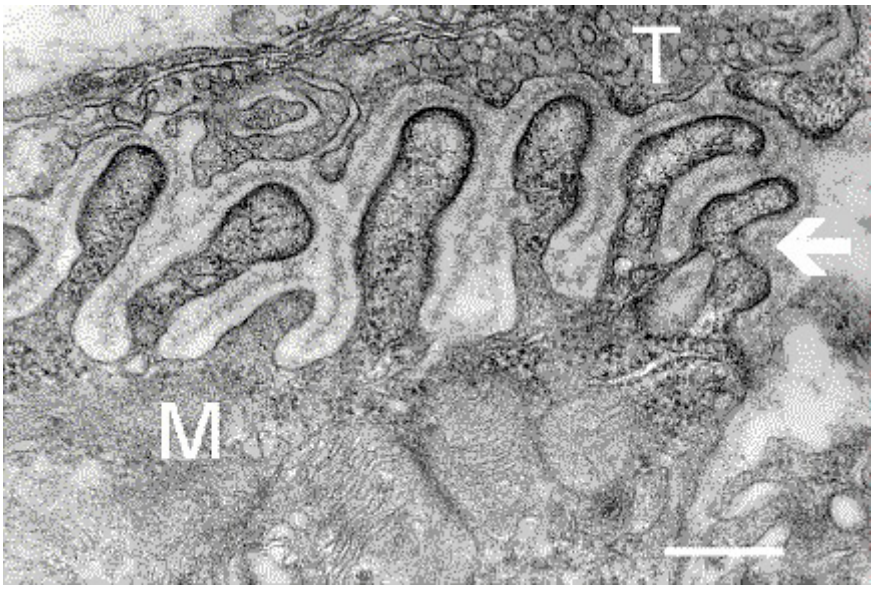
ACh was the first neurotransmitter discovered and chemically isolated, a feat which earned two researchers the shared Nobel Prize in Physiology or Medicine in 1936. One of the two scientists, a German pharmacologist named Otto Loewi, stimulated the vagus nerve connected to an isolated frog heart, which caused the heart rate to slow down. When he put the surrounding solution on top of another heart, he observed that the second heart also slowed down, despite having no physical connection to the first heart. From this, he concluded that a chemical released by the vagus nerve is able to decrease heart rate. This chemical was first called *Vagusstoff*, the German word meaning Vagus substance. Today, we know it as acetylcholine.

ACh is able to act at ionotropic and metabotropic receptors, and activity at both receptor classes is essential for normal function. The ionotropic receptors of the nervous system are called **nicotinic acetylcholine receptors (nAChRs)** because they can be activated by nicotine in addition to acetylcholine. These ionotropic receptors are ligand-gated sodium channels and are therefore excitatory. On the other hand, the metabotropic receptors are called **muscarinic acetylcholine receptors (mAChRs)** since they are activated by the chemical muscarine found in some species of mushrooms. MAChRs can be coupled with either G_s or G_i, so they can be either excitatory or inhibitory.

ACh is the main neurotransmitter that the nervous system uses in order to communicate with the muscles at the neuromuscular junction (NMJ). Here, ACh is released by motor neurons, where it activates nicotinic acetylcholine receptors on muscle cells, causing them to constrict, or flex. On the other hand, muscarinic acetylcholine receptors are located in the heart, and their activation causes a decrease in heart rate (as Otto Loewi demonstrated with the isolated frog heart preparation.)

In the central nervous system, ACh is involved in a wide variety of processes, including attention and learning. One of the first theories to explain the symptoms of Alzheimer's disease looked at a loss of ACh-producing neurons in the basal forebrain that become more severe as the disease worsens. It has since been demonstrated that Alzheimer's disease is more complex than this early hypothesis.

Fig 5.20 An electron microscope image of the neuromuscular junction showing vesicles (T, top) forming a synapse with the muscle cell (M, bottom). Acetylcholine is the main neurotransmitter used in muscle control at the PNS.

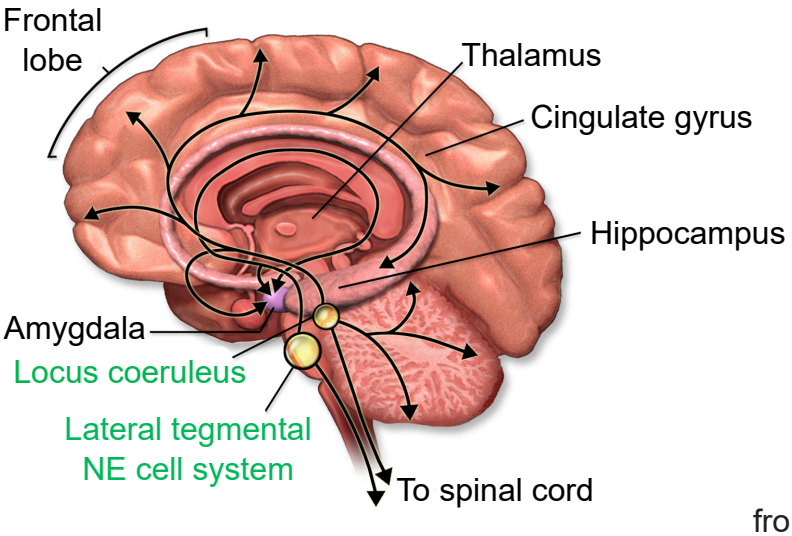


Norepinephrine

Norepinephrine (NE) is a neurotransmitter that is synthesized from a molecule of dopamine by the enzyme **dopamine beta-hydroxylase**. Norepinephrine-producing cells are localized in the pons of the brain stem, a structure

called the locus coeruleus. The locus coeruleus is very small, but these neurons send projections widely throughout the brain.

Fig 5.21 Norepinephrine in the brain is synthesized by small populations, but these cell bodies project widely across several other areas.



Outside of the brain, we think of norepinephrine as being responsible for triggering the sympathetic nervous system response of the body, the "fight-or-flight" reaction that prepares the body for physical activity in times of fear or acute stress. These norepinephrine-producing nerve cells reside in the sympathetic ganglia, a clump of nerve cells that run parallel to the spinal cord, one on each half of the body. These neurons project out towards the internal organs.

Receptors for NE are classified into two main categories, alpha or beta. There are subtypes within each category, giving us five major receptors for NE: alpha-1, alpha-2, beta-1, beta-2, and beta-3. All five of these receptors are metabotropic, and some are excitatory while others are inhibitory. Our internal organs express these noradrenergic receptors. Clinically, the "beta blockers" are a class of drugs that inhibit beta-adrenergic receptors; the resulting action is a decrease in blood pressure. Conversely, some beta-agonists are used as bronchodilators for asthma.

Norepinephrine also functions in the brain to modulate behaviors including alertness and attention.

Atypical neurotransmitters

Although we generally think of neurotransmitters as neurochemicals that function as described above, there are a few atypical neurotransmitters that don't quite fit the mold of the other chemical signals.

Neuropeptides

Neuropeptides are a class of large signaling molecules that some neurons synthesize. Neuropeptides are different from the traditional neurotransmitters because of their chemical size. Monoamines like DA, NE, or 5-HT have a molecular weight around 150-200, while one of the smaller neuropeptides, enkephalin, has a molecular weight of 570. One of the largest, dynorphin, has a molecular weight greater than 2,000. Because of their large size, neuropeptides have to be packaged in dense-core vesicles very close to the site of production near the nucleus rather than in clear vesicles right at the terminal.

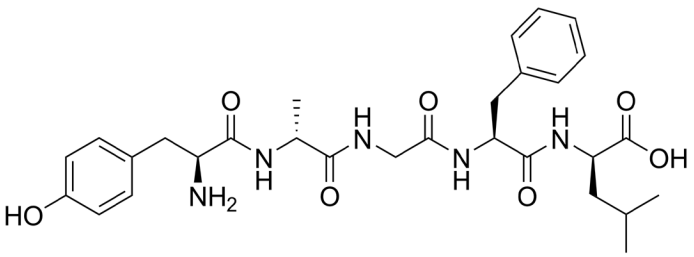


Fig 5.22 Enkephalin, one of the smaller neuropeptides, is very large compared to other neurotransmitters. Enkephalin is an agonist for both δ and μ opioid receptors.

Neuropeptides such as enkephalin and dynorphin are agonists at a class of receptors called the **opioid receptors**. These opioid receptors fall into four main types. The three classical opioid receptors are named using Greek letters: δ (delta), μ (mu), and κ (kappa), and the fourth class is the nociceptin receptor. All of these receptors are inhibitory metabotropic receptors which signal using the $G_{\alpha i}$ protein.

These receptors are expressed in several brain areas, but expression is particularly heavy in the periaqueductal gray, a midbrain area that functions to inhibit the sensation of pain. Drugs that activate the opioid receptors, like morphine, oxycontin, or fentanyl, are the most effective clinical treatments that we know of for acute pain. Unfortunately, these same drugs also represent a tremendous health risk, as opioid drugs can be lethal in overdose and have a high risk of substance use disorder.

Endocannabinoids (eCBs)

The eCBs are a class of lipid-based neurotransmitters. They are unusual neurochemicals in a few ways. Instead of sending information from the axon of one neuron to the dendrite of the next neuron (anterograde signaling), eCBs allow the postsynaptic dendritic component to communicate with the presynaptic axon terminal. Since they communicate information in the "opposite" direction of classic neurotransmitter signaling, eCBs are called retrograde signaling molecules. Secondly, eCBs are not packaged into vesicles and released by fusion processes. Instead, eCBs are synthesized de novo, meaning they get created right when they needed and used at that moment. The two most well-characterized eCBs in humans are called 2-AG and AEA.

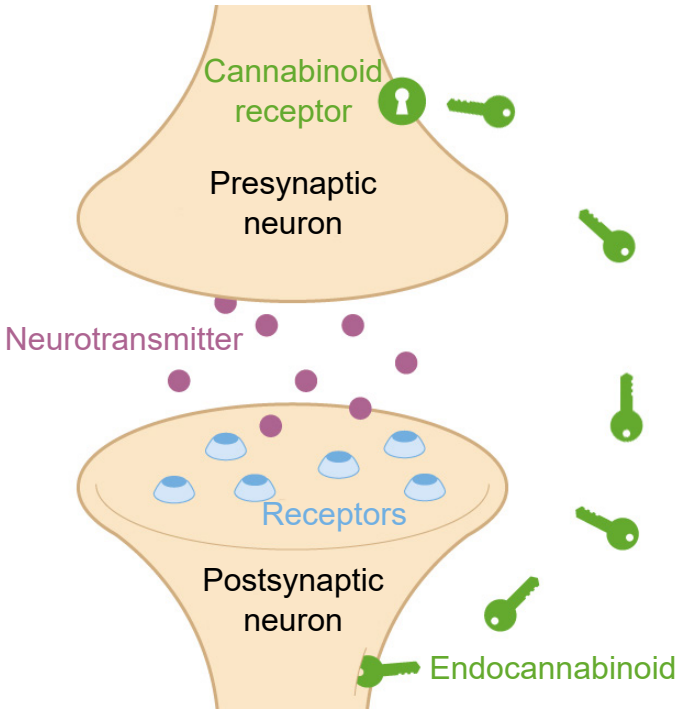


Fig 5.23 ECBs are synthesized from the postsynaptic cell membrane and signal through presynaptic cannabinoid receptors

ECBs activate one of two receptors, CB1 and CB2. Both of them are inhibitory metabotropic receptors that couple with Galphai. Generally, CB1 receptors are found in the nervous system, while the CB2 receptors are found elsewhere in the body, such as in the immune system.

The eCB system is widely used by various systems in the body. It is estimated that eCB receptors are the most abundant GPCRs in the whole body.

These substances were named because they are endogenous chemicals that are functionally similar to compounds found in plants of the genus *Cannabis*. One reason cannabis is used is because of its ability to interact with our eCB receptors.

Nitric oxide

The nervous system is capable of signaling via the gas nitric oxide (NO). This gasotransmitter

is not stored in vesicles but rather is synthesized as it is needed. NO is formed when the amino acid arginine is degraded by the enzyme NO synthase (NOS).

Because NO is a gas, it easily permeates across cell membranes. Therefore, the receptors for NO do not need to be transmembrane proteins expressed on the cell surface. Instead, the receptor for NO is an intracellular receptor called **soluble guanylate cyclase (sGC)**. SGC works through a signaling pathway that is different from other metabotropic receptors so far described. SGC is linked with the signaling molecule cyclic GMP (cGMP), which activates protein kinase G. PKG can either be excitatory or inhibitory, depending on the intracellular components.

Chapter summary

Neurons communicate with one another in a variety of ways. Anatomically, neurons are separated by a small extracellular gap called the synapse. This synapse may directly connect the intracellular cytoplasm, as in an electrical synapse. Alternatively, the gap may be much larger, and chemicals that get released and diffuse across the synapse in order to signal to the following neuron. These chemicals, the neurotransmitters, are stored in vesicles, tiny spheres that are located in the presynaptic axon terminal. The release of these chemicals is very closely regulated, and neurons have several mechanisms that regulate vesicular fusion.

Following release, the neurotransmitters diffuse across the synapse and can bind to the active site on transmembrane proteins called receptors. Upon binding a molecule of neurotransmitter, these receptors physically change shape, resulting in ion flow across the membrane (ionotropic receptors) or a change in the activity of intracellular signaling molecules (metabotropic receptors). Binding of neurotransmitter changes the excitability of neurons.

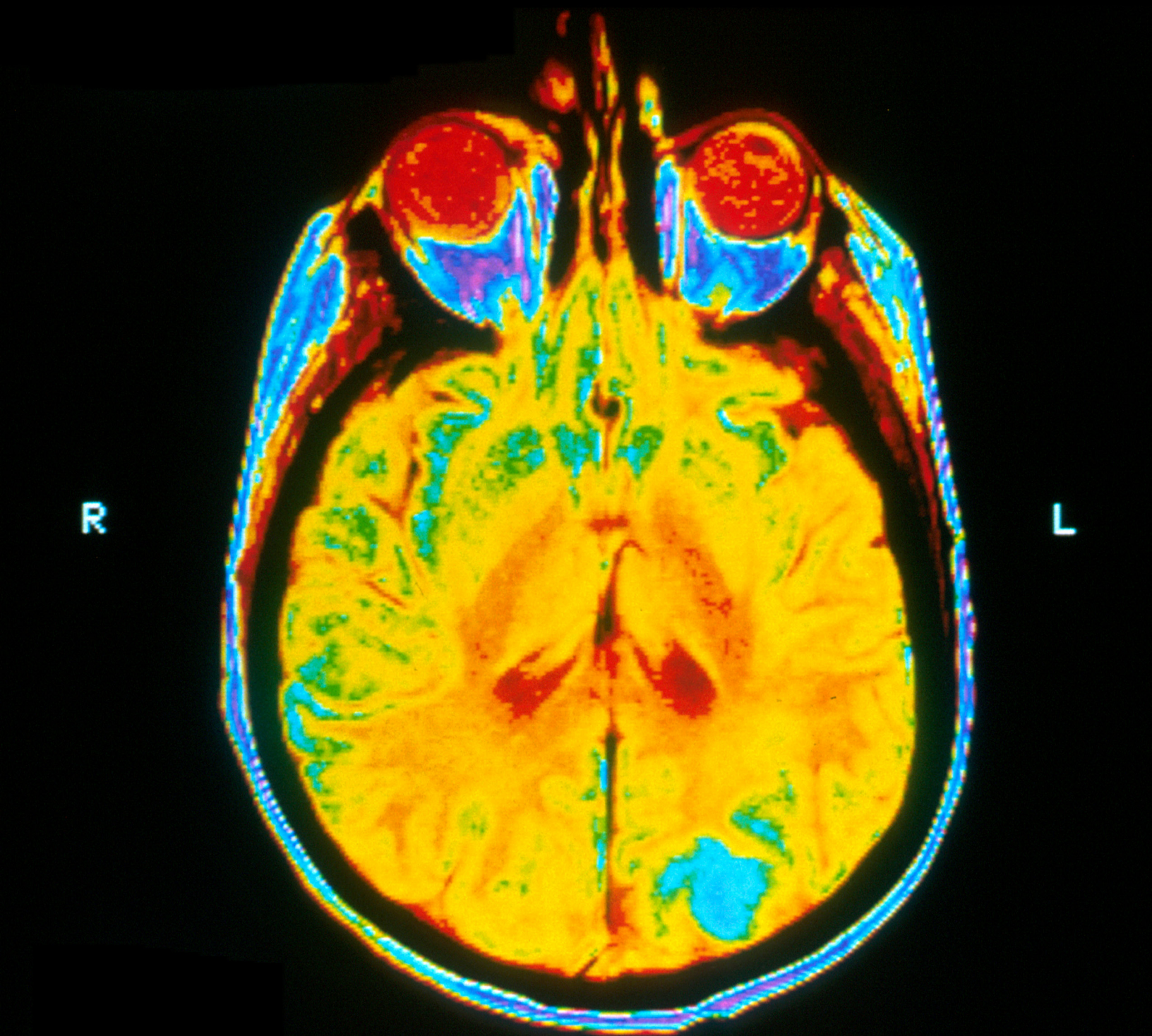
We have so far identified more than 100 neurotransmitters. Many of them are small molecules that are packaged in vesicles, which then diffuse from the presynaptic neuron to the postsynaptic neuron, such as acetylcholine, glutamate, or GABA. However, there are some atypical neurotransmitters such as neuropeptides, endocannabinoids, and nitric oxide that have different methods of communication.

The Open Neuroscience Initiative is funded by a grant from the Vincentian Endowment Fund of DePaul University.

Cover: Image found at https://www.hippopx.com/en/people-letters-envelope-pen-dip-pen-writings-ink-62437
5.1 https://upload.wikimedia.org/wikipedia/commons/7/78/Gap_cell_junction_keys.svg modified by Austin Lim
5.2 https://upload.wikimedia.org/wikipedia/commons/e/e4/A_skybridge_at_IUPUI_campus.jpg
5.3 https://upload.wikimedia.org/wikipedia/commons/e/ed/Charcot-marie-tooth_foot.jpg
5.4 (Left) https://upload.wikimedia.org/wikipedia/commons/4/4c/Synapse_diag1.svg modified by Austin Lim
(Right) https://commons.wikimedia.org/wiki/File:Synapse_diag4.png#file modified by Austin Lim
5.5 Tao C-L, Liu Y-T, Zhou ZH, Lau P-M and Bi G-Q (2018) Accumulation of Dense Core Vesicles in Hippocampal Synapses Following Chronic Inactivity. Front. Neuroanat. 12:48. doi: 10.3389/fnana.2018.00048
5.8 https://upload.wikimedia.org/wikipedia/commons/2/28/Opening_of_a_Fusion_Pore_during_Exocytosis.png modified by Austin Lim
5.9 https://upload.wikimedia.org/wikipedia/commons/1/1e/Botulinum_Toxin_Mechanism.png modified by Austin Lim
5.10 https://upload.wikimedia.org/wikipedia/commons/0/0c/0310_Exocytosis_cleaned.png modified by Austin Lim
5.11 https://upload.wikimedia.org/wikipedia/commons/e/e7/1226_Receptor_Types.jpg modified by Austin Lim
5.12 https://upload.wikimedia.org/wikipedia/commons/6/6d/GPCR-Zyklus.png
5.13 https://upload.wikimedia.org/wikipedia/commons/8/8f/CREB_cAMP_neuron_pathway.svg modified by Austin Lim
5.14 https://upload.wikimedia.org/wikipedia/commons/3/31/Activation_protein_kinase_C.svg modified by Austin Lim
5.17 https://upload.wikimedia.org/wikipedia/commons/d/de/Dopamine_pathways.svg modified by Austin Lim
5.18 Data from K.N. Segovia, M. Correa, J.D. Salamone, Slow phasic changes in nucleus accumbens dopamine release during fixed ratio acquisition: a microdialysis study, Neuroscience, Volume 196, 2011, Pages 178-188, ISSN 0306-4522, https://doi.org/10.1016/j.neuroscience.2011.07.078. Reprinted with permission 2/15/2020
5.19 https://upload.wikimedia.org/wikipedia/commons/4/44/Sir_William_Richard_Gowers_Parkinson_Disease_sketch_1886.svg
5.20 https://upload.wikimedia.org/wikipedia/commons/a/aa/Electron_micrograph_of_neuromuscular_junction_%28cross-section%29.jpg
5.21 https://upload.wikimedia.org/wikipedia/commons/4/48/Norepinephrine_Part_1.png modified by Austin Lim
5.23 https://upload.wikimedia.org/wikipedia/commons/2/27/Sistema_endocannabinoide.png modified by Austin Lim

Chapter 6:

Methods of Neuroscience



Austin Lim, PhD (DePaul University)
Claire Sun (Mount Sinai)
Editor: Katie L. Willis, Ph.D. (University of Oklahoma)

Most of the information you learn in neuroscience textbooks were discovered by people who applied the scientific method to systematically answer a research question. In order to come to their conclusions, these neuroscientists used a variety of techniques, borrowed from biology, medicine, chemistry, physics, engineering, psychology, and many other academic disciplines. Neuroscience researchers use the wide range of scientific methodological strategies to answer: "How does the nervous system work?"

In this chapter, we will describe a few of the methods that are used to ask and answer questions in neuroscience. These methods can be divided roughly into four major categories: Imaging anatomy, imaging function, imaging cells, and finally manipulating the nervous system. Within each of these categories, the techniques will start with the "big picture" view, and by the end of each section we zoom down to the level of molecules and genes.

Just a minor caveat about these techniques. Many of these strategies can be used independently, but some may be used simultaneously.

Figure 6.1 Being able to zoom in allows us to see small details, but information from the big picture is missing.

There are three major components that can be described for some of the techniques:

1. A description of how the method works. Each of these methodological strategies rely on using the knowledge from other fields of study and applying that knowledge to a question about the nervous system. Knowing how the methods work allow us to decide what types of data can be collected and when each method will be appropriate to use.

2. Some of the main advantages of the technique. Not every method is applicable for studying every question. Instead, it would be beneficial for neuroscientists to choose the most appropriate strategy for a particular research question given the circumstances.

3. Some shortcomings or limitations of the technique. As above, no strategy is perfect. Methods that look at the "big picture" aren't able to look at microscopic level detail. There are limitations for every technique and being aware of these limitations inform us about caveats in interpreting data.

Although almost all these techniques can be used in either humans or non-humans, some techniques are best adapted for humans and others for non-humans. Using a technique on a inappropriate species can lead to poor results. For example, humans can lay still for tens of minutes at a time during an imaging scan session, while rats or mice will not be able to do this simple behavior (without restraint or anesthesia). Before

About resolution

Many of the imaging techniques are concerned with resolution. Resolution in neuroscience is very similar to resolution in digital photography or computer monitors. A picture taken in high resolution contains more information than one in low resolution, and can therefore be more valuable. Imaging strategies consider two different types of resolution, spatial and temporal.

Spatial resolution refers to the ability to differentiate two points in space from each other. An imaging method with high spatial resolution means that two signals very close together can be identified as two different signals instead of one. Spatial resolution is usually measured in units of distance or volume. Electron microscopy (described in 6.3.1) offers the highest spatial resolution of the imaging techniques.

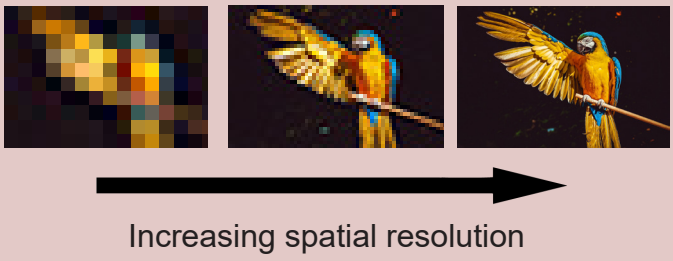


Figure 6.2 Methods with higher spatial resolution are better for identifying the location of different parts of the image. Precisely identifying the parrot's beak in the leftmost image (low resolution) is difficult, but is easy in the rightmost image (high resolution).

Temporal resolution refers to the ability to distinguish two events in time from one another. Temporal resolution is measured in units of time. Functional imaging techniques with high temporal resolution, such as electrophysiology, can differentiate two signals as close together in the hundreds of microseconds range.

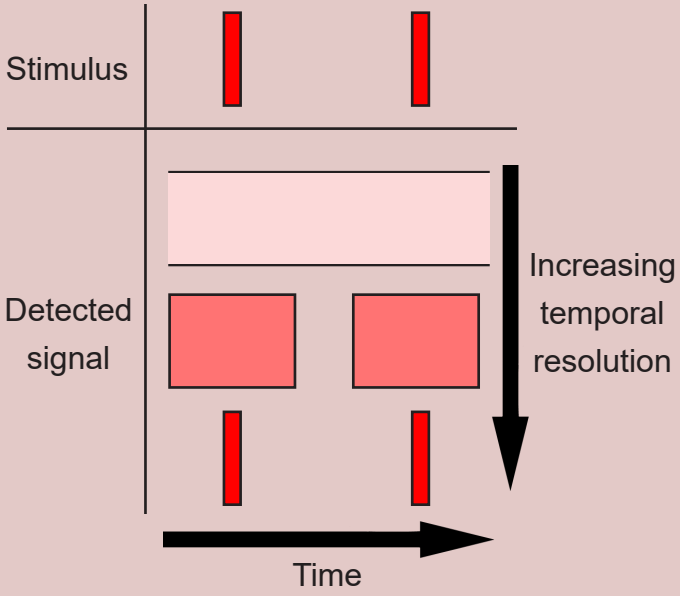


Figure 6.3 Methods with higher temporal resolution are better for precisely detecting when a signal occurs. With low temporal resolution techniques (top of detected signals), two discrete events spaced apart by time appear as one event.

getting started with our list of research methods, we should ask ourselves "Is it better to study the question in humans or in non-human animals?"

The case for human subjects

The main reason we conduct studies in humans is to understand the human nervous system. It is knowledge of the human condition that leads to the development of therapies that prevent, cure, or otherwise improve the quality of patient's lives. And while curing diseases in non-human animals is the purpose of veterinary medicine, curing humans is the primary end goal of biomedical research, and that goal is only truly met by testing on humans. A cure developed for rodents might give us a clue about the human

case, but without seeing how well it works in humans, the therapy won't save human lives.

Humans are great at following directions without much training, unlike non-humans - as anyone who had tried to house train a puppy can attest. We can follow simple sets of directions that untrained non-humans find difficult, such as "lie still". During functional activity scans, we can perform extremely complex higher-order cognitive tasks, such as "imagine that we give you 20 dollars, and we will take away a dollar when you answer a question incorrectly" (rats and chimpanzees have no concept of currency). We can describe our feelings in writing and self-report on our symptoms, both of which are tasks that are impossible in non-humans.

Believe it or not, recruiting human subjects for neuroscience or psychology studies is usually cheaper than studying that same question in nonhuman animal models. Collecting data from undergraduates is sometimes free, since many are required to participate in some number of hours of psychology studies to pass introductory level classes. Although the cost per patient differs depending on the nature of the study, many patients get paid as little as $10 an hour. On the other hand, housing colonies of mice, rats, or monkeys gets very expensive, very quickly.

The case for non-human subjects

But, it is often impossible to test every question in humans. Instead of always studying humans, scientists often use non-human **model organisms**, the most common organisms being the worm C. elegans, fruit flies (Drosophila melanogaster), zebrafish (Danio rerio), song birds, mice, rats, and macaque monkeys. The closer we move towards the human, the more similarities the model organism shares with us. Of the commonly used model organisms, macaque monkeys are the non-humans that are most similar to humans. We share 93% of our genetic material with macaques, but we still have different metabolic and physiological processes, and our behaviors

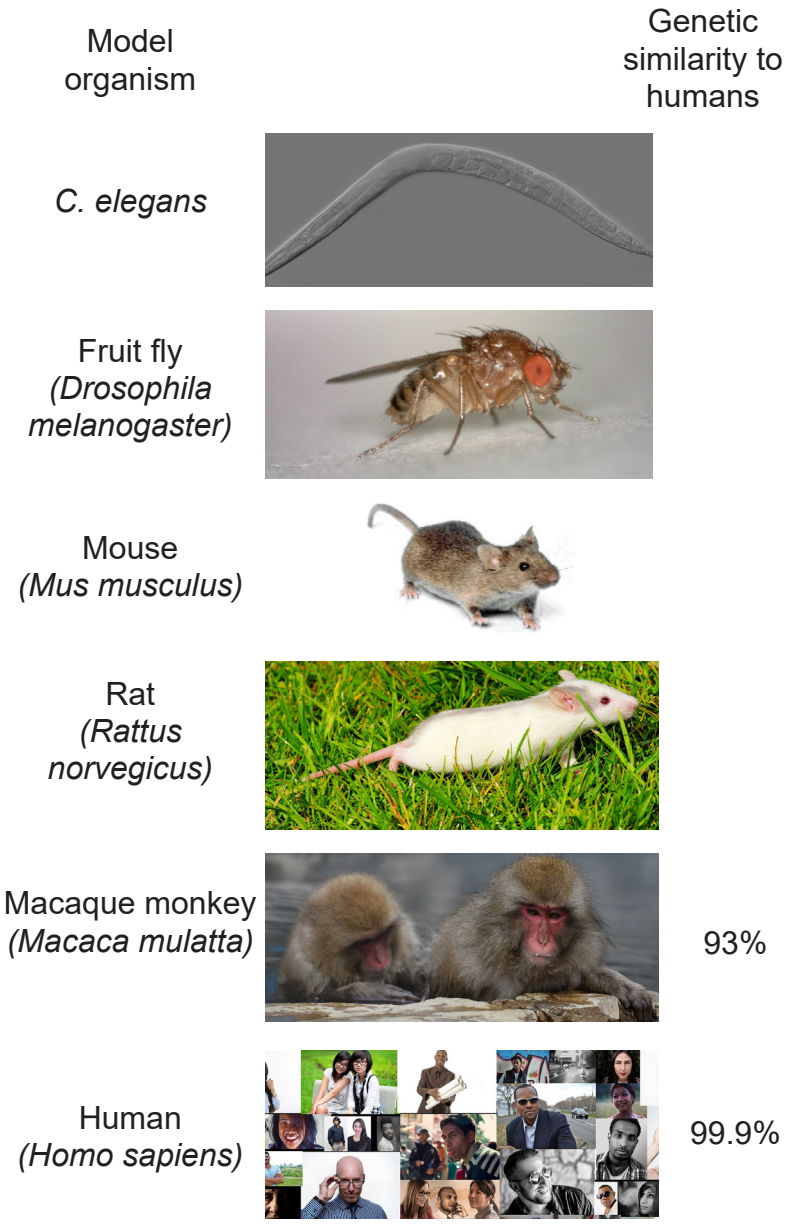


Figure 6.4 A short list of animal models that are often used in neuroscience research and their genetic similarity to humans. Humans are 99.9% genetically identical to other humans.

are much different from theirs.

Ethical constraints prevent us from performing experiments that may cause physical or psychological harm if performed in humans. We would never conduct a test on humans to assess what concentration of neurotoxin leads to brain damage (these experiments aren't done very frequently in nonhumans anyway.) Invertebrates, such as worms and fruit flies are not heavily regulated by ethics oversight committees, allowing scientists to conduct a wider set of experiments on these animals.

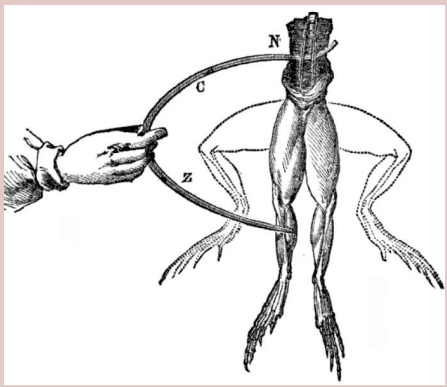
Experimental preparations

Performing an experiment in an intact, living organism, whether human or nonhuman, is described as an ***in vivo*** (Latin meaning "within life") preparation. The main strength of this strategy is that the data collected here are more predictive of the human condition, which is one of the main goals of biomedical research. However, the in vivo preparation has challenges, because thousands of variables within a living system are uncontrolled or still unknown. There are also very strict ethical limitations on the nature of experiments that can be done in vivo.

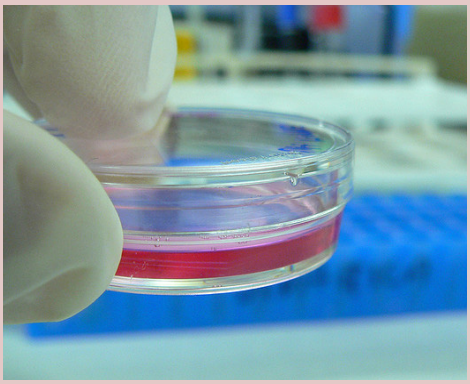
On the other hand, an ***in vitro*** (Latin meaning "within glass") preparation is an experiment performed on cultured cells or isolated molecules of DNA, RNA, or protein. These preparations have the opposite strengths and weaknesses of in vivo preparations. They allow for extremely good control over variables, but the results are less reliable in translating to a therapy. The regulations on these experiments are much more lax compared to in vivo experiments; most of the regulatory guidelines are to protect the experimenter rather than the patient or the experimental subject.

Falling in between these two preparations is an ***ex vivo*** experiment. In this kind of experiment, a section of the living organism is taken, such as a slice of brain, a tissue biopsy, or a detached frog leg. The strengths and limitations of these experiments are somewhere in between that of the other two preparations.

in vivo



ex vivo



in vitro

Increasing control over variables

Decreasing ability to predict therapeutic potential

Decreasing strictness of ethical regulations

Nonhuman animals perform all kinds of behaviors that humans never experience. These behaviors are unique to nonhumans and can only be studied in nonhumans of course. We will never be able to understand flying (birds) or slithering (snakes) as a means of locomotion by studying humans.

One thing that makes human studies difficult is how weird we are. We all come from very different backgrounds, bringing our own set of experiences, flaws, and biases into every research study. Your current mood and mind state can alter behaviors profoundly - just ask any one who gets “hangry.” In nonhuman studies, many of these variables are closely monitored. When we raise model organisms in the lab, almost every aspect of their life is controlled, from their food source, living conditions, and their day-night cycles, which eliminates several sources of variability that affect human studies.

Another factor that makes human studies challenging is that we are WEIRD: The acronym meaning that most patients in psychology studies are Western, Educated, and from Industrialized, Rich, and Democratic countries. One study estimates that about 96% of psychology studies use WEIRD subjects, whereas these people are only 12% of the total global population. This creates a bias in sample selection, since WEIRD people perform differently on behavioral tasks compared to others. Even our susceptibility to visual illusions is affected by our WEIRD-ness.

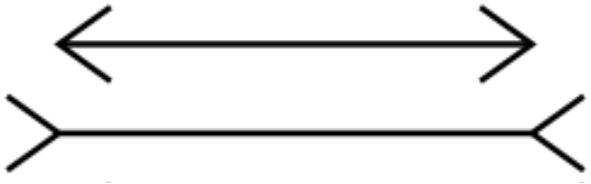


Figure 6.5 Many WEIRD people are susceptible to the Muller-Lyer illusion, and we perceive the lower line to be longer than the top line even though they are the same length.

Chapter 6 outline

6.1 Imaging brain activity

The earliest methods of analyzing nervous system anatomy were crude: manual dissection of the brain post-mortem (after death). In the 1860s, these methods received attention clinically with the studies conducted by French neurologist and surgeon Paul Broca. One of Broca's patients, nicknamed "Patient Tan," had severe language deficits and was only able to say "tan." Broca performed an autopsy after Patient Tan died, and observed a very localized injury in his brain that was likely the cause of Tan's language deficits.

At the same time, a pseudoscientific fad was gaining in popularity. Franz Gall, a German neuroanatomist, published a treatise describing **phrenology**, the belief that we can predict personality traits based on the shape of a person's head and the bumps on the outside of the skull. According to Gall, behaviors such as romanticism, individuality, and cautiousness are a result of differences in brain anatomy that push the skull outwards.

Today, we know that phrenology is a fad unsupported by the rigorous methodology of science. But there was a valuable lesson that emerged from phrenology which still persists in part today. Both phrenology and Broca's observation support **localization of function**, the idea that specific areas of the brain are important for certain functions. Today, we also

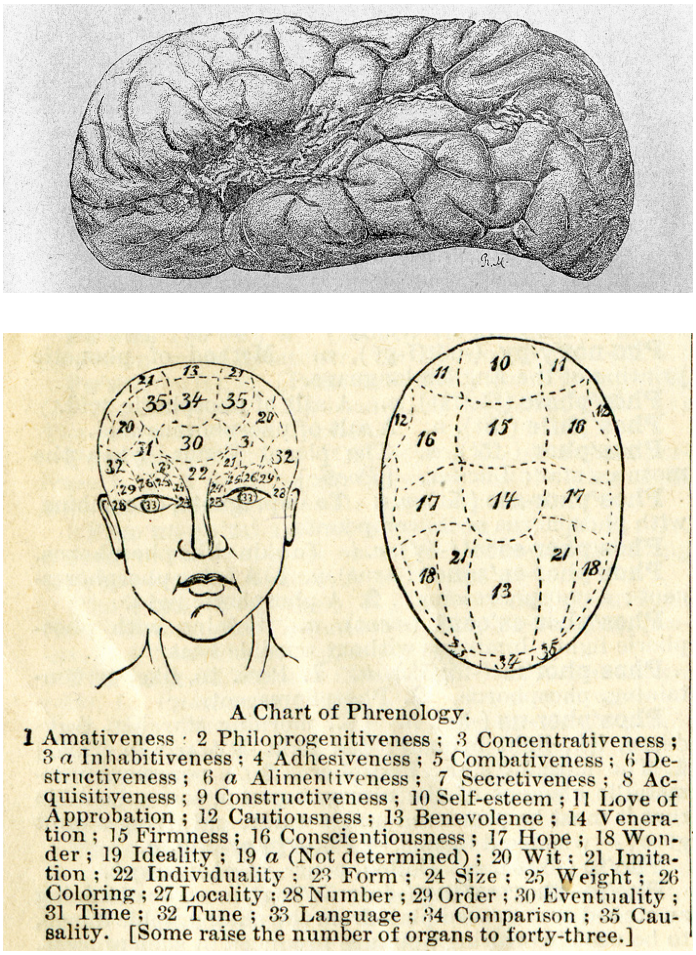


Figure 6.6 Drawing of Patient Tan's brain after Broca's autopsy (top). Tan's language disorder was likely a result of the highly localized damage to a region of his left hemisphere. Chart showing the localization of different behaviors according to the pseudoscience practice of phrenology (bottom).

think about the connections between two areas as being important for healthy brain activity. The techniques discussed below were developed to allow scientists to see some aspect of the anatomy of the nervous system, either gross anatomical differences or connectivity.

6.1.1 Computerized tomography scan (CT scan or CAT scan)

Example questions answered:
"Does the patient have a brain tumor, and where is the brain tumor located?"
"Are the meninges intact?"

The CT scan relies on X-ray technology that was developed in 1895. X-rays are high energy beams of electromagnetic radiation that are capable of passing through many physical objects. Traditional two-dimensional X-rays, such as those used to image a broken bone or tooth decay, use radiographic film to detect where the X-rays get blocked. When an X-ray passes unimpeded, it causes the film to darken. But, wherever the X-rays are blocked, the film remains white. Therefore, material that is more dense (such as bone) appears as white, less dense material (such as the air surrounding the body or CSF) appears dark. Other tissue are some shade of gray in between.

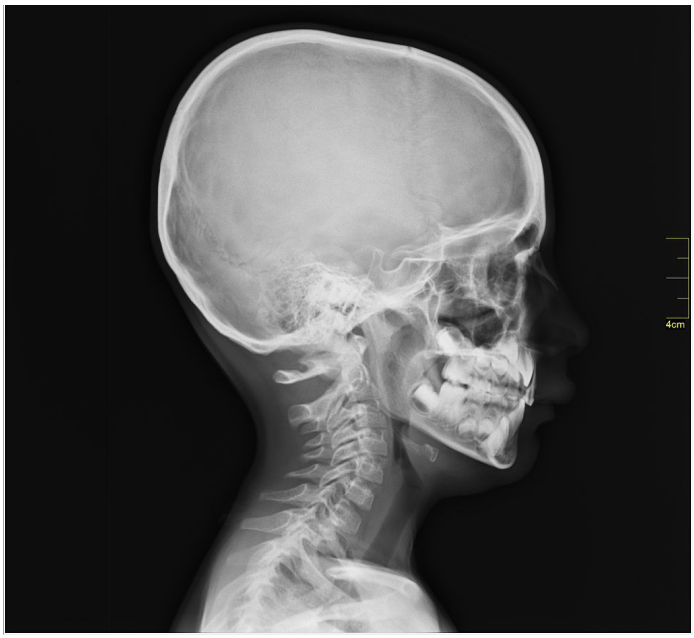


Figure 6.7 In X-ray images, dense materials (bone) appear as brighter while less dense smaterials (air) appear as black.

The **CT scan** is essentially a three-dimensional X-ray that revolves around the person as they move through the scanner. Instead of using radiographic film, the CT scan uses a computer that detects the passage of X-rays, located directly across the emission source of the X-ray. Instead of a flat, two dimensional-image, the CT scan uses an X-ray gun that revolves around the person's body as they advance slowly through a large circular hole. The computer is then able to compile the series of two-dimensional images and turn them into a three dimensional reconstruction that can be used to see the brain from any projection. CT scans give us a spatial resolution of about 0.5 mm. CT scans are generally used clinically to assess diagnostic changes over several days (such as before and after tumor removal or to

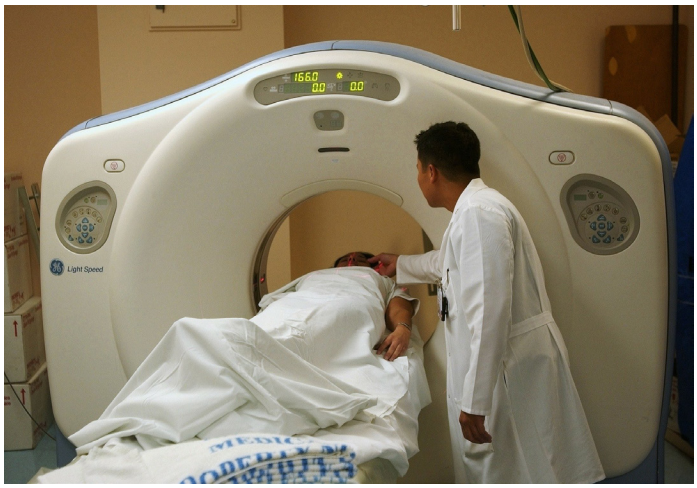


Figure 6.8 The patient lies on a table that moves through the middle of the CT scanner.

determine if an intracranial bleed has healed), so temporal resolution is not a major consideration.

As an anatomical analysis that can easily identify tissues of different density, it is great for identifying and diagnosing particular brain conditions. Brain tumors can be visualized in a CT scan, since they are identified by an increase in tissue density compared to normal brain tissue. Hydrocephalus, an abnormal and potentially deadly expansion of the CSF-filled ventricles, can be quickly identified by this analysis. Meningitis, an inflammation of the meninges, may present as increased contrast in the CT scan.

The big advantage of the CT scan is that it is noninvasive. You can use a CT scan in order to diagnose and identify the cause of a condition while a person is still alive, and hopefully work towards developing an intervention.

It is also a relatively quick technique. A full head CT scan takes only minutes, which allows for a rapid diagnosis of anatomical structures.

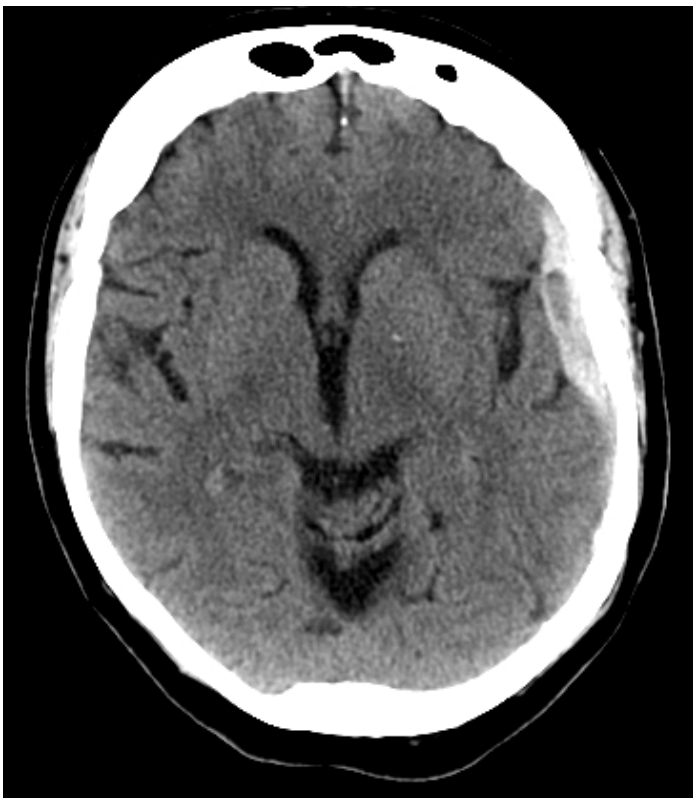


Figure 6.9 What do you notice that is abnormal in this CT scan?

However, X-rays are highly mutagenic. Prolonged exposure to X-rays dramatically increases the risk of developing various cancers, since X-rays interfere with the process of DNA replication. It is estimated that the radiation exposure in a single head CT scan is similar to the background exposure of X-rays in a few months. When a CT scan is prescribed, the diagnostic information gained from a CT scan is more important than the risks from increased radiation exposure.

6.1.2 Diffusion tensor imaging (DTI)

Example questions answered:
"Is the volume of the white matter tract medial longitudinal fasciculus important for normal language processing?"
"Does spinal cord compression cause neurological deficits?"

While a CT scan is great for detecting gross anatomical anomalies like tumors or intracranial bleeding, it has a difficult time with subtle anatomical changes like differences between gray matter and white matter tracts. A technique for identifying these differences was proposed in 1994, called **diffusion tensor imaging (DTI).**

DTI quantifies white matter because of the morphological features of white matter. More specifically, DTI uses MRI technology (see section 6.2.3 for more information) to detect and quantify the movement of water molecules, which moves differently in gray matter than white matter. A single molecule of water in the middle of a cup can move in any direction with equal probability. This type of motion is called **isotropic diffusion**. However, the movement of a water molecule in biological tissue is not completely random, largely because

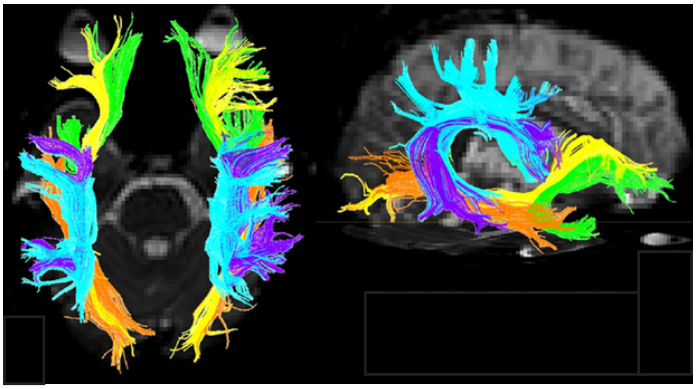


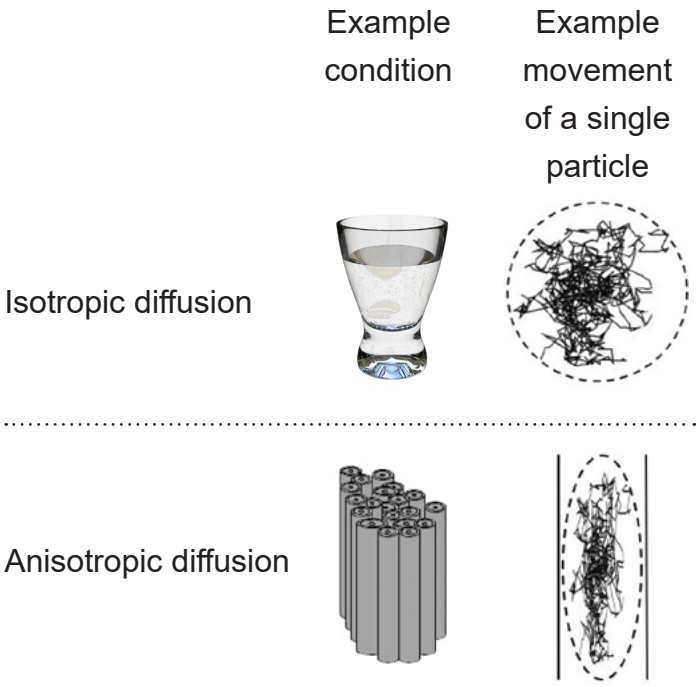
Figure 6.10 Diffusion tensor imaging can be used to visualize tracts of white matter, indicated by different colors.

the brain is made up of heterogeneous tissue. White matter is very different from gray matter. A water molecule is more easily able to diffuse along the same direction as a tract of white matter, but has a difficult time moving perpendicularly across such tissue. The difference in molecular motion is called **anisotropic diffusion**. DTI uses the premise that anisotropic diffusion is observed in white matter tracts.

DTI can give spatial resolution on the order of millimeters. It is generally not used for monitoring changes over time, so temporal resolution is not a major consideration.

Axonal projections are directional, with the soma at one end and the axon terminal at the other. One of the shortcomings of DTI is that it cannot give us information about the directionality of the axonal projections.

Figure 6.11 In isotropic diffusion (top), a single particle is able to move in any direction randomly. However, in anisotropic diffusion (bottom), a single particle is more likely to move along a certain pathway, aligning with the cellular architecture.



6.1.3 CLARITY

Example questions answered:
"Do neurons in layer 5 of motor cortex send axonal projections into the spinal cord?"

The methods described above only provide rough anatomical information, and therefore have limits to their value. CLARITY is a revolutionary anatomical technique that gives microscopic level analysis. First published in 2013, CLARITY was developed at Stanford by the lab of Dr. Karl Deisseroth. The value behind CLARITY is that we are now able to visualize connectivity in the brain at a microscopic level, giving us amazing spatial resolution at the order of a microns.

One of the problems with visualizing the structures of the brain at a microscopic scale is that cells have a lot of membrane, which is mostly made up of lipids. Instead of letting light pass unimpeded, lipids refract light in somewhat unpredictable ways. Even cytosol has all sorts of particles floating around in it that cause light to bend and bounce around. These lipids and other cellular particles prevent light from passing through nervous tissue, minimizing our ability to use transmitted light to see the brain.

In CLARITY, the brain is first flushed with chemicals that form a gel matrix that surrounds every cellular structural component, from the spines on the dendritic arbor all the way to the axonal terminals. Then, using a chemical detergent, the cellular lipids get solubilized and washed away while the gel matrix remains unaffected. This allows us to now see the “mold” that remains around where the cell membranes used to be. By washing away the light-deflecting lipids, we can now see where the connections were, causing the brain to appear transparent.

Unlike the CT scan, CLARITY can’t be applied in an intact living organism. CLARITY is an extremely destructive process that destroys tissue. The surrounding gel matrix is like a mold of a hand. You can see all the anatomical features of the hands - fingers, wrinkles, nails, and so on. But the mold is just an image, unable to function. The mold of the hand cannot grip things, cannot feel things, and does not have any bones / tendons / muscles. In using CLARITY, all function is destroyed.

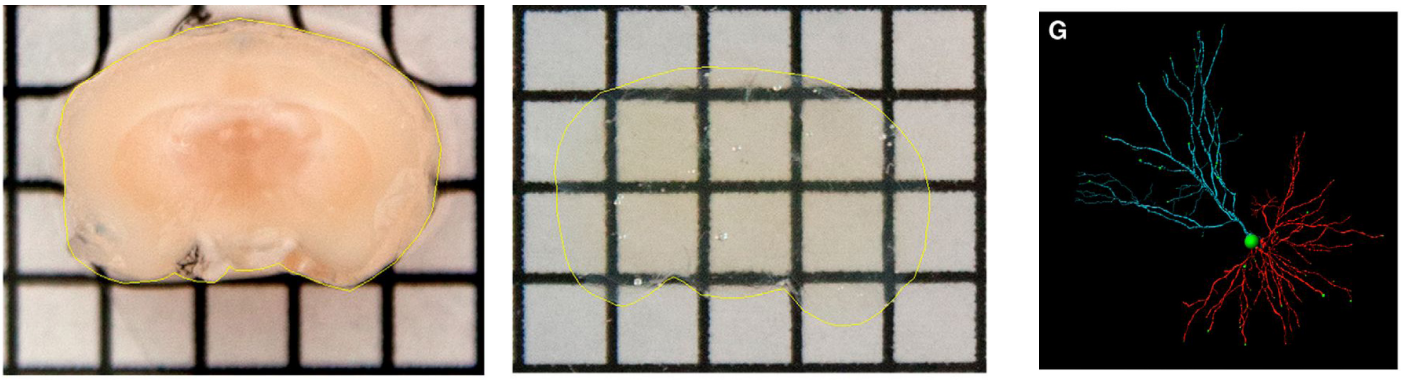


Figure 6.12 With CLARITY, a whole animal brain (left) can be made transparent (middle), allowing for visualization of individual cellular structures (right).

6.2 Imaging brain function

Anatomical analyses are certainly a meaningful way to study aspects of the nervous system. But being able to measure and record brain function is important for answering different types of questions. After all, the brain of a recently deceased person might look identical anatomically to that of a living person, but they would exhibit significantly less function.

6.2.1 Electroencephalography (EEG)

Example questions answered:
"At what times of the night does the sleeping brain exhibit synchronized neuronal activity?"
"How does brain activity change when a person is having a seizure?"

The **electroencephalogram (EEG)** is one method by which we can observe electrical activity of the brain. When neurons send action potentials, several charged ions move across cell membranes, which causes a change in the electrical charge of the adjacent area. Some of these charges are very large, especially when several millions of neurons are sending action potentials at the same time. These currents produced by neurons of the outer surface of the brain (cortex) can be so large, that they are detectable outside the skull at the surface of the head. The EEG works by recording these changes at different areas of the brain. The technique was first used in the 1920s by psychiatrist Hans Berger, who invented the technique. The theory behind the EEG is the same used in an electrocardiogram, or ECG, to detect the electrical activity of heart cells.

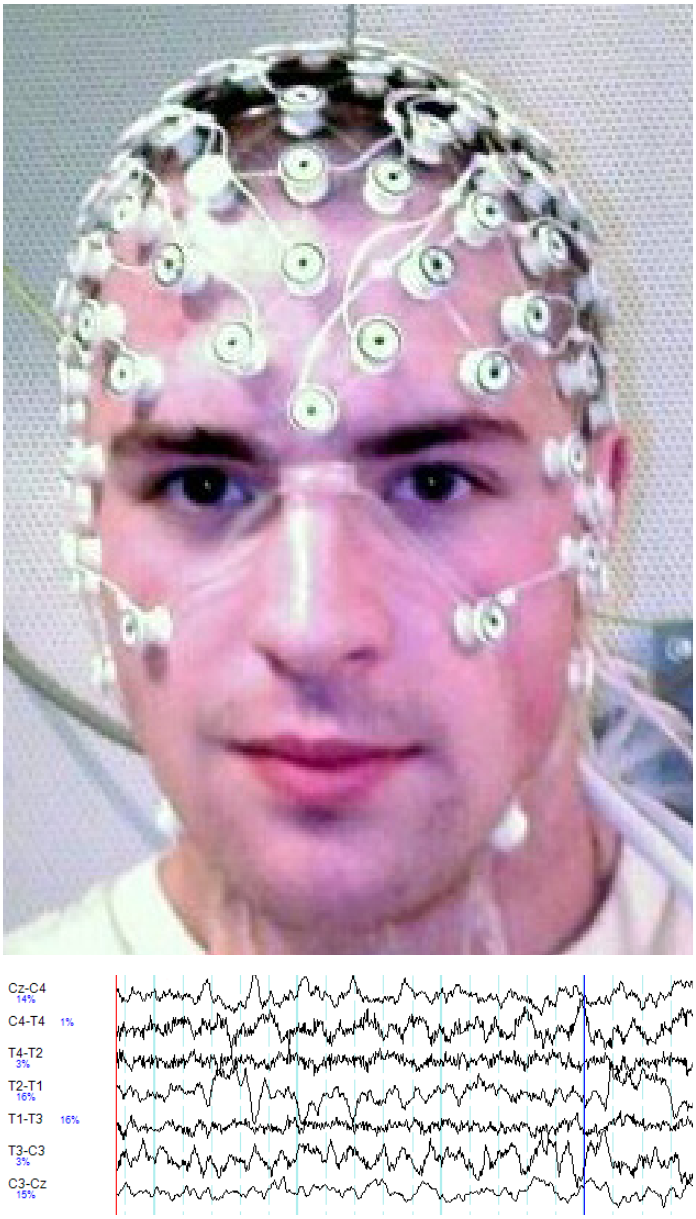


Figure 6.13 The electroencephalogram (EEG) is a noninvasive measure of neuronal activity (top). The information gathered (bottom) reflects changes in electrical potential measured at the scalp.

To perform an EEG, a gel is applied onto small patches of the scalp. This gel is basically a sodium chloride gel that acts as a conductor, allowing the currents to be picked up by a series of adhesive electrodes. EEG systems used in

hospitals for diagnosis or in laboratories for research studies may have anywhere between 20 to 128 electrodes. Each electrode is sensitive enough to detect voltage deflections as small as 10 microvolts. Each electrode is placed on a specific area of the scalp, and the information from each electrode runs into a computer. From there, the computer can then perform calculations to determine which cortical areas to the brain exhibit what patterns of activity.

EEG software is capable of examining the rapidly-fluctuating electrical potentials and dissecting out the several components buried within the complex signals. From one waveform, it can pick out high frequency components called beta waves (between 13 and 30 Hz), low frequency components called delta waves (between 0 and 4 Hz), and all frequencies in between.

The EEG is a noninvasive functional analysis method. Nothing permanent is done to the person when they get an EEG because it is only detecting information. Wearing a cap of electrodes may be annoying, but it is a harmless procedure. Because of the experimental set up for EEG, there are many cases where using an EEG is preferable. The EEG is a standard part of a polysomnogram, a series of tests that can be done to evaluate sleep disorders. In fact, our best and most reliable characterization of the phases of sleep rely on brain activity measures as recorded by an EEG (Chapter 13). Relatedly, EEGs are also used clinically when a person is under anesthesia as a tool to evaluate their level of unconsciousness.

Not only is the EEG noninvasive, but it is also relatively cheap - at least in comparison to many of the other methods described in this chapter. It does not require heavy equipment and is mobile. All the machinery needed to run an EEG can be contained within a backpack.

EEG is also useful for the diagnosis of a variety of brain disorders that are a result of aberrant cortical neural activity. The most well known of these is epilepsy. In epilepsy, when a person has a seizure, it is believed that large masses of neurons are sending action potentials across the cortex, producing very large signals that can be detected by scalp electrodes. Migraine, a debilitating condition, may be due to waves of unusual neural activity that sweep across the cortex. Other applications of EEG may include helping diagnose Alzheimer's disease, depression, and possibly ADHD.

The EEG has great temporal resolution. Because it is detecting changes in currents, it is capable of sampling in the range of 10,000 Hz, meaning that for each electrode, 10,000 data points can be collected per second. Since action potentials happen on the time scale of milliseconds, being able to calculate changes in electrical potentials at such speeds gives us the ability to assess brain activity very precisely.

But, the EEG has poor spatial resolution, only detecting signals that originate in the outer most layer of the brain. More electrodes increase spatial resolution, but even at 128 electrodes, an EEG only has spatial resolution to the level of about 7 cubic centimeters.

6.2.2 Positron emission tomography (PET scan)

Example questions answered:

"Which areas of the brain decrease in activity when a person experiences mild cognitive impairment?"

"Do drug-dependent people have a high density of opioid receptors?"

The **positron emission tomography (PET)** scan is an application of nuclear medicine best known for its applications in the medical setting for the diagnosis of cancer. Before the PET scan, a radioactive compound called a **tracer** is injected into the bloodstream. This tracer is a compound where an atom is substituted with a radioactive isotope, such as tritiated hydrogen (^{3}H) or fluorine-18. The PET scanner itself is a large circular device that looks similar to the CT scanner. The tracer is chemically unstable, and it spontaneously emits positrons. When those positrons interact with the electrons of nearby molecules, two gamma particles are emitted in perpendicular directions. These gamma particles are then detected by the PET scanner as the person moves through the machine.

A common tracer that is used for imaging brain function is **fluorodeoxyglucose-F18**, or **FDG**. FDG is a radioactive analog of glucose, one of the main sources of cellular energy. Highly

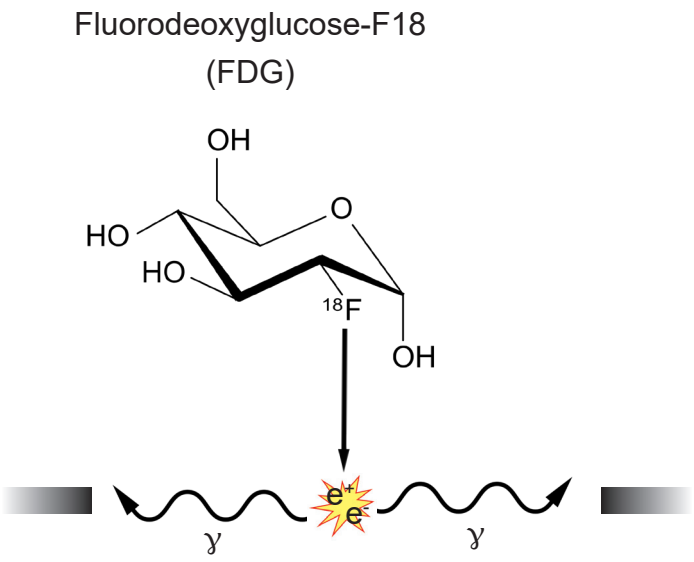


Figure 6.14 When a positron (e^+) emitted by the tracer compound (FDG) interacts with nearby electrons, gamma rays (γ) are emitted at perpendicular angles which are detected by the PET scanner (gray boxes).

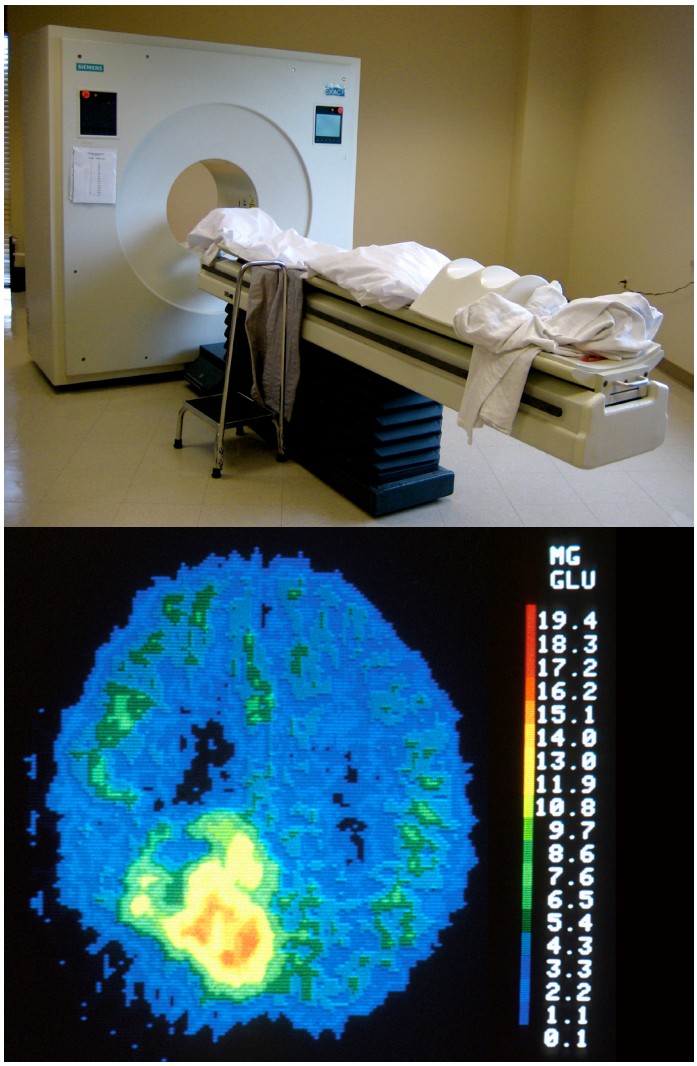


Figure 6.15 The positron emission tomography (PET) scanner (top) is quiet compared to the CT scan or MRI. It can generate data sets (bottom) that visualize the approximate location of highly-energetically demanding tissue (yellow and orange), such as a brain tumor.

metabolically active areas of the body take FDG into the cells. Energetically demanding areas of the brain require more energy to do their functions (the same reasons that tumors, rapidly-reproducing patches of tissue, appear very bright in PET scan analyses). In order to produce the energy needed for increased neuronal activity, the brain changes perfusion levels by dilating blood vessels in order to bring more glucose to those areas. Therefore, when an area of the brain

increases in energetic demand, that change can be detected by identifying the increase in glucose movement.

PET scans can be effective at diagnosing and identifying the location of tumors in the nervous system. It also provides an overall picture of brain activity, which may be useful in diagnosing disorders of cognitive deficits, like dementia associated with Alzheimer's disease or Pick's disease. PET scans have also been used to image the activity of specific brain areas as a person performs behavioral tasks, but this use of PET scan has largely been replaced by functional magnetic resonance imaging (fMRI; see section 6.2.3).

A second application of PET scanning is to visualize levels of receptors in vivo. This experimental approach relies on the nature of the interaction between specific radioactively-labeled compounds and receptors (such as [11C] raclopride and dopamine receptors, or [18F] ASEM and acetylcholine receptors). Once the radiolabeled compound binds to the receptor, the PET scan is able to detect the location of the radioactive signal and the strength of that signal.

The downside of the PET scan as a diagnostic tool is similar to a limitation of the CT scan. A person is exposed to radioactive compounds and gamma wave radiation, which are potentially mutagenic.

In images produced by a PET scan, it is often difficult to identify boundaries between tissue, even between dramatically different internal organs. To make up for this deficit, PET scans are frequently performed simultaneously with an anatomical analysis like a CT scan.

PET scans generally have very poor spatial and temporal resolution. With PET scanning, you can really only see differences between areas if the volume is in the range of 5-10 cm^3. Anything smaller than a cubic centimeter will be impossible to detect with current PET scanners. As to the speed of the PET signal, it often takes tens of seconds or minutes before a change in signal can be observed and detected.

6.2.3 Functional magnetic resonance imaging (fMRI)

Example questions answered:

"Do neurons in the right hemisphere cingulate gyrus increase in activity when a person sees their loved ones?"

"Which areas of the brain change in activity when a person is planning a motor action?"

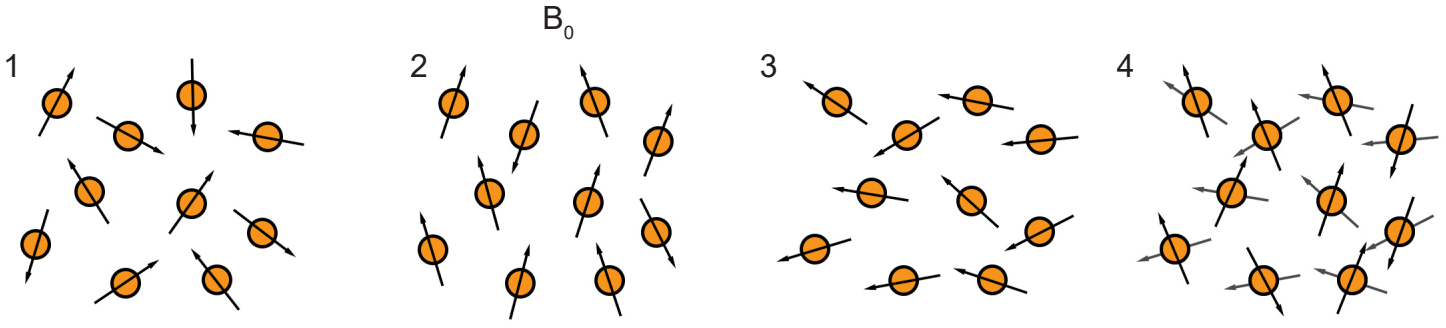


Figure 6.16 (1) Protons spin around an axis, aligning randomly at rest. (2) In the presence of a magnetic field B_0, their spin aligns with or directly opposite of the magnetic field. (3) Exposure to radiowaves knocks the protons out of their alignment into a high energy state. (4) As the protons return back to their alignment with the magnetic field, they release energy.

The functional magnetic resonance imaging (fMRI) technique is probably the most well-known method of studying brain activity. Because fMRI can be performed while a person is engaged in a task, many research studies use fMRI as a means to correlate behavior with activity patterns in specific parts of the brain.

An fMRI machine basically consists of two main components. As with the CT scan or PET scan described above, the fMRI machine is a circular tunnel through which a person on a table moves. As the person moves through the scanner, an extremely powerful magnet revolves around their head. The power of a typical magnet used in a hospital fMRI may be as powerful as 10,000 gauss (1 Tesla), strong enough to lift a car. The more powerful fMRI machines can be as powerful as 100,000 gauss (10 Tesla). The stronger the magnets, the better the spatial resolution that the machine can produce (our current best spatial resolution is on the order of millimeters). The second component is a device that emits radio waves. Just like the magnet, the radio emission device revolves around the head.

Both the magnetic field and the radio waves are crucial for the fMRI to work. They both interact with protons, the charged subatomic particles that have a small polarized direction because they spin around an axis. Initially, when these protons enter into a strong magnetic field, each proton charge aligns either with or directly against the direction of the field. Then, when hit by a radio wave, the protons lose their alignment, going into a high energy state. The protons then fall back to a low energy state as they return to alignment with the magnetic field. This process is what is being detected by the fMRI.

As it turns out, the protons in oxygenated hemoglobin are sensitive to the magnetic field (diamagnetic) while deoxygenated hemoglobin is not (paramagnetic). Because of this difference, we are able to use sensitivity to a magnetic field as a measure of changes in oxygenation levels. Like the PET scan, the fMRI hinges on the idea that more active areas of the brain have different metabolic demands than less active areas of the brain. When there is more activity in one area of the brain, the neurons in that area need more oxygen, and the adjacent blood vessels react by dilating. This change in blood flow is detected by the fMRI, and is called the **blood oxygenation level-dependent signal**, or **BOLD signal**. Temporal resolution of the fMRI is limited by the speed of blood vessel dilation, which occurs on the order of seconds to tens of seconds.

The main reason fMRI is useful in so many research applications is that you are able to visualize brain activity real-time during the performance of complex behavioral tasks. You can present specific visual stimuli to a person in an fMRI scan and evaluate which parts of the brain changes in activity. For example, seeing pictures of faces causes increased blood flow into the fusiform face area. You can ask a person to perform a gambling task and evaluate the areas

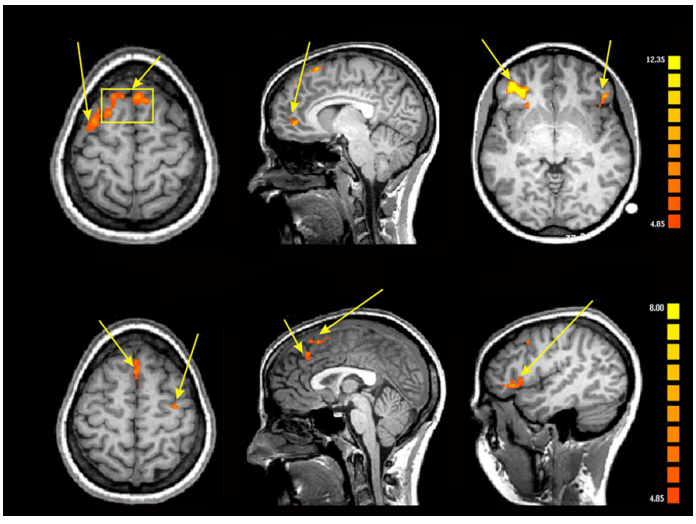


Figure 6.17 The fMRI scan detects the location of blood flow changes as a person pays attention to different visual stimuli (top vs. bottom).

of the prefrontal lobe that are responsive to risk taking.

Although the technique has a great capacity for analyzing brain function, the nature of the machine itself presents limitations. The scanning tunnel is very small and claustrophobic, making it difficult to safely study anxiety or panic disorder without endangering the patient. The machine can also be very loud, which is not trivial if you are interested in studying younger patients. The use of a tremendously powerful magnetic field presents a different set of limitations. At the risk of severe injury or death, the patient entering the scanner cannot have any magnetosensitive implants, such as metallic aneurism clips, intrauterine devices, or shrapnel. Even older generation tattoos have trace amounts of metal that cause burns when exposed to the magnets of the fMRI machine.

The data collected by fMRI can be very difficult to analyze and are frequently subject to false positives. The change in BOLD signal measured in an fMRI is very small: one study estimates that perfusion increases by only 0.4% even under the greatest activation. Using standard fMRI analysis methods, an infamous study demonstrated that even a dead salmon in an fMRI exhibits brain activity that is similar to those reported in other fMRI studies.

FMRI also assumes that increased blood flow is directly correlated with the amount of neural activity, which may not always be the case.

6.3 Imaging the cells of the nervous system

The methods thus far discussed are limited in how closely we can see the components of the nervous system. For example, it could be useful to try and see the specific cellular features of a healthy neuron and compare that with the features of a neuron that has been damaged by multiple sclerosis. To be able to visualize the morphology of structures smaller than hundreds of microns, you'll need a different set of techniques.

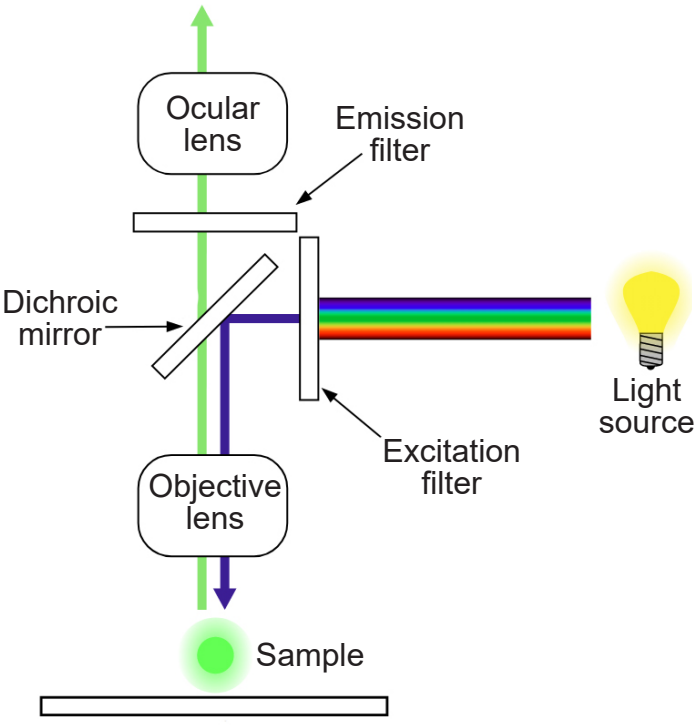
6.3.1 Microscopy

Example questions answered:
"What is the diameter of the soma of a granule cell?"
"How are the retinal neurons at the fovea different from the neurons in the periphery?"

Among the techniques described in this chapter, short of physical dissection or autopsy, microscopy may be the oldest. Microscopy began with the Dutch scientist Antonie van Leeweunhoek in the late 1700s. Originally, microscopy was more of a novelty for entertainment rather than a piece of equipment fit for a laboratory. Using a series of precisely-crafted lenses and a bright source of light, Leeweunhoek was able to see tiny animals living in water.

Microscopy gives us the ability to see things at a level of resolution that would otherwise not be possible. Compared to today's microscopes, the earliest scopes provided poor magnification, able to bring us only about 40 times closer than the unaided eye (these days, a standard laboratory microscope provides a level of magnification up to 400 or 1,000 times closer). Other types of microscopes, such as the **electron microscopes** developed in the 1930s, use an electron emission device in conjunction with high speed detectors to visualize structures that are on the order of nanometers, giving us a zoom of up to a million times!

A different form of microscopy relies on the fluorescent properties of molecules such as **green fluorescent protein (GFP).** Fluorescence microscopes use light sources such as filtered light or lasers which emit light at specific



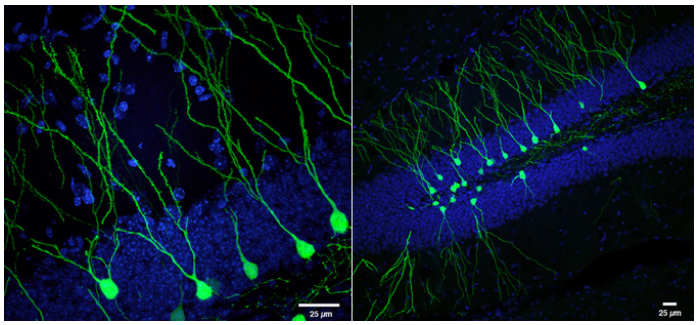


Figure 6.18 The fluorescence microscope (diagram, top) can be used to excite GFP (green) expressed in hippocampal neurons (bottom).

wavelengths. Photons at specific wavelengths activate the proteins, which in turn emit light at a different wavelength. The emission of this second color of light can then be detected.

Microscopy is a standard part of a neuroscience research kit. Most of the techniques below will rely on microscopy to visualize some component of the nervous system.

6.3.2 Staining

Example questions answered:
"Where is the white matter located in the brain stem?"
"What is the morphology of dendritic arbor of a cerebellar Purkinje cell?"

Staining is an imaging method that is often used in conjunction with microscopy. Thin slices of brain tissue are exposed to various chemical processes. The chemicals that are used for staining have different affinities for parts of cells. For example, there are stains that specifically stain myelin (Luxol Fast Blue), some that only stain neurons (cresyl violet), and some that stain only the genetic material found in cell nuclei (DAPI).

One of the most influential biological stains was developed in the late 1800s by the Italian anatomist Camillo Golgi. The **Golgi stain** is a silver-based stain that filled every single part of the neuron, from the dendrites to the axons, turning the cells black. The stain is only taken up by less than 1% of neurons, and because of the low rate of staining, individual neurons could be traced in their entirety and drawn. This stain led to a contentious battle between Golgi and a rival neuroanatomist Santiago Ramon y Cajal over the nature of the connections between neurons.

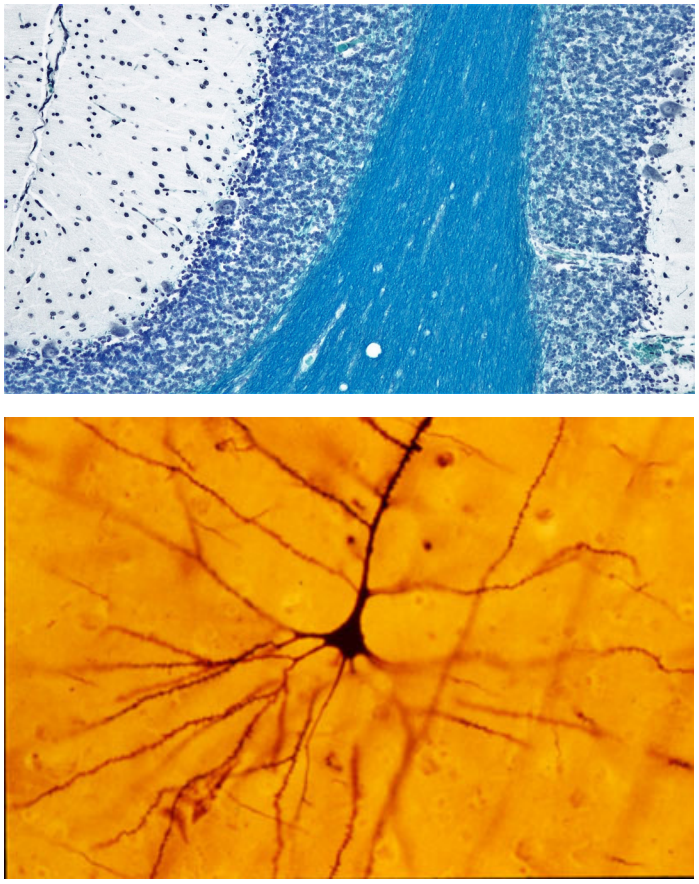


Figure 6.19 A slice of cerebellum, with the myelin stained with Luxol Fast Blue and the neurons stained with cresyl violet (top; 100x magnified). A pyramidal cell treated with a Golgi stain (bottom).

For staining to work, the tissue needs to be subjected to a series of chemical processes. First, the tissue needs to be **fixed**. Fixation is a chemical process that is accomplished by exposing the tissue to a chemical like **paraformaldehyde (PFA).** The most effective way to expose every part of the body to fixative is to "hijack" the endogenous circulatory system by flushing fixative through the arteries, a process called **perfusion**. Chemically, fixatives cause adjacent proteins to become covalently bonded (cross-linked), a process which causes the proteins to become unchangeable - they become "fixed" in time. These fixatives are very harsh chemicals, and usually kill microorganisms and inactivate

the endogenous enzymes that normally degrade biological tissue. (As a side note, fixatives are particularly nasty carcinogens that can permeate easily through latex gloves.)

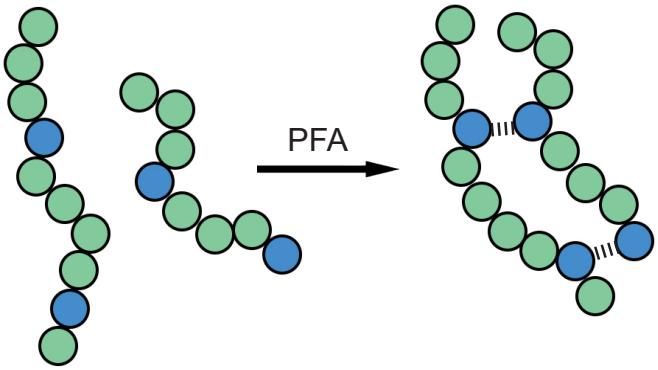


Figure 6.20 A chemical fixative such as PFA creates covalent bonds between lysine (blue) amino acid residues of proteins.

After fixation, devices such as a **microtome** or **cryostat** are used to slice the brain into sections as thin as 10 microns. Stains do not always reliably pass all the way through thick sections of tissue such as an intact brain. With thin slices, chemical stains are able to permeate through the depth of the tissue.

Figure 6.21 A microtome (left) or a cryostat (right) are devices for taking brain sections. The microtome fits on a tabletop, and cuts accurately down to 100 uM thin. The cryostat is larger, and can cut 10 uM thin slices.

Certain staining methods can also be used to determine projections of neurons. These techniques are collectively called **tract tracing** methods. An **anterograde trace** stains cells from the soma to the axon terminal, while a **retrograde trace** is taken up by the axon terminals and stains the soma. These strategies are helpful for visualizing the nature of connectivity in the nervous system.

Staining cannot be done on living tissue; it can only work on fixed tissue.

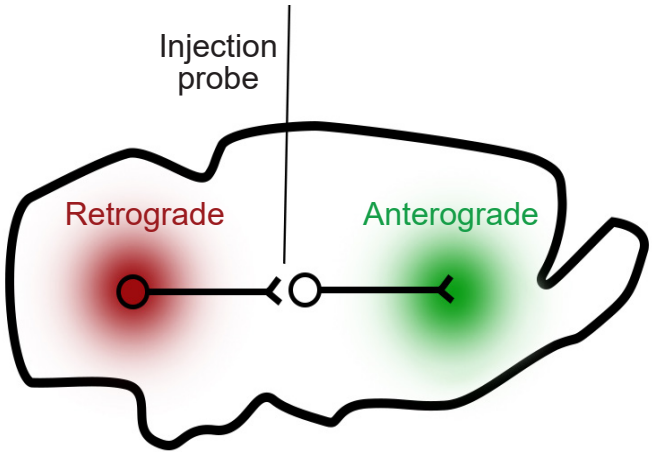


Figure 6.22 In tract tracing, injection of different staining dyes can help us identify the connectivity of neurons. If an injection site contains somata, the dye is transported in an anterograde direction towards the area where the axon terminals are located (green). If the injection site contains axon terminals, the dye is taken up and transported in a retrograde direction towards the cell bodies (red).

6.3.3 Immunohistochemistry staining

Example questions answered:
"Where is the protein actin located in a neuron?"
"Is the protein PSD-95 expressed in the same areas as the NMDA receptor protein?"

Immunohistochemistry (IHC) is a modification of the staining methods that gives

increased specificity over the other chemical stain methods presented above. IHC is used because it is good for identifying the location of specific proteins at a subcellular level. In other words, you can use IHC to figure out where specific neurons are located inside or on the cell.

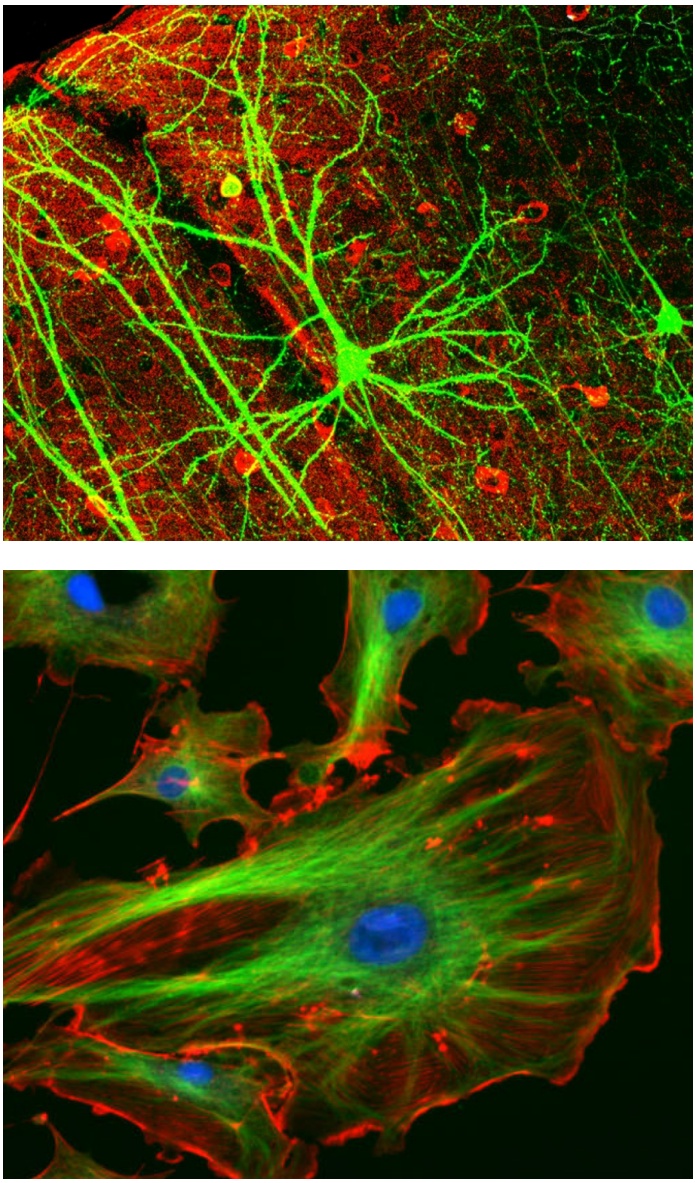


Figure 6.23 At low magnification (top), immunohistochemistry in conjunction with fluorescence microscopy can help us identify the morphology and location of different populations of cells. At high magnification (bottom), we can identify the location of different proteins within a cell, such as tubulin (green), which makes up the microtubules.

As with other staining methods, IHC begins with thinly prepared sections of fixed tissue. Instead of exposing the tissue to various chemical stains, IHC utilizes the molecular properties of **antibodies**, proteins that function in the immune response. Each antibody has a part of their Y-shaped structure that is able to bind to other targets (which are called **antigens**) with extremely high specificity. This other target could be a specific protein, such as serotonin receptors or amyloid precursor protein, or an enzyme such as choline acetyltransferase. Antibodies are created with a particular target in mind, so when the antibody is bathed over the brain tissue, the antibody will adhere only to that target.

The IHC protocol often uses two different antibodies. The first antibody, called the **primary antibody**, is selected because of its ability to bind to the antigen that you are interested in studying. For example, if you wanted to search for the location of a protein called NeuN, the primary antibody would be anti-NeuN. Afterwards, the tissue will be exposed to the **secondary antibody**.

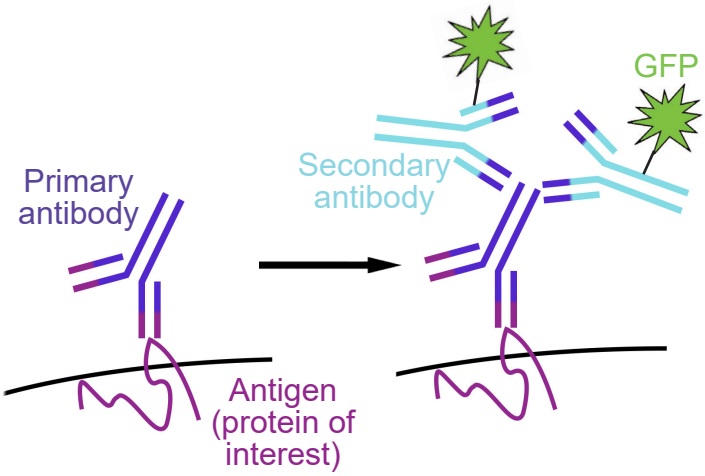


Figure 6.24 The primary antibody (purple) is chosen because it binds to a specific protein of interest (magenta). In a second step, the secondary antibody (light blue), conjugated with a fluorophore (in this case GFP), binds to the primary antibody.

This second antibody is chosen because its target is something that is found on the primary antibody. As an additional feature, the secondary antibody is conjugated with a light sensitive **fluorophore**, or light-producing molecule. Activation of this fluorophore on the secondary antibody under fluorescence microscopy is then used to identify the location of the primary antibody.

It is also possible to use antibodies and fluorescence microscopy to visualize the elements of cells in a culture dish (in vitro). The theory behind this technique is the same as in IHC, but the name is different. Instead, this approach is called immunocytochemistry (cyto = cell, while histo = tissue). This immunolabeling strategy can also be applied to a whole brain after it has been made transparent through CLARITY (section 6.1.3).

One of the main difficulties with immunochemistry is that the antibodies often exhibit **nonspecific binding**. Although antibodies are used because of their ability to recognize and differentiate between similar-looking antigens, often they bind to the wrong targets. When they do, it creates a false positive, causing a fluorescent signal to appear when there is actually no protein present. Nonspecific binding also causes an increase in background noise, making it more difficult to identify a genuine fluorescent signal, indicating presence of the protein of interest. Various steps can be taken during the immunochemistry procedure that minimize the likelihood of nonspecific binding, such as thorough rinsing and exposure to blocking agents.

For an immunochemistry protocol to work, an antibody must exist for the specific target of interest. However, an antibody has not been developed against every single protein or enzyme. Another issue is that some proteins are structurally very similar to other proteins, and an antibody has not yet been developed that can differentiate the two. For example, the structure of the dopamine receptor type 2 is so similar to type 3, that there are no antibodies that can target one without the other.

6.4 Changing nervous system activity

Visualizing the gross anatomy or cellular structures of the nervous system allows us to ask very meaningful scientific questions. However, sometimes you are interested in controlling some part of the nervous system, either by increasing or decreasing activity. Using the following neural manipulation strategies, it becomes possible to establish causation between some aspect of the nervous system (neuron, structure, neurotransmitter, gene, etc.) and function.

Early neural manipulation strategies trace back to the late 1700s with the accidental discoveries of the Italian biologist Luigi Galvani. With a dissected frog leg on his lab bench, Galvani discovered that he could cause the muscles to twitch by electrically activating the nerves that innervated the muscles of the leg. He called this observation "**bioelectricity**," which is the idea that endogenous electrical activity is important for muscle control, and by extension, the activity of the whole organism.

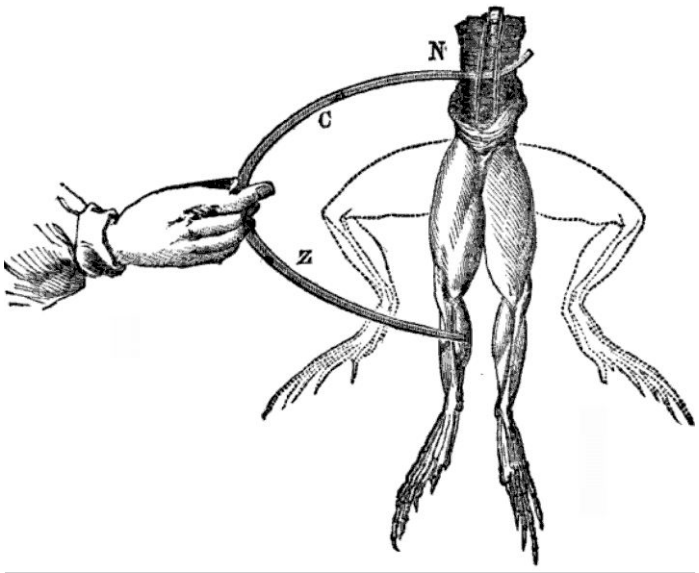


Figure 6.25 Luigi Galvani observed that electrically activating a nerve caused muscular activity in the frog leg. The phrase "galvanized to action" is a reference to Galvani's discovery.

Neural manipulation became possible in humans following the work of the brain surgeon Wilder Penfield. In the 1950s, Penfield worked to develop a treatment for epilepsy. In this procedure, the patient is given a local anesthetic while the skin above their skull was resected, allowing access to the brain. The surface of the brain has no pain receptors, so brain surgery can be done while the patient is completely awake. With the patient's brain exposed, Penfield used a brief electrical pulse delivered by an electrode placed at the surface of the brain. The stimulation activates descending motor systems, which caused muscle contractions in very specific muscle groups of the patient. For example, stimulating the dorsal most part of the motor cortex caused a twitch of the upper leg and hip muscles. By moving incrementally across the motor cortex, he was able to draw a graphical representation of which areas of the motor cortex controls which muscles.

The techniques presented here are all methods by which we can induce some type of change in nervous system activity.

6.4.1 Electrophysiology (Ephys)

Example questions answered:
"How frequently do neurons in the striatum fire action potentials?"
"Does increasing cortical activity change the excitability of spinal cord neurons?"

Electrophysiology relies on using specialized equipment, informally called a **rig**, to measure and manipulate the electrical properties of the nervous system. Originally,

electrophysiology was used to detect the difference in electrical potential between the inside of the neuron and the outside. The earliest electrophysiology techniques examined the giant axon of the *Loligo* squid. The axon itself is so large with a diameter around 1 mm, the inside of these neurons are very easy to access by running a conductive probe into the cell. Two researchers, Alan Hodgkin and Andrew Huxley, earned a Nobel prize in 1963 for the work that came out of these electrophysiological techniques.

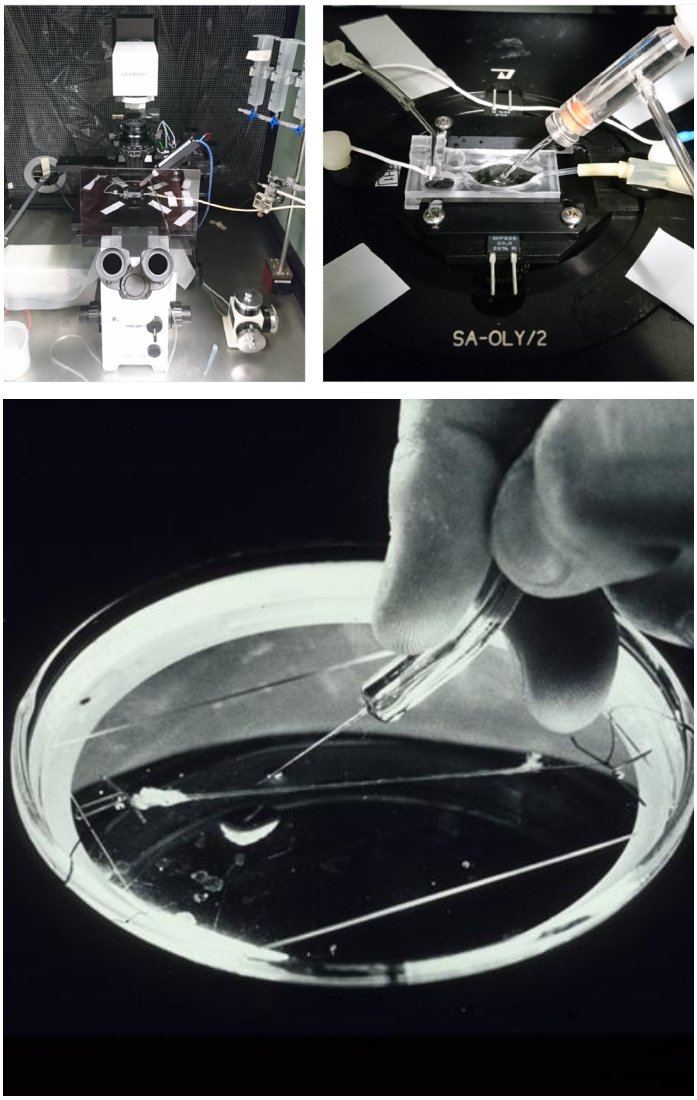


Figure 6.26 An electrophysiology rig is the setup used for recording and controlling the electrical activity of neurons (top). The isolated squid giant axon (bottom) was one of the earliest preparations that informed us about the electrical properties of neurons.

Electrophysiology on a cellular level usually requires the concurrent use of a microscope to get resolution down at the level of the neuron, which may be on the order of tens of microns in diameter. At this level of magnification, tiny glass micropipettes filled with an electrolyte solution can then be inserted into neurons, and electrical currents can then be detected and recorded. These currents can be the currents associated with the movement of charged ions, such as sodium during an action potential.

Electrophysiology can also be used to manipulate the activity of neurons. Instead of simply detecting currents, electrophysiology gives you direct physical control over the electrical properties of the neuron, which gives us the ability to manipulate different components of the nervous system.

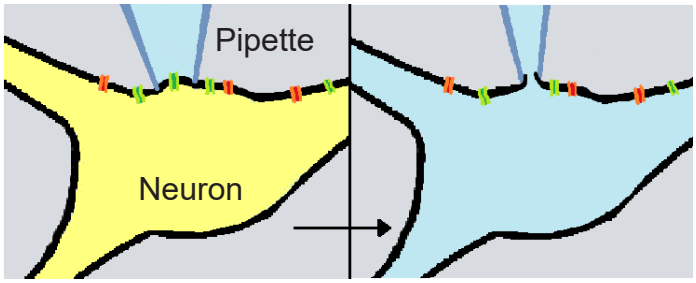


Figure 6.27 Rupturing the cell membrane with a microscopic glass pipette allows the experimenter to monitor and control the electrical conditions of the neuron.

Electrophysiology can be a very versatile technique. Because there are many strategies that count as electrophysiology, you can ask questions at many levels, from behavior and neuronal circuitry to individual neurons and ion channels. With electrophysiology, you can record the patterns of neuronal activity as a rat performs on a behavioral task, and stimulate neurons to see if you can modify the behavior. On the other extreme, you can examine the activity of individual ion channels, to determine the conditions when

they open and close.

One weakness lies in the highly experimental nature of electrophysiology. There are many different preparations that can be used in studying the electrical activity of the nervous system: intact anesthetized, thin slice of living brain, neuronal culture, or even frog oocytes that express neuronal receptors. But, each step you move away from the awake and behaving animal, the more difficult it becomes to draw inferences based on our results. What we gain in control over the experimental preparation, we lose in ability to generalize about data.

6.4.2 Transcranial magnetic stimulation (TMS)

Example questions answered:
"Which part of the motor cortex sends signals to the contralateral foot?"
"Can activation of parts of the prefrontal cortex decrease the symptoms of depression?"

In physics, there is a close relationship between electricity and magnetism. The movement of magnets can generate electric currents, and conversely, electricity can generate magnetic fields, a process called **induction**. A brain stimulation technique called transcranial magnetic stimulation (TMS) relies on induction to generate electrical currents through the skull. The TMS machine itself is a handheld coil of wires in a loop. Passing electrical current through that loop produces a magnetic field. The magnetic field generated then induces an electrical current at some distance away from the coil. By placing the coil at the surface of the scalp, we can electrically activate small areas of brain tissue. For instance, using TMS above the motor cortex causes muscle contractions, and using TMS above the occipital lobe causes the perception of flashes of light.

TMS is believed to have moderate benefits for several different psychiatric conditions, potentially making it a meaningful therapeutic intervention. Activation of motor cortex may help alleviate chronic pain conditions, Parkinsonian symptoms, and improve contralateral motor function following stroke injury. Activation of prefrontal cortex can decrease anxiety, antidepressive symptoms, cigarette craving or consumption, and schizophrenia. Stimulation of

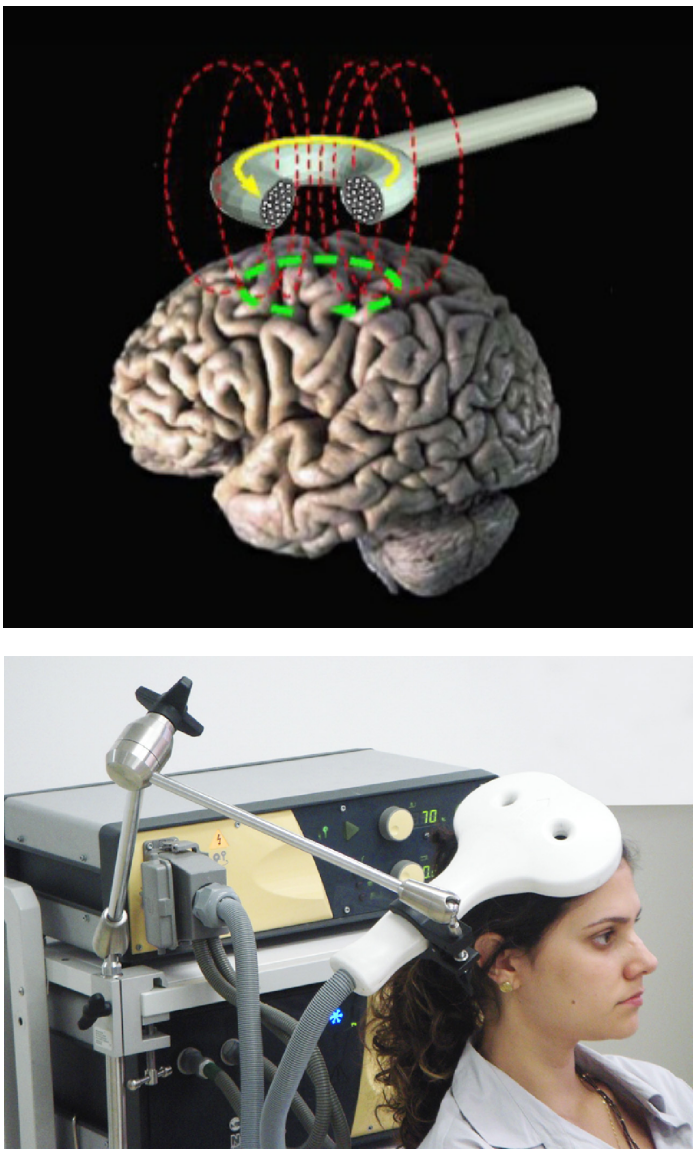


Figure 6.28 Running current through a magenetic coil generates an electric field via induction (top), which can activate or inactivate neural circuits in TMS (bottom).

Introduction to Computational Neuroscience

Computational neuroscience is a branch of neuroscience that focuses on explaining neural mechanisms through computational methods. This multi-disciplinary field draws from other fields such as electrical engineering, physics, and computer science to better understand how the nervous system processes information. Computational neuroscience strives to answer a wide range of questions in neuroscience research, including computation in single neurons, computation in a neural network, neuron development, sensory processing, motor control, memory and plasticity, as well as broader questions about vision, consciousness, and cognition. Sometimes, using computational approaches can be described as being an ***in silico*** preparation.

Computational neuroscience provides tools and methods for developing models, which are used for answering different categories of questions.

1. Descriptive models characterize what nervous systems do. These models qualitatively describe data through neural encoding as well as extract information through neural decoding, the reconstruction of information that has already been encoded and represented in the brain by a network of neurons. Reconstruction is the researcher's ability to predict sensory stimuli purely from the action potentials from neurons. The objective of neural decoding is to characterize the electrical activity of neurons in relation to responses in the brain. Neural encoding characterizes the relationship between the stimulus and the neuronal responses and the relationship with electrical activity.

2. Mechanistic models are used to determine how nervous systems relates with known anatomy and physiology. These models simulate the behavior of a single neuron or network of neurons.

3. Interpretive models lead to an understanding of why the nervous system operates in a particular way. These models are used to examine neural cognition and the underlying nervous system functionality.

There is debate with how computational models can be used to best reflect the nervous system. A model that focuses largely on defining biological processes into mathematical terms may stray from its intended purpose and not accurately reflect neural behavior. In other cases, a mathematical model may oversimplify neural behavior and may lead to false conclusions. Thus, there is a focus on biologically plausible neural networks to best model the nervous system.

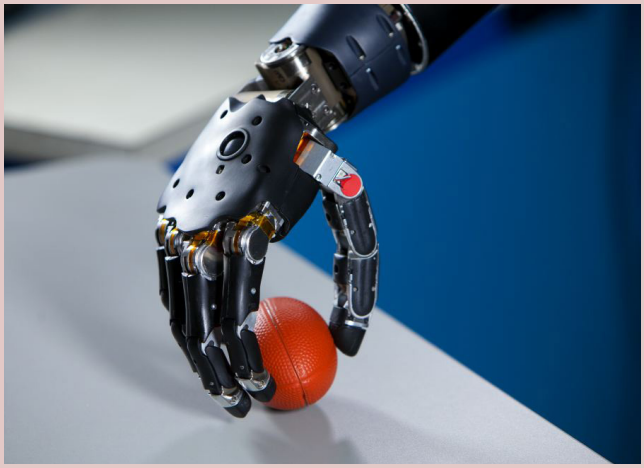


Figure 6.29 Advances in computational neuroscience have led to the development of robotic protheses that read neural signals, and predict the corresponding motor output.

other brain regions can be antiepileptic, and may also minimize auditory hallucinations or tinnitus (ringing in the ears.)

The technique itself is completely noninvasive. TMS can be delivered simply with placement of the electric coil on the surface of the head.

TMS may have some unexpected consequences, such as temporary headaches, localized pain, changes in hearing, and bizarre changes in somatosensation. The most severe side effect so far recorded is seizures, which can be very dangerous. Use of TMS is still highly experimental.

As with fMRI, the magnetic fields generated by a TMS device can cause dangerous interactions with any magnetosensitive implants, such as deep brain stimulation devices, cochlear implants, or aneurysm clips.

6.4.3 Genetic modification

Example questions answered:
"Do rats behave differently if they do not synthesize the hormone leptin?"
"How do mutated voltage-gated sodium channels change their neuronal action potential firing?"

Since the rise of molecular genetics techniques in the 1970s, researchers have discovered some of the underlying genetic commonalities between healthy people and those with neurological or psychiatric conditions. The genomes of many animals have been completely mapped out including the nematode C. elegans (1998), drosophila (2000), mouse (2002), human (2003), rat (2004), and the macaque monkey (2007).

Knowledge of their genomes, and the capability to manipulate their genomes, has allowed researchers to create animal models of a variety of human conditions. For example, we can use gene excision to remove a small section of the genetic code, such as the part that codes for a certain protein. This strategy is called **knock-out**, and is helpful for identifying the function of a genetic sequence. Alternatively, we can use gene insertion to put in exogenous genes into an animal. For example, we can generate mice that express the human version of a gene. We call these humanized mice. This type of modification is called a **knock-in**. There is an entire spectrum of genetic modifications that can be performed, including knock-downs (moderate decrease in function), upregulations (some increase in function), or conditional knock-outs (an organism that is normal until exposed to certain chemicals).

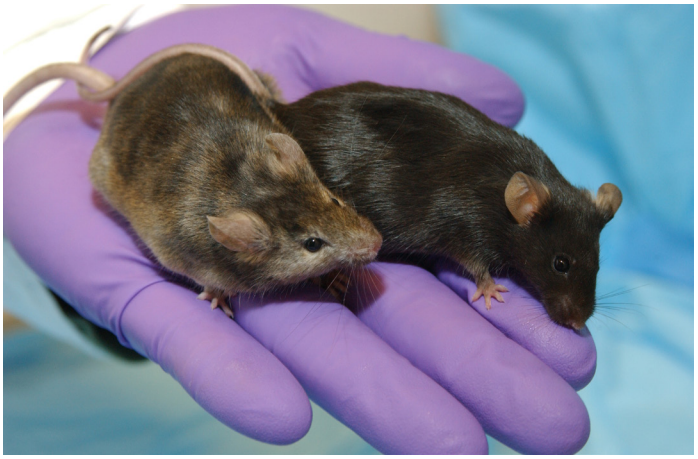


Figure 6.30 A mouse with a hair-related gene knocked out (left) compared to a wild type mouse (right).

A recent advance in genetic modification technology published in 2012 called the **CRISPR-Cas9** system allows for targeted editing of the genome. The main advantage provided by CRISPR-Cas9 is the simplicity, efficiency, and precision of the method. While still highly experimental, there are potential therapeutic applications for CRISPR-Cas9 in treating a variety of genetic diseases.

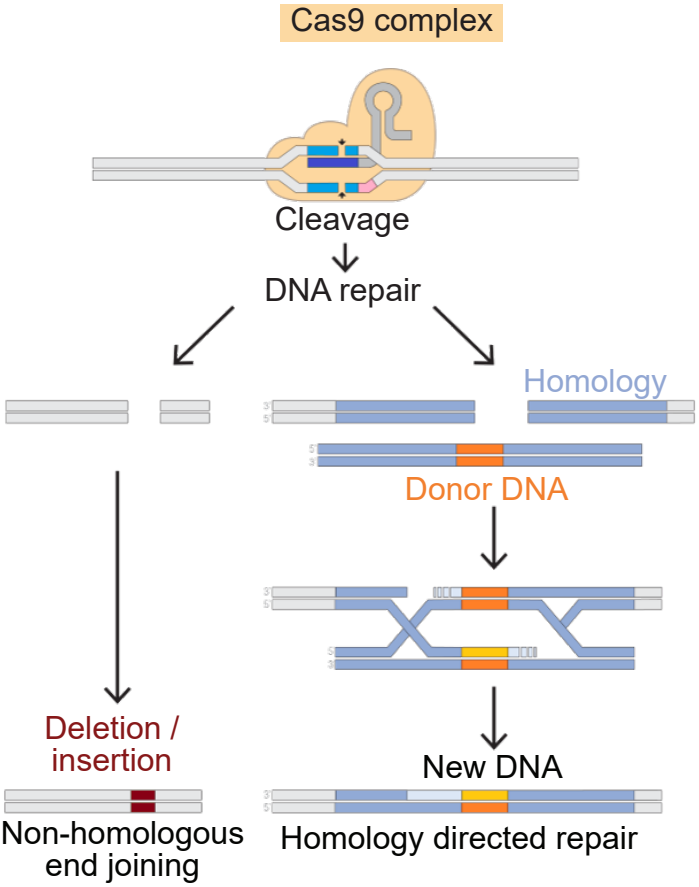


Figure 6.31 The CRISPR-Cas9 system is a powerful method of gene editing.

With the combined help of dedicated laboratories across the world, a wide variety of genetically-modified mice are commercially available for research purposes. Many scientific advances have come from these research models. It's also possible to develop an animal with multiple genetics crosses, which allow for several complex questions to be answered.

A shortcoming of the information gained from these genetic modification techniques is the difficulty in generalizing the findings beyond the specific genetic strain. For example, consider a knock-in approach to model a brain disorder in mice. If a therapy is developed that helps these animals, it is certainly a starting place that may lead to a human therapy. But, there is no certainty that the therapy will also be effective in humans.

Genetic modifications may also produce unexpected side effects as a result of changes to the genetic code. It is difficult to predict how a specific gene manipulation will interact with other aspects of the animal's physiology.

6.4.4 Optogenetics / Chemogenetics

Example questions answered:

"How does activating cortical glutamatergic drive affect firing rate of neurons in the striatum?"
"Does inactivation of GABA-ergic circuits in the amygdala change anxiety behaviors?"

One of the problems with early neural stimulation strategies was a lack of specificity. For example, many brain regions are home to a hodgepodge of cell populations that produce and release different neurotransmitters. In activating these areas, it was impossible to know if the effect of stimulation was due to activity of a specific population of neurons. Additionally, we were unsure if the results of electrical stimulation were due to the activity of that specific brain area where the stimulator was placed, or if it was due to the activation of axonal fibers that simply pass through that area. This was described as the **fibers of passage problem**.

The following two strategies, optogenetics and chemogenetics, were recent advances in neuroscience techniques that increase the specificity of brain stimulation. As of now, both of these techniques are exclusively used in nonhumans for research purposes.

Optogenetics

A revolutionary strategy developed in 2006 by researchers from Stanford University, optogenetics use light sensitive-ion channel proteins called **channelrhodopsin**, or **ChR2**. The light-sensitive component of ChR2 is a protein that is derived from the rhodopsin molecule, the same proteins that allow photoreceptors in the visual system to detect light. The ion channel component is a cation-permeable transmembrane pore that can allow Na+ ions to move across the cell membrane. When a photon of blue light hits the ChR2 molecule, the protein physically changes shape, causing the cation channel to open, enabling inward flow of positively charged sodium ions. Therefore, ChR2 gives us the ability to excite neurons with flashes of light.

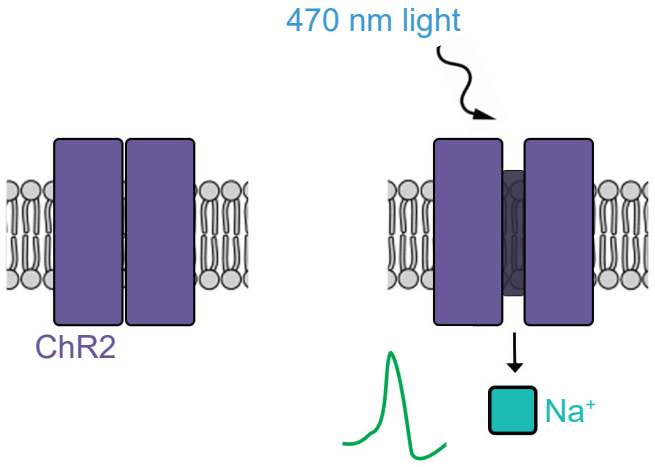


Figure 6.32 ChR2 is a transmembrane protein that is opened by blue light, which can cause action potential firing.

The ability for ChR2 to change structure in response to light is very fast, and the subsequent movement of changed ions down their electrochemical gradient is even faster. A brief 0.5 millisecond flash of blue light is enough to trigger sufficient inward Na+ current to induce a single action potential. Once the blue light is turned off, the channel closes in less than a millisecond. The speed with which ChR2-expressing neurons can respond gives optogenetic strategies a temporal resolution on the order of milliseconds.

ChR2 is possibly the most popular optogenetic molecule, but not the only one. Since its adoption into neuroscience, there have been several other proteins with similar light-sensitive characteristics. Halorhodopsin is a light-sensitive pump that uses energy from yellow-green wavelengths to actively push chloride into the cell, which inhibits neurons. Archaerhodopsin is a light-sensitive proton pump that extrudes H+ ions, which inhibits neurons.

ChR2 also has very high specificity, since optogenetics is used in conjunction with knock-in genetic modification technology or viral gene delivery to insert ChR2 into unique cell populations. This strategy allows researchers to isolate the effect of activation of a specific cell population within an area of the brain that has many different cell types.

Chemogenetics

Chemogenetics is an alternative strategy to selectively activate specific cell populations. Instead of using light to activate special receptors, these techniques use various chemical agonists. Chemogenetic receptors are usually G-protein coupled receptors that are activated only by exogenous drugs, but not by any endogenous neurotransmitters. The most well-characterized chemogenetic receptor is called **DREADD**, which stands for designer receptor exclusively activated by designer drugs. When the receptor is exposed to the designer drug, it can trigger the activation of the intracellular signaling pathway that can either excite or inhibit the neuron, depending on the nature of the DREADD. So, when a DREADD

protein is inserted into a neuron, there are no changes in neuronal activity at rest. But once the receptor is exposed to the endogenous drug, the DREADD is activated, which changes the activity of the neuron.

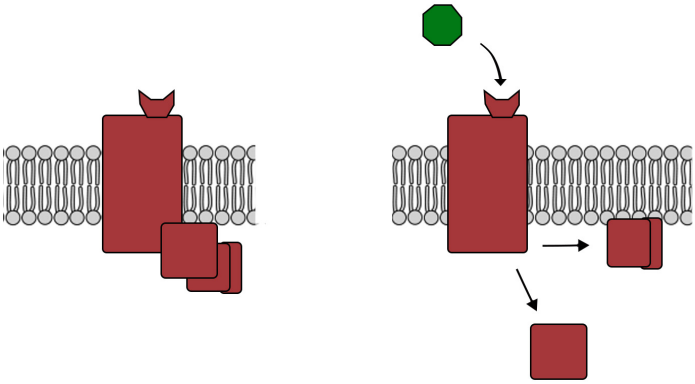


Figure 6.33 A chemogenetic system such as DREADD allows for control over intracellular signaling pathways upon exposure to an exogenous ligand.

Because chemogenetics requires the drug to enter the system and subsequent GPCR activation before a change in behavior can be observed, DREADD offers significantly less temporal resolution than optogenetics (in an ex vivo preparation - seconds to tens of seconds; in a behaving animal - minutes).

Optogenetics / chemogenetics can be used simultaneously with other techniques. Combining optogenetics with electrophysiology techniques has taught neuroscience about the nature of the connectivity between brain areas. Behavioral testing with simultaneous optogenetic or chemogenetic stimulation has given us insight into the roles that are played by specific neural circuits or neurotransmitters. These techniques are quickly advancing our understanding of the nervous system, and will continue to do so in the future.

The Open Neuroscience Initiative is funded by a grant from the Vincentian Endowment Fund of DePaul University.

Cover: https://upload.wikimedia.org/wikipedia/commons/f/fb/Brain_MRI.jpg
6.1: Image by Pete Linforth from Pixabay https://pixabay.com/photos/countryside-tree-landscape-sunlight-2175353/
6.2: https://pixabay.com/photos/ara-parrot-yellow-macaw-bird-3601194/
6.4: https://upload.wikimedia.org/wikipedia/commons/c/cc/Adult_Caenorhabditis_elegans.jpg
https://commons.wikimedia.org/wiki/File:Standing_female_Drosophila_melanogaster.jpg
https://commons.wikimedia.org/wiki/File:Mouse_white_background.jpg
https://pixabay.com/photos/rat-garden-out-curious-sweet-1379317/
https://pixabay.com/photos/animal-monkey-3892155/
https://pixabay.com/photos/personal-network-social-media-3139194/
Experimental preparations: https://commons.wikimedia.org/wiki/File:54986main_mouse_med.jpg
https://upload.wikimedia.org/wikipedia/commons/c/c7/Galvani-frogs-legs-electricity.jpg
https://upload.wikimedia.org/wikipedia/commons/0/05/Cell_culture.jpg
6.5: https://upload.wikimedia.org/wikipedia/commons/f/fe/M%C3%BCller-Lyer_illusion.svg
6.6: https://wellcomecollection.org/works/vyfjw978?wellcomeImagesUrl=/indexplus/image/L0028667.html
https://commons.wikimedia.org/wiki/File:1895-Dictionary-Phrenolog.png
6.7: https://commons.wikimedia.org/wiki/File:Medical_X-Ray_imaging_WYJ07_nevit.jpg
6.8: https://pixabay.com/photos/bethesda-naval-medical-center-80380/
6.9: https://commons.wikimedia.org/wiki/File:Ct-scan_of_the_brain_with_an_subdural_hematoma.jpg modified by Austin Lim
6.10:
https://upload.wikimedia.org/wikipedia/commons/0/09/The_diffusion_tensor_tractographies_of_neural_tracts_for_language_fnhum-07-00749-g001.png
6.11: https://pubs.niaaa.nih.gov/publications/arh27-2/146-152.htm modified by Austin Lim
6.12: Optimization of CLARITY for Clearing Whole-Brain and Other Intact Organs Jonathan R. Epp, Yosuke Niibori, Hwa-Lin (Liz) Hsiang, Valentina Mercaldo, Karl Deisseroth, Sheena A. Josselyn, Paul W. Frankland eNeuro 8 May 2015, 2 (3) ENEURO.0022-15.2015; DOI: 10.1523/ENEURO.0022-15.2015
6.13: https://upload.wikimedia.org/wikipedia/commons/b/bf/EEG_cap.jpg
https://upload.wikimedia.org/wikipedia/commons/f/f1/Gen-20s_SWK.png modified by Austin Lim
6.14: https://commons.wikimedia.org/wiki/File:Fluorodeoxyglucose_18-F_skeletal.svg
6.15: https://upload.wikimedia.org/wikipedia/commons/7/7b/PET_scan.jpg
https://upload.wikimedia.org/wikipedia/commons/2/2a/Pet_scan_of_brain_tumor.jpg modified by Austin Lim
6.17: https://upload.wikimedia.org/wikipedia/commons/2/2a/FMRI_BOLD_activation_in_an_emotional_Stroop_task.jpg modified by Austin Lim
6.18: https://upload.wikimedia.org/wikipedia/commons/e/e1/FluorescenceFilters_2008-09-28_cs.svg modified by Austin Lim
https://upload.wikimedia.org/wikipedia/commons/9/9e/DGCs_in_GFP.jpg
6.19: https://upload.wikimedia.org/wikipedia/commons/5/5e/Cerebellum_Cross_Section_Medulla_%2842040992452%29.jpg
https://upload.wikimedia.org/wikipedia/commons/6/6d/GolgiStainedPyramidalCell.jpg

6.21: https://upload.wikimedia.org/wikipedia/commons/9/9c/Cryostat_microtome.jpg
https://upload.wikimedia.org/wikipedia/commons/4/46/Leica_microtome_2.JPG
6.23: https://upload.wikimedia.org/wikipedia/commons/d/dc/PLoSBiol4.e126.Fig6fNeuron.jpg
https://upload.wikimedia.org/wikipedia/commons/0/09/FluorescentCells.jpg
6.25: https://upload.wikimedia.org/wikipedia/commons/c/c7/Galvani-frogs-legs-electricity.jpg
6.26:
https://upload.wikimedia.org/wikipedia/commons/1/13/Patch_rig.png
https://upload.wikimedia.org/wikipedia/commons/1/19/Giant_Axon_of_Squid_%2814356033761%29.jpg
6.27: https://upload.wikimedia.org/wikipedia/commons/9/9c/Whole_Cell_Patch_Clamp_Steps.png modified by Austin Lim
6.28: https://upload.wikimedia.org/wikipedia/commons/6/67/Transcranial_magnetic_stimulation.jpg
https://upload.wikimedia.org/wikipedia/commons/3/38/Neuro-ms.png
6.29: https://upload.wikimedia.org/wikipedia/commons/5/5f/Brain-Controlled_Prosthetic_Arm_2.jpg
6.30: https://upload.wikimedia.org/wikipedia/commons/e/e1/Knockout_Mice5006-300.jpg
6.31: https://upload.wikimedia.org/wikipedia/commons/0/04/DNA_Repair-colourfriendly.png Modified by Austin Lim

Chapter 7:

Sensation and Perception: The Visual System

Austin Lim, PhD (DePaul University)
Editor: Dana Simmons, PhD (University of Chicago)
Allison Morales, OD (Illinois College of Optometry)

At over 400 miles of explored passages, the Mammoth Cave system in Kentucky is the longest known cave system on Earth. During the cave tour, the guides momentarily extinguish all the lights. The void that follows is overwhelmingly eerie. Opening and closing your eyes makes no difference. You may have difficulty figuring out which way is up and down, and might even lose your balance when you tilt your head.

Humans are remarkably dependent on the visual system to gain information about our surroundings. Consider how tentatively you walk from the light switch to your bed right after turning off the lights!

The visual system is complex and consists of several interacting anatomical structures. Here, we will describe the process of how photons of light from our surroundings become signals that the brain turns into representations of our surroundings.

Sensation vs. perception

The following 3 chapters are focused on sensation and perception. By definition, the words "sensation" and "perception" are very similar. However, in psychology and neuroscience, the two have slightly different meanings. Both terms are related to processes of getting information from the environment. "**Sensation**" is the detection of stimuli, whereas "**perception**" is the interpretation of those stimuli. The quantity of stimuli that you *sense* far outnumbers what can be realistically *perceived*. For example, imagine you are at a loud party, having a conversation with someone. You sense so many things here: all the friends and strangers you see, the smells of sweat or food, and the sounds of music and various voices. But while speaking to your friend, all you perceive is the person in front of you while all the other stuff fades into the background.

Perception is also shaped by past memories. Consider the following scenario. You are at home, studying by yourself quietly, when you hear footsteps and the jingling of keys. This process is sensation. Based on the fact that you have a roommate, and you know they have a nightly routine of studying late at the library before returning home, you interpret those sounds as your roommate coming home. This process is perception.

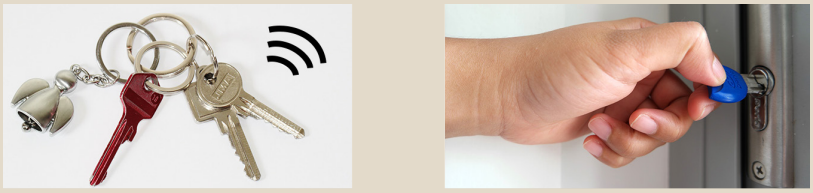


Figure 7.1 Sensation refers to the ability to detect stimuli, while perception is concerned with interpretation of those stimuli.

Chapter 7 outline

7.1 The Eye

Visual sensation starts at the level of the eye. The eye is an organ that has evolved to capture **photons**, the elementary particle of light. Photons are unusual because they behave as both particles and as waves, but neuroscientists mostly focus on the wave-like properties. Because photons travel as waves, they oscillate at different frequencies. The frequency at which a photon oscillates is directly related to the color that we perceive. The human visual system is capable of seeing light in a very narrow range of frequencies. On the short end, 400 nm wavelengths are observed as violet, while on the long end, 700 nm wavelengths are red. Ultraviolet light oscillates at

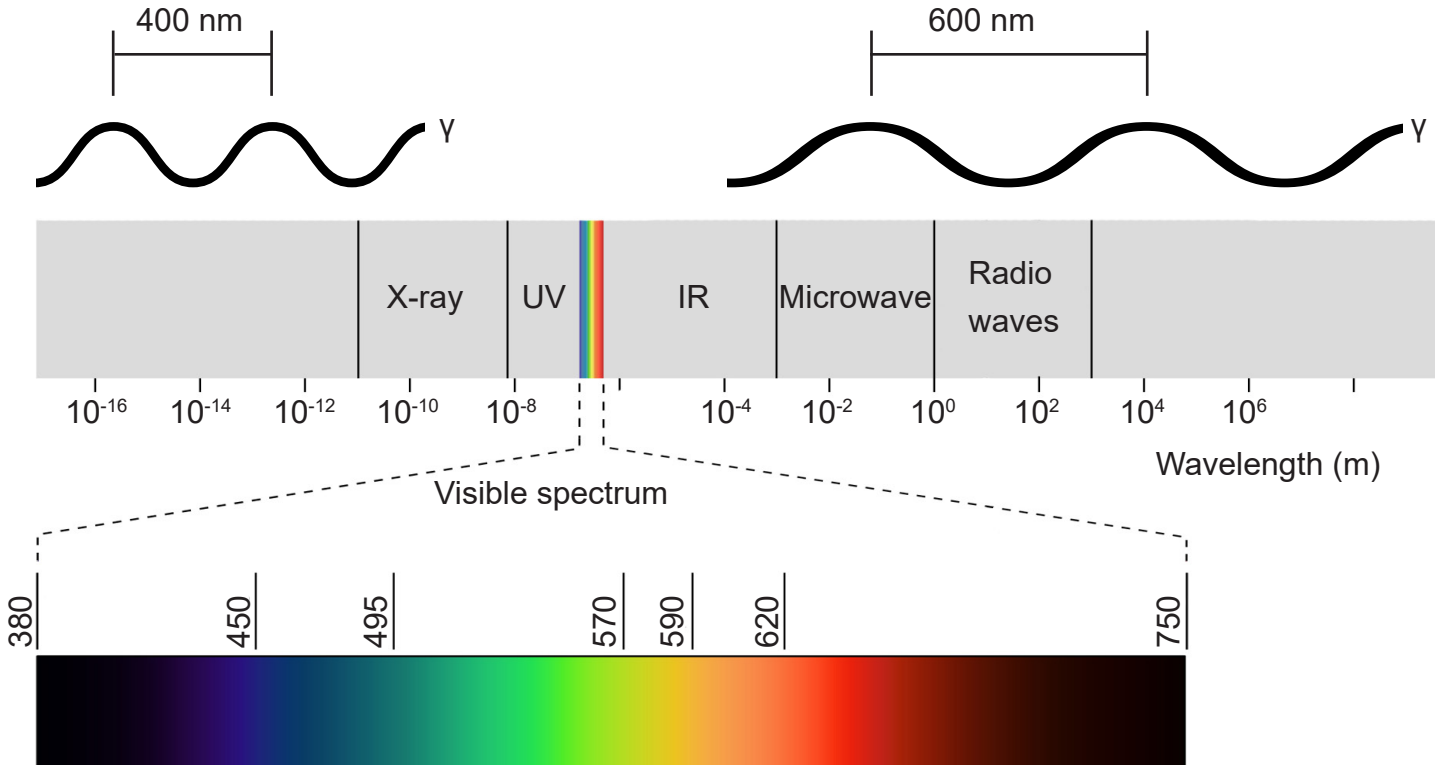


Figure 7.2 Photons (γ) oscillate at different frequencies. If the oscillation repeats every 400 nanometers, we perceive that color as violet (top left), but if it oscillates at 600 nm, we perceive those wavelengths as orange (top right). Humans are only able to detect a narrow range of particles within the electromagnetic spectrum, between 400 and 700 nanometers (bottom).

a wavelength shorter than 400 nm, while infrared light oscillates at a wavelength longer than 700 nm. Neither ultraviolet nor infrared light can be detected with our eyes.

For us to see light, three basic conditions need to be met. First photons need to be present. When there are no photons of light, as in the depths of the Mammoth caves, we are unable to detect anything.

Secondly, photons of light must reflect off objects in the surrounding world. If photons cannot bounce off the object, the eye cannot collect the photons. Black appears colorless since it absorbs more photons, while pure white reflects photons. The wavelengths that reflect off objects determine the color that we perceive.

Third, photons need to reach the back of the eyeball. Photons pass through several anatomical structures before the nervous system processes and interprets them.

The anterior, or front-most part of the eye, is the **cornea**. The cornea refracts, or bends, the incoming rays of light so that they converge precisely at the retina, the posterior most part of the eye. If the light rays fail to properly converge, a person would be near-sighted or far-sighted, and this would result in blurry vision. Glasses or contact lenses bend light before it reaches the cornea to compensate the cornea's shape.

After passing through the cornea, light enters through a hole in the center of the eye called the **pupil**. The diameter of the pupil can change depending on ambient light conditions. In the dark, the pupil dilates (also called mydriasis), or gets bigger, which allows the eye to capture more light. In bright conditions, the pupils constricts (also called miosis), or gets smaller, which decreases the amount of light that enters the eye. Two muscles, the pupillary dilator muscle and the pupillary sphincter muscle, are responsible for change in pupil size. The **pupillary dilator muscle** causes dilation and is downstream of the sympathetic branch of the autonomic nervous system, while the **pupillary sphincter** causes constriction and is activated by the parasympathetic nervous system.

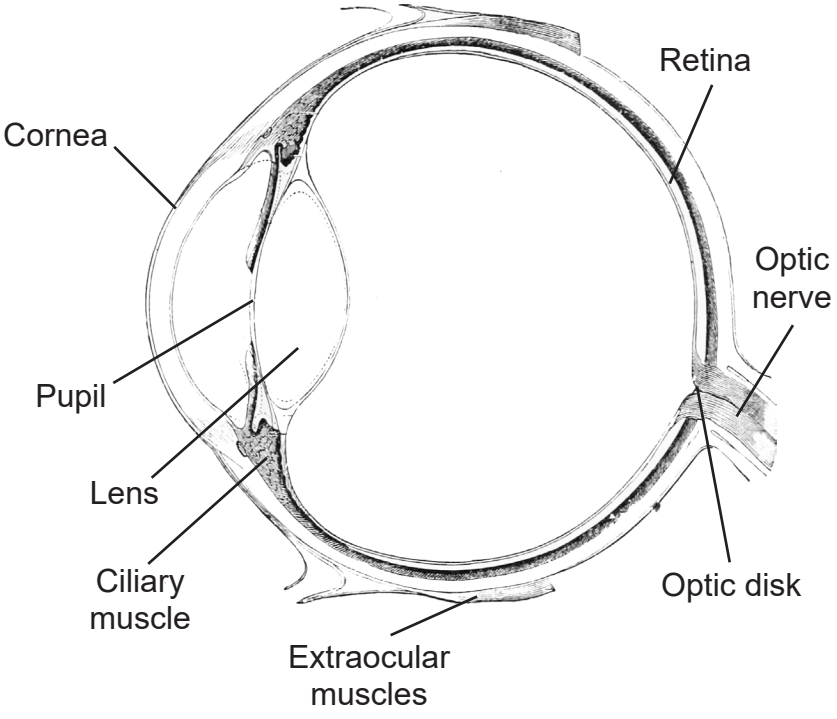


Figure 7.3 Anatomy of the eyeball.

The next structure that light passes through is the **lens**. Like the cornea, the lens refracts light so that the rays converge on the back of the eye. The lens of the eye is shaped like a camera lens or a magnifying lens, convex on both sides (biconvex). Unlike the cornea, the lens is capable of adjusting its shape, a process called **accommodation**. A circular muscle that surrounds the lens, called the **ciliary muscle**, changes the shape of the lens depending on the distance of the object of focus. Whenever you focus your eyes on something close, such as holding your hand a few inches

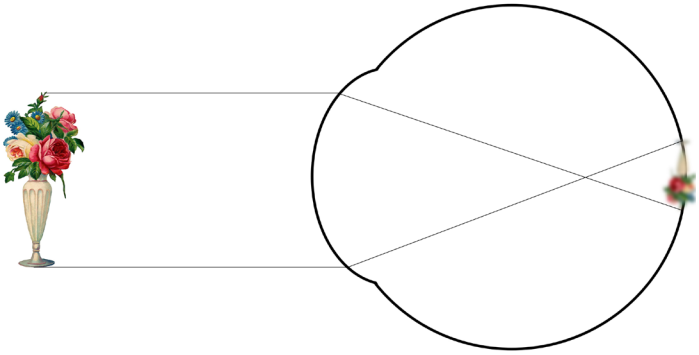


Figure 7.4 If the light refracts incorrectly, the image does not focus on the retina and our field of view will be blurry.

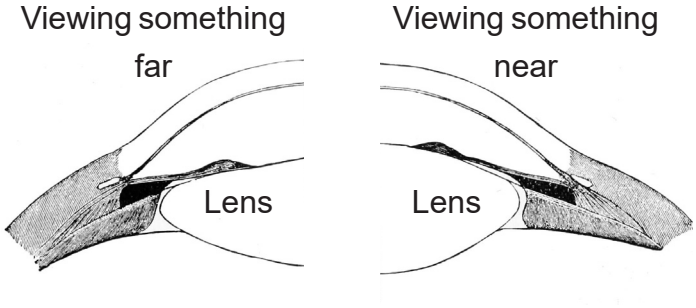


Figure 7.5 The lens is a biological tissue that flattens when we are focusing on something far away, and gets thicker when we focus up close. This process is called accommodation.

away from our face, the lens becomes thicker in the middle. Conversely, when you focus on the distant horizon, the lens flattens.

Another interesting consequence of the shape of the lens is the way it bends light. Think back to when you've held a magnifying lens in front of you as you look through it: The world you see through the lens is both upside down and reversed. This is exactly what the lens in our eyes does as it refracts light. As we look straight ahead, light from the upper part of our vision, like the ceiling, gets projected to the bottom of the eyeball, while light from the floor passes through the lens and refracts to the upper part of the eyeball. Additionally, light from the left part of our environment is refracted onto the right part of the eye, and vice versa.

We use the term **visual field** to describe the portion of our surroundings that can be seen without moving your eyes. If you look straight ahead, information on the left half of our vision are in our left visual field, which is projected onto the right half of both eyes. Conversely, objects in the right half of our vision are in our right visual field, which is projected as images onto the left half of both eyes.

There are six muscles that control eye movement. These muscles, plus the muscle that controls upper eyelid movement, are called the **extraocular muscles**. These seven muscles receive innervation from three cranial nerves, specifically cranial nerves 3, 4, and 6.

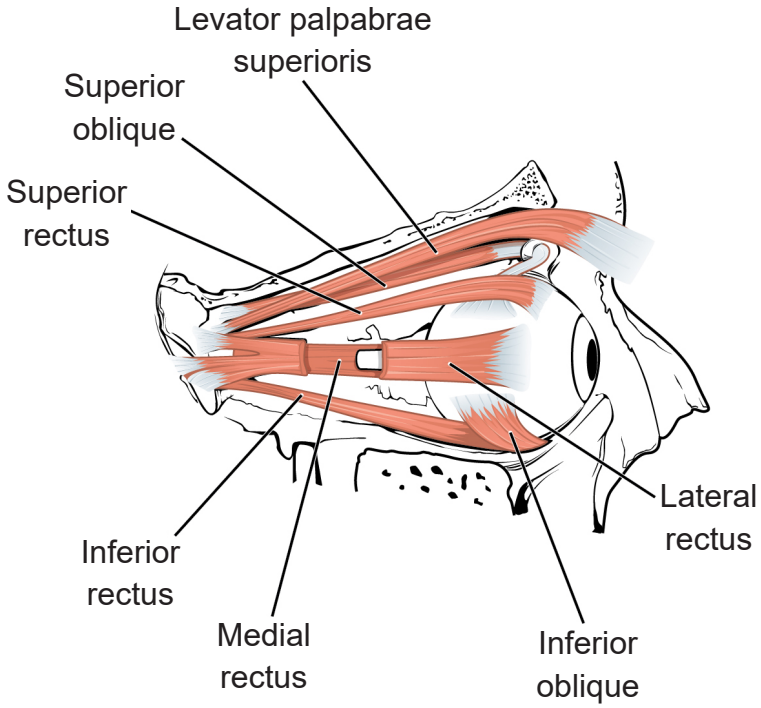


Figure 7.6 The extraocular muscles are a series of seven muslces, six of which control eyeball movement and one (leavtor palpabrae superioris)which controls the eyelid.

7.2 The Retina

The back of the eyeball is the **retina** (ret-n-UH). The medial half of the retina is also referred to as the **nasal hemiretina** since it is closer to the nose, while the lateral half of the retina is called the **temporal hemiretina** since it is closer to the temporal bone of the skull. The terms "upper hemiretina" and "lower hemiretina" are commonly used as well.

In the center of the retina is a small pit called the **fovea**. Light from only about 2 degrees from the center of the visual field gets focused onto the fovea. Compared to the rest of the retina, the fovea is where we have the highest acuity, or the clearest vision. As you are reading this text, light information from only 4 or 5 letters lands on the fovea.

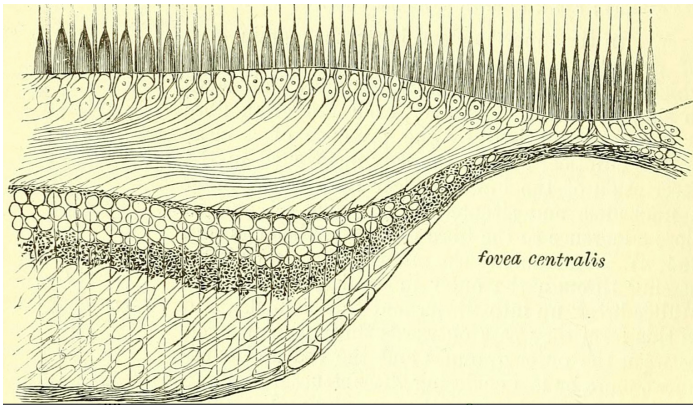


Figure 7.7 The fovea appears as a "pit" with the neuronal components swept away from the center of our vision.

The retina is the beginning of the nervous system's involvement with the visual system. It consists of five layers of neurons (described below). The pathway of information moves from the back of the eye towards the front, before exiting the back of the eye in an area called the **optic disk**. Visual information is passed directly through three of the five neuronal layers before being shuttled into the brain via the **optic nerve**. The other two layers of cells modify that route of communication. In the next paragraphs, we'll take an in-depth look at the role of each of the five layers of neurons in the retina.

In the order of the pathway of communication, the five neurons of the retina are:

7.2.1 Photoreceptor cells

Photoreceptors are the first cells in the neuronal visual perception pathway. They are the cells that detect photons of light and convert them into neurotransmitter release, a process called **phototransduction**.

Morphologically, photoreceptor cells have two parts, an outer segment and inner segment. The outer segment contains stacks of membranous disks bounded within the neuronal membrane. These membranous disks contain molecules called **photopigments**, which are the light-sensing components of the photoreceptors. Hundreds of billions of these photopigments can be found in a single photoreceptor cell.

The inner segment contains the nucleus and other organelles. Extending from the inner segment is the axon terminal. Photoreceptors are classified into two categories, named because of their appearance and shape: **rods** and **cones**.

Rod photoreceptor cells

Visual information from our peripheral vision is generally detected by our rod cells, which are most densely concentrated outside the fovea.

Rod cells are organized to have high **synaptic convergence**, where several rod cells (up to 30) feed into a single downstream route of

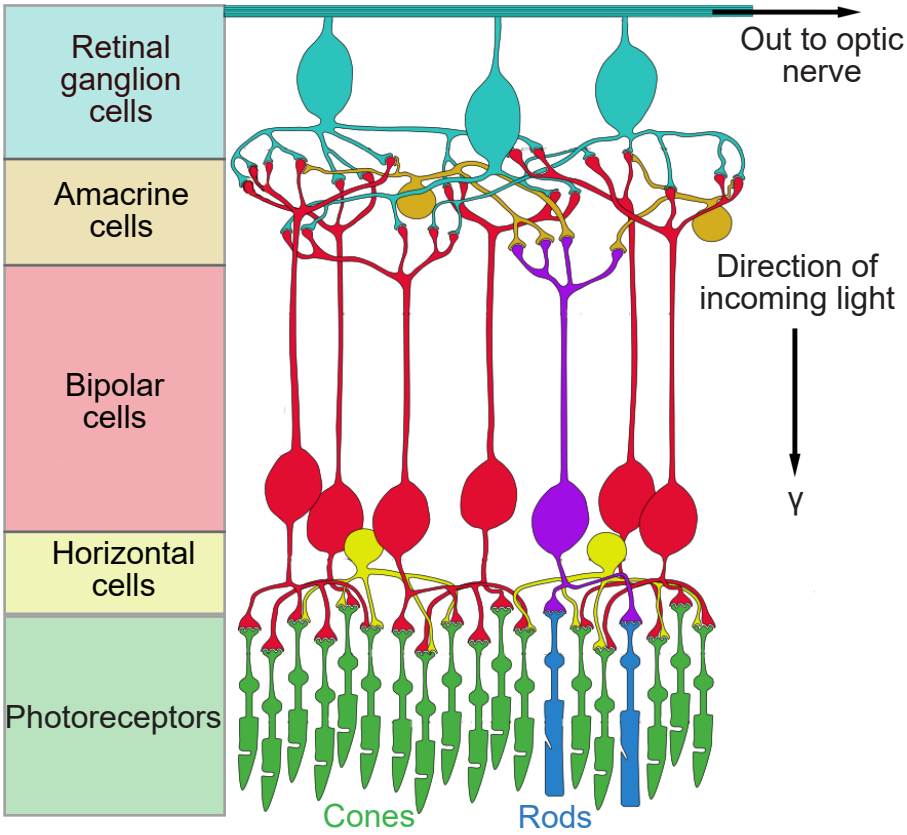


Figure 7.8 The five neuron layer of the retina.

communication (the bipolar cells, to be specific - see 7.2.3). An advantage of a high-convergence network is the ability to add many small signals together to create a seemingly larger signal. Consider stargazing at night, for example. Each rod is able to detect low levels of light, but signals from multiple rod cells, when summed together, allows you to recognize faint light sources as a star. A disadvantage of this type of organization is that it is difficult to identify exactly which photoreceptor is activated by the incoming light, which is why accuracy is poor when seeing stimuli in our peripheral vision. This is one of the reasons that we cannot actually read text in our peripheral vision or see the distinct edges of a star.

Rod cells are most sensitive to light that has a wavelength of 500 nm (blue-ish green). Light at other wavelengths still causes changes in the way rod cells respond, but to a lesser degree. Rod photoreceptors are maximally active in low-light conditions, which is why our surroundings appear to have a blue-ish tint at night. This phenomenon is known as the **Purkinje shift**, named after the Czech anatomist Jan Purkinje who observed that his favorite brightly-colored flowers appeared dull and more blue as the sun was setting.

Cone photoreceptor cells

Cone photoreceptor cells allow for high-acuity vision. They are most densely packed at the fovea, corresponding to the very center of your visual field. Despite being the cell population that we use for our best vision, cone cells make up the minority of

Figure 7.9 Demonstration of the Purkinje shift. At night, normal bright colors appear blue-ish due to increased activity of the rod photoreceptors.

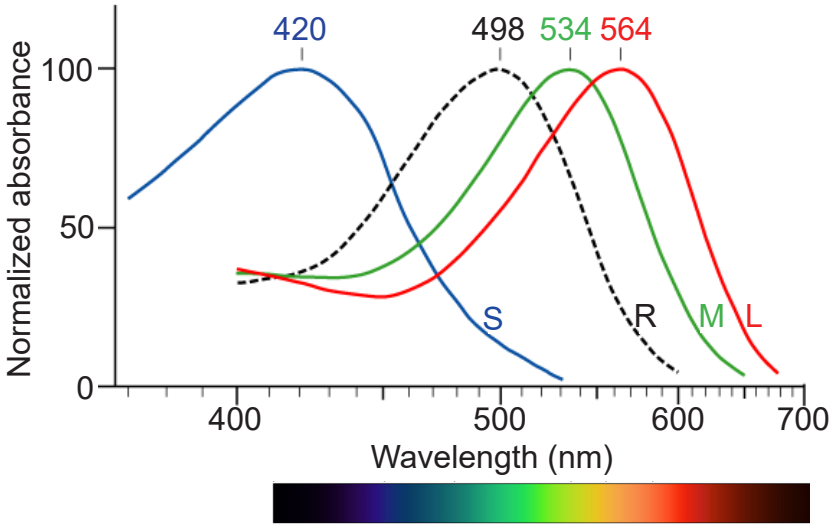


Figure 7.10 Different photoreceptor cells respond to different frequencies of light.

photoreceptors in the human retina, outnumbered by about 20-times more rod cells.

Unlike rod cells, cone cells have very low synaptic convergence. In fact, at the point of highest visual acuity, a single cone photoreceptor communicates with a single pathway to the brain. The signaling from low-convergence networks is not additive, so they are less effective at low light conditions. However, because of this low-convergence organization, cone cells are highly effective at precisely identifying the location of incoming light.

Cone photoreceptors are responsible for processing our sensation of color (the easiest way to remember this is cones = color). The typical human has three different types of cone photoreceptors cells, with each of these three types tuned to specific wavelengths of light. The short wavelength cones (S-cones) respond most robustly to 420 nm violet light. The middle wavelength cones (M-cones) exhibit peak responding at 530 nm green light, and the long wavelength cones (L-cones) are most responsive in 560 nm red light. Each of these cones is activated by other wavelengths of light too, but to a lesser degree. Every color on the visible spectrum is represented by some combination of activity of these three cone photoreceptors.

The idea that we have two different cellular populations and circuits that are used in visual perception is called the **duplicity theory of vision**, and is our current understanding of how the visual system perceives light. It suggests that both the rods and cones are used simultaneously and complement each other. One component, called the **photopic vision** system, uses cone photoreceptors of the retina, and is responsible for high-acuity sight and color vision in daytime. Its counterpart, called **scotopic vision**, uses rod photoreceptors and is best for seeing in low-light conditions, such as at night.

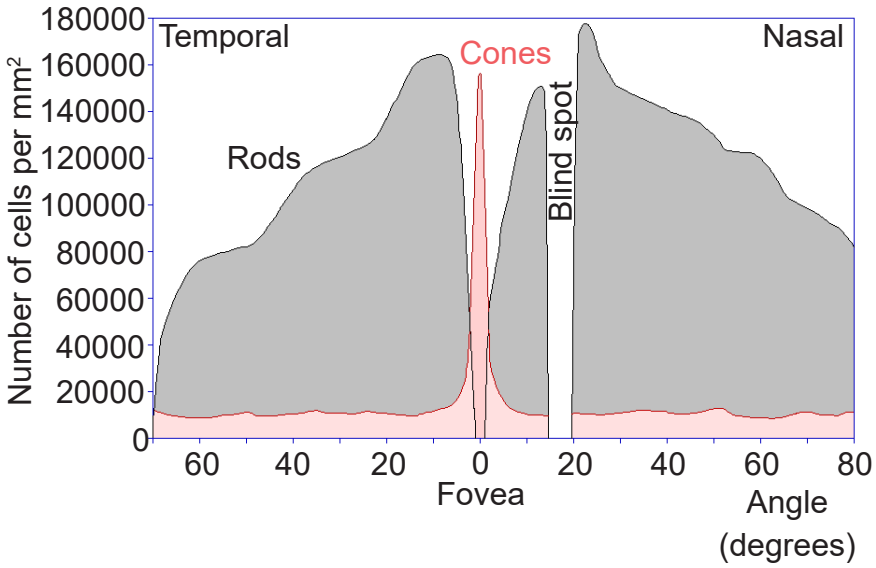


Figure 7.11 The cone photoreceptors are dense in the fovea, while the rod photoreceptors are mostly in the periphery.

Clinical connection: Color vision deficiency

Color vision deficiency is a very common condition resulting from a dysfunction in cone photoreceptor cells. As you know, each person has three types of cone cells. However, if one of the three types of cone cells fails to respond to the correct wavelengths of light, the person will lose the ability to differentiate between certain colors on the visible light spectrum. For example, lacking functional L-cones causes a person to have difficulty distinguishing purple from blue.

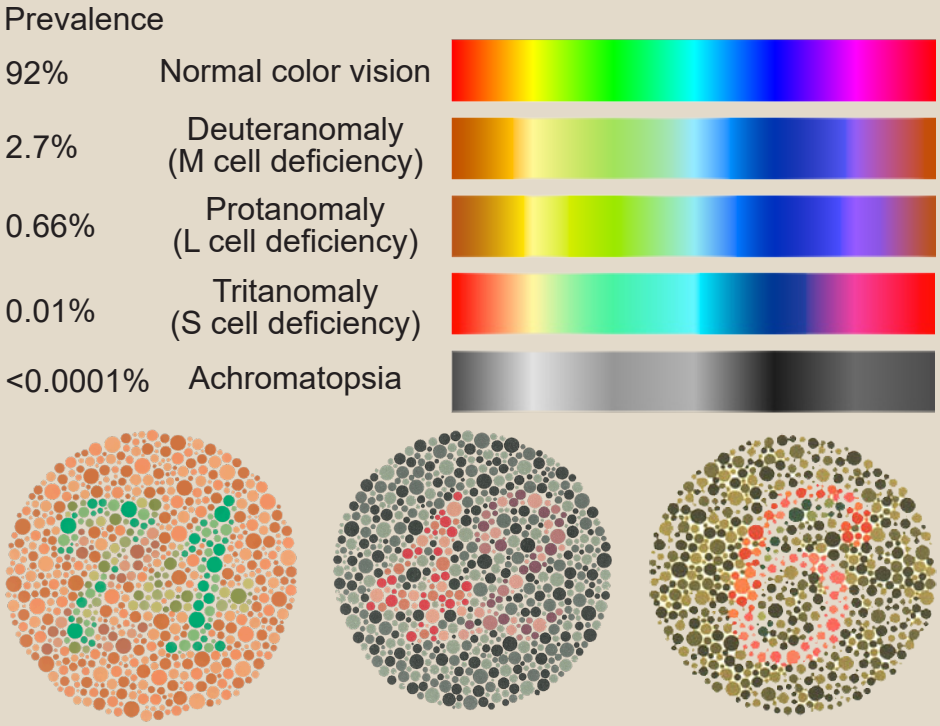


Figure 7.12 Deficits of the the different photoreceptors lead to seeing the visible spectrum in atypical ways (top). Color vision deficiency can be assessed with Ishihara plates, a series of images. An inability to see the 74 (bottom left), the 42 (bottom middle), or a 6 (bottom right) may indicate some form of color vision deficiency.

Color vision deficiency can either be acquired (as a symptom of disease or as a side effect of exposure to certain chemicals) or inherited. Most cases are inherited, and the genes associated with color vision deficiency are often recessive and located on the X-chromosome. Because these genes are sex-linked, color vision deficiency is 15 times more likely to occur in men than in women with a prevalence of about 8% of males and less than 1% in females.

On the molecular and cellular scale, the neuronal encoding of light information is counterintuitive. We normally think of neurons as being excited when they are activated, releasing more neurotransmitter upon experiencing a stimulus. Photoreceptors, on the contrary, decrease in excitability when they are exposed to light. Additionally, they do not fire action potentials. Instead, their membrane potential fluctuates, which determines the amount of glutamate release: depolarization causes more glutamate release, and hyperpolarization causes a decrease in glutamate release.

In the dark, photoreceptor cells have leak sodium channels that allow for extracellular sodium to enter those neurons, which is called the **dark current**. This current causes cellular depolarization. When a photon of light strikes the photoreceptor, the cell responds by decreasing its intracellular stores of a signaling molecule called **cGMP**, which in turn decreases the dark current. Light causes the neuron to

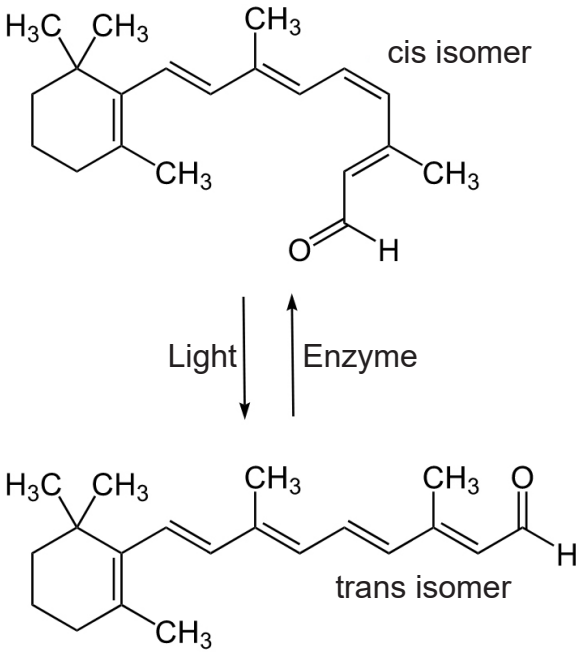


Figure 7.13 11-cis retinal can absorb a photon, causing retinal to enter the high energy state, 11-trans retinal.

become hyperpolarized, which inhibits neurotransmitter release.

These neurons sense light because of the presence of the photopigments found in the outer section of the photoreceptors. The photopigment molecule has two components. **Retinal** is a chemical synthesized from Vitamin A. In the dark, retinal exists as 11-cis-retinal, which has a low energy molecular configuration. Light can be highly energetic. When a photon hits retinal, it causes the chemical to change to all-trans-retinal, which is the high energy state configuration. While in the all-trans-retinal configuration, cGMP levels are decreased, therefore blocking the dark current, which causes hyperpolarization.

The other component is a protein called an **opsin**. Opsins exist in different forms, and they cause retinal to respond

Sidebar: Lateral inhibition

Bring your vision over to one of the white "intersections" in the Hermann grid below. In your peripheral vision, do you see faint fuzzy outlines of gray circles at the other white intersections?

This phenomenon happens because of the circuitry of the horizontal cells. In being activated, they block adjacent signaling pathways from also sending information forward. Because horizontal cells inhibit adjacent neurons at the same "level," we describe this phenomenon as **lateral inhibition**. It is hypothesized that lateral inhibition evolved to allow for better **edge detection**, the ability to easily differentiate the outlines of objects.

Now bring your focus to one of those fuzzy gray blobs. You will notice that the grayness disappears immediately when we look directly at it, but a new gray blob appears where you were just looking at. This happens because the size of the receptive field for each photoreceptor-to-bipolar synapse differs at the fovea from the periphery. Foveal photoreceptors have very little convergence onto their bipolar cells, so they have very high precision, and are therefore less susceptible to the illusion as seen above.

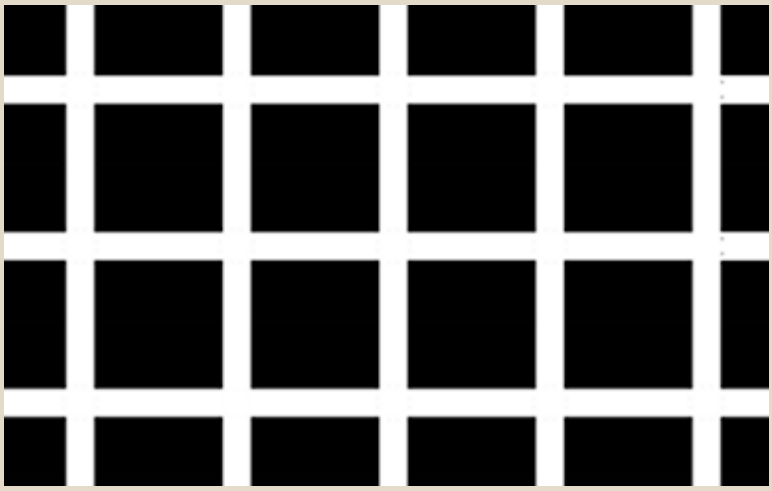


Figure 7.14 Gray blobs appear at the intersections of this Hermann grid because of lateral inhibition driven by horizontal cell activity.

differently depending on the wavelength of light: some opsins cause the photoreceptors to be maximally activated at 420 nm, others at 530 nm, and so on.

7.2.2 Horizontal cells

Horizontal cells are interneurons that do not directly participate in the pathway of signaling in visual perception. Rather, they communicate laterally, receiving synaptic inputs from photoreceptors, and forming synaptic connections with the axons of nearby photoreceptors. Like most interneurons in the nervous system, horizontal cells are inhibitory.

Photoreceptors release glutamate onto the horizontal cells, which causes the horizontal cells to depolarize. More glutamate is released when the photoreceptors are in the dark. In the light, the amount of glutamate released is decreased, which causes horizontal cells to hyperpolarize. This hyperpolarization then decreases neurotransmitter release onto nearby photoreceptor axons. Less inhibition onto these photoreceptors causes depolarization, which decreases signaling from these adjacent photoreceptors.

This complex pattern of neurotransmission can be summarized more simply: activation of a single photoreceptor causes inhibition of adjacent photoreceptors. The activity of horizontal cells helps us isolate the origin of a light signal by minimizing the noise around it.

7.2.3 Bipolar cells

Bipolar cells are the second type of neuron that is directly responsible for visual perception, receiving visual input signals from photoreceptor cells. They are named because of their morphology: from the cell body, they extend projections in two opposite directions. Their dendrites receive neurotransmitter signals from photoreceptors, and they send axonal projections to the next two layers of neurons in the retina (amacrine cells and retinal ganglion cells).

In mammals, bipolar neurons receive inputs exclusively from either rod or cone photoreceptors. In other words, there are no bipolar neurons that receive both rod and cone photoreceptor input.

As we've already discussed, photoreceptor cells release glutamate. At most synapses in the nervous system, glutamate is excitatory. However, at synapses between photoreceptors and bipolar cells, glutamate activates metabotropic mGluR6 receptors, which are inhibitory. In the light, there is a decrease in glutamate release from the photoreceptors, and this leads to an increase in bipolar cell activity.

7.2.4 Amacrine cells

Amacrine cells, like the horizontal cells, are interneurons that modify the communication pathway rather than directly being involved with light sensation. Amacrine cells receive inputs from bipolar cells, and in turn, modulate the activity of nearby neurons. Like most interneurons, amacrine cells synthesize and release the inhibitory neurotransmitter GABA.

Amacrine cells exist in several different shapes and sizes. They help us detect directional motion, aiding us in knowing when to transition between our scotopic or photopic vision systems. Some amacrine cells also perform circadian regulatory functions.

7.2.5 Retinal ganglion cells

Retinal ganglion cells are the third and last cell type that directly conveys visual sensory information, receiving inputs from the bipolar cells. The axons of the retinal ganglion cells

bundle together and form the optic nerve, which then exits the eyeball through the optic disk. These axons are unmyelinated before leaving the optic disk – if they were myelinated, the lipids that make up the myelin sheath would diffract light and cause imperfections in vision.

Unlike the two cells before them (the photoreceptors and bipolar cells), retinal ganglion cells communicate using action potentials.

Although most retinal ganglion cells receive light information through communication downstream of the photoreceptors and bipolar cells, a small population of retinal ganglion cells (1-2%) are photosensitive. These cells contain a molecule called **melanopsin**, a light-sensitive opsin molecule just like the photopsin or rhodopsin found in photoreceptors. These cells contribute significantly to our body's ability to adjust the circadian rhythms based on external light cues (see chapter 12 for more information.)

Sidebar: The blind spot

Take a look at the image below. Close your right eye and bring your face close to the page, focusing on the plus sign. Slowly move your head away from the page. At a distance of about one and a half feet away, the dot will disappear!

The dot disappears under these specific conditions because visual information from the dot falls directly onto the blind spot. The blind spot is about 15 degrees temporal from the fovea. We don't notice the blind spot in everyday life because usually both eyes are working simultaneously, so when something enters the blind spot of the right eye, it appears in the visual field of the left eye. The blind spot only comes into effect when using monocular vision, or when we only see with one eye. The blind spot is elliptical, and you can demonstrate this using the drawing above: Bring the dot into your blind spot as before, but this time, rotate the page, pivoting around the plus sign. The dot reappears as it moves above or below the blind spot.

Now, repeat with the figure below. Hold the page at the same distance where you noticed the blind spot. How would you describe what happens to the break in the line? While we are unable to physically detect any light that falls onto the blind spot, our brains fill in the gap in sensation by "filling in" the blind spot with the adjacent visual stimuli.

Figure 7.15 The blind spot can be observed in monocular vision with stimuli like these.

The retina is not completely uniform across the entire back of the eye. There are a few spots of particular interest along the retina where the cellular morphology is different: the **fovea** and the **optic disk**.

There are two cellular differences that explain why the fovea is the site of our best visual acuity. For one, the neurons found at the fovea are “swept” away from the center, which explains why the fovea looks like a pit. Cell membranes are made up mostly of lipids, which distort the passage of light. Because there are fewer cell bodies present here, the photons of light that reach the fovea are not refracted by the presence of other neurons.

Secondly, the distribution of photoreceptors at the fovea heavily leans toward cone type photoreceptors. Because the cone cells at the fovea exhibit no convergence, they are most accurately able to pinpoint the exact location of incoming light. On the other hand, most of the photoreceptors in the periphery are rod cells. With their high-convergence circuitry, the periphery of the retina is suited for detecting small amounts of light, though location and detail information is reduced.

Another anatomically interesting area of the retina is an elliptical spot (1.9mm vertical by 1.75 mm horizontal) called the optic disk. This is where the optic nerve exits the eye. At this part of the retina, there is an absence of photoreceptor cells. Because of this, we are unable to perceive light that falls onto the optic disk. This spot in our vision is called the **blind spot**.

7.3 The Optic Nerve

The **optic nerve**, or **cranial nerve II**, exits the posterior end of the eyeball, and travels posteriorly along the ventral surface of the brain. Like all other cranial nerves, the optic nerve is paired, meaning there is one on each half of the body - one for each eye. Both optic nerves merge at a spot called the **optic chiasm**, then diverge yet again as they travel posteriorly towards the thalamus.

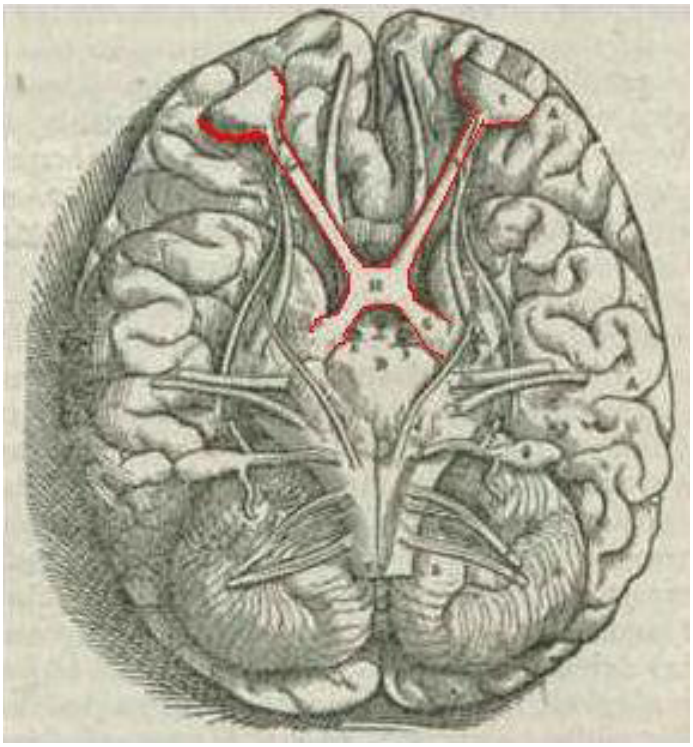


Figure 7.16 Cranial nerve II, or the optic nerve (highlighted in red), exits each of the eyeballs and meet at the optic chiasm before diverging.

The axonal connections in the optic nerve are not quite as simple as its anatomical appearance. From each optic nerve, some of the nerve fibers cross the midline, headed towards the contralateral hemisphere. Other nerve fibers meet at the optic chiasm, but then project into the ipsilateral hemisphere. Visual information from the nasal hemiretina projects contralaterally, so the left nasal hemiretina crosses at the optic chiasm to project into the right hemisphere of the brain. On the other hand, visual information from the temporal hemiretina projects ipsilaterally (left temporal hemiretina projects into left hemisphere, and right temporal hemiretina projects into right hemisphere).

The easy way to keep track of this unusual system is to remember that all information from the left visual field enters the right hemisphere of the brain, while visual information from the right visual field enters the left hemisphere of the brain. Because of the light-refracting properties of the lens, objects in the left visual field get represented onto the right half of each eyeball.

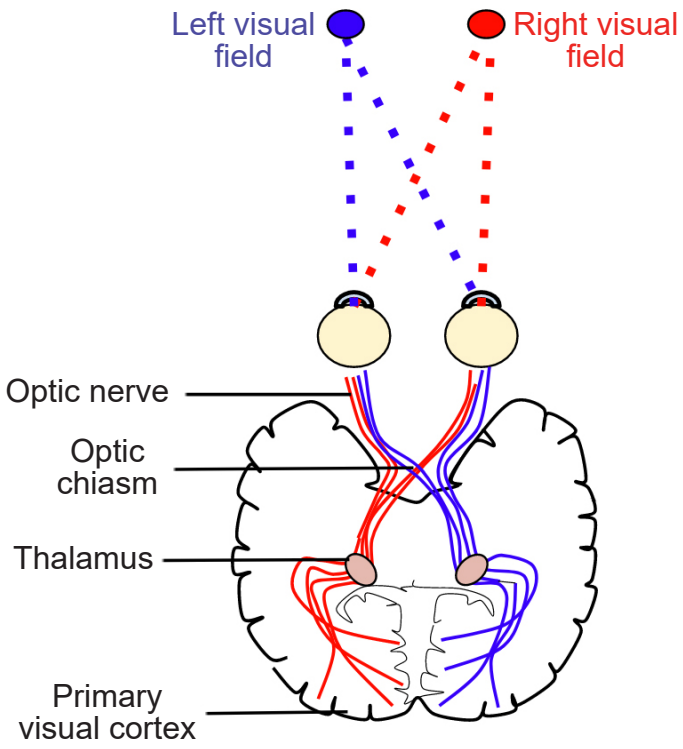


Figure 7.17 Visual information from the left visual field (blue) is carried into the right hemisphere of the brain, and visual information from the right visual field (red) is carried into the left hemisphere.

The right half of the left eye is the nasal side, so this information crosses contralaterally. The right half of the right eye is the temporal side, so this stays ipsilateral. This information then passes towards the thalamus for the beginning of central processing.

Not all of the axons convey direct visual information into the thalamus for visual perception. One pathway of the optic nerve sends information into a midbrain structure called the **pretectal area**. The information sent into the pretectum is related to some of the more "primitive" visual functions. For example, the ability for the eyes to follow something unconsciously as it moves across your field of view is called **smooth pursuit**, and this reflex is partly carried out by the optic nerve pathway into the pretectal area.

Another signaling pathway that branches from the optic nerve is the **retinohypothalamic tract (RHT).** It does not carry any conscious visual information. The RHT conducts light information from the intrinsically-photosensitive retinal ganglion cells. Instead of sending axonal projections towards the thalamus, this information pathway forms a synapse into a region of the hypothalamus called the **suprachiasmatic nucleus, or SCN**. This structure functions to help the body adapt its sleep-wake cycle in the face of changing day-night patterns (Chapter 12).

Clinical correlation: Glaucoma

The optic nerve is particularly sensitive to injury. A condition called **glaucoma** can cause optic nerve destruction, leading to blindness. Glaucoma affects up to 3 million Americans, but can be easily detected during an annual eye exam. Glaucoma can result from several different causes including high intraocular pressure, physical trauma, medications, or a natural genetic weakness of the nerve. High intraocular pressure is the greatest risk factor for glaucoma, since hypertension compresses the axons, which can also decrease the blood flow to the optic nerve. Almost all the first-line therapies to treat glaucoma decrease intraocular pressure.

7.4 Visual perception in the brain

The first synapse of the optic nerve is formed in the thalamus at a subregion called the **lateral geniculate nucleus, or LGN** (originating from the Latin *genu*-, meaning "knee"). The LGN is divided into six layers, named 1 through 6, going from ventral to dorsal. Histologically, the LGN is made up of three different populations of neurons. The magnocellular cells (*magn*- means large, as in "magnify" or "magnanimous"), or M cells, are large-diameter neurons (~20 μM) that are found in layers 1 and 2. The parvocellular cells (*parv*- means small), or P cells, are small-diameter (~10 μm) neurons found in layers 3 through 6. Ventral to each layer is a row of very small neurons, called koniocellular cells (*konio*- means dust), or K cells.

We know that the visual information from the temporal retina synapses ipsilaterally, while information from the nasal retina synapses contralaterally. The specific synaptic inputs into the LGN follow a specific organization depending on the origin of the retinal ganglion neurons. Temporal retinal ganglion neurons synapse only onto layers 2, 3, and 5 of the ipsilateral LGN, while nasal retinal ganglion neurons form synapses exclusively onto contralateral layers 1, 4, and 6.

The outputs of the LGN are a series of axonal bundles called the **optic radiations**. These optic radiations are divided into two main bundles, the upper and lower divisions. The upper divisions carry information from the lower visual field, whereas the lower divisions carry information from the upper visual field.

From the LGN of the thalamus, the optic radiations project to the occipital lobe at the caudal (posterior) end of the brain. Once visual information travels into the cortex, the process is

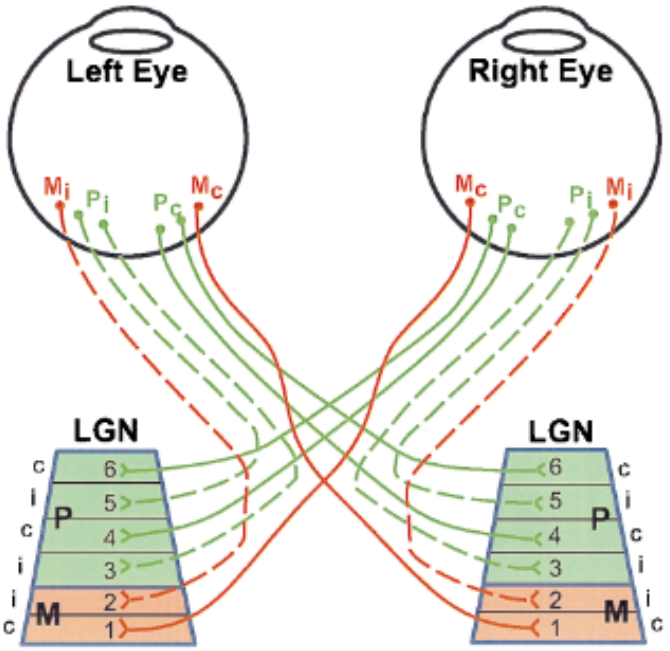


Figure 7.18 Synaptic circuitry showing the inputs into the layers of the lateral geniculate nucleus of the thalamus.

less about sensation and mostly about perception. For example, many of the post-optic radiation signaling pathways are concerned with questions like attention, object orientation, and recognition. Many visual perception functions are performed by multiple brain regions, and no single area is responsible for any one compartmentalized function. In other words, there is still much work to be done on understanding the precise function of the areas of the visual cortex.

The outputs of the LGN are axons which form synapses in the **primary visual cortex**, which is also called **V1**. Alternatively, V1 is also called the **striate cortex** because it has a large white stripe that can be seen in surgical dissection. This white stripe is the bundle of incoming optic tract axons, which are heavily myelinated.

Each neuron in V1 receives visual information from a specific patch of retinal cells.

Clinical connection: Anatomy of vision loss

Health care professionals can diagnose the location of an injury based on the deficits in a patient's visual field. To record the sites of vision loss, they may use a pair of circles which show which areas of the visual field are lost. For example, if a person is missing their right eye, all information from the right eye will be missing and the right circle will be colored in black.

If the visual information-conveying pathway is damaged, a person may suddenly lose half of their visual field resulting in a condition called **hemianopsia** (*hemi-* meaning half, *a-* meaning none, and *-opsia* referring to vision). There are a few forms of hemianopsia that correlate to specific injuries of the optic nerve. For example, injury to one branch of the nerve after the optic chiasm results in homonymous hemianopsia, the loss of either the left or right visual field of both eyes.

Injuries may damage some optic radiations while sparing the others. This results in a condition called **quadrantanopsia**, resulting in a loss of a fourth, or one quadrant, of a person's visual field. If the upper division of the right optic radiation is injured, the person will lose the upper left visual field of both of their eyes: left superior quadrantanopsia.

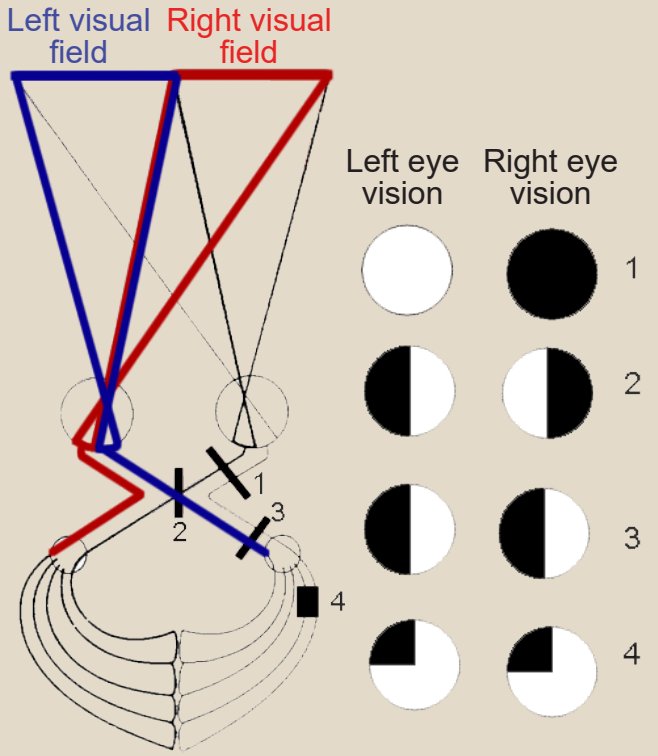


Figure 7.19 Injury to different parts of the optic tract (numbers) leads to specific deficits in vision (right).

This organizational pattern, where a section of retinal inputs map onto neurons of a specific section of V1, is called **retinotopic organization**. For example, the dorsal part of V1 receives information from the bottom half of the visual field. Visual information from the fovea, despite being only 1% of the total visual field, takes up about half of all neurons in V1.

After processing in V1, visual information is passed along to the **secondary visual cortex**, or **V2**. V2 represents the beginning of the visual association cortices. Following V2 are other visual cortical areas that contribute to various aspects of visual perception. For example, V3 and V5 help us comprehend motion, V4 contributes to color perception, and V6 helps with understanding our position in our surroundings.

Dual stream hypothesis

After V2, visual information passes through two streams of communication: the dorsal stream and the ventral stream.

The **dorsal stream** is described as the "where" pathway because these structures help us identify where objects are located in the space around us. The dorsal stream includes area V5,

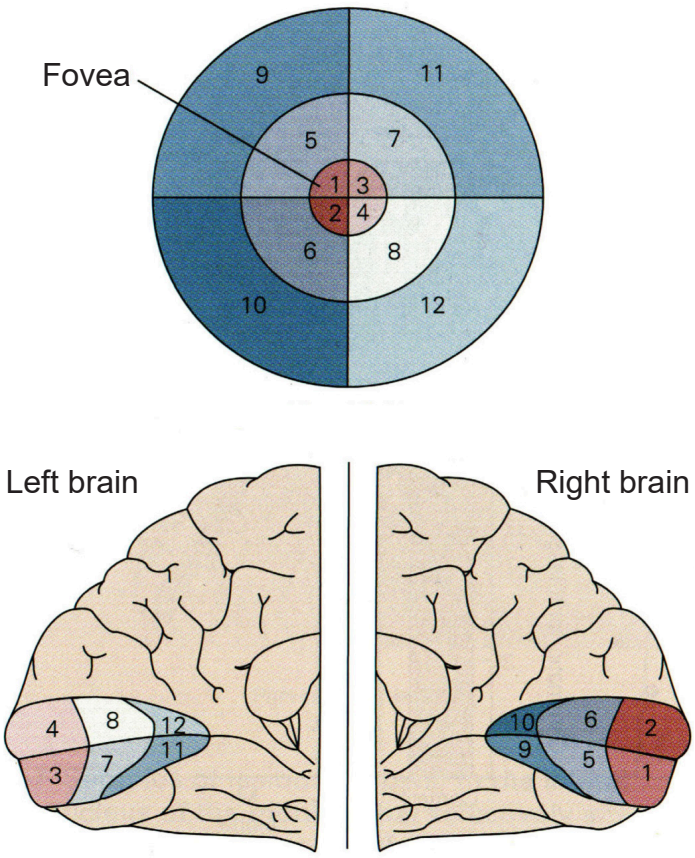


Figure 7.20 Retinotopic organization of the visual field. Note that adjacent areas of the visual field are represented in adjacent areas of the cortex, and that the visual information in the fovea is represented very heavily in the cortex.

which contributes to perception of motion. These structures also guide us when we move through our environments, contributing to our sense of spatial awareness. For example, a task such as reaching out to grab an object in front of you uses a combination of these features, so this task is guided largely by dorsal stream structures.

The **ventral stream** is the "what" pathway, and helps in the identification of objects that we see. Ventral stream structures are important for visual memory. The fusiform face area (FFA) and the parahippocampal place area (PPA) are especially important for the storage of facial stimuli and scenery or buildings, respectively. When the visual system senses these complex stimuli, those signals get processed through these ventral stream pathways. These incoming stimuli are compared with the memories stored in the ventral stream, and this comparison contributes to our capacity for perception and identification.

The two streams are not independent of each other. Rather, successful organisms require the melding of both components of visual perception. For example, imagine you are a prehistoric organism living in a food-scarce environment. Approaching a small berry tree, you would use ventral stream structures to correlate the berries with memories: Did these berries taste good and give me the calories I need to survive? Or did these berries make me violently ill, and are therefore probably poisonous? If they are the delicious berries that I want, I will use the dorsal stream structures to take note of their precise location so I can reach out for them and pick the berries . In this example, proper interaction of the dual streams contributes to goal-driven actions.

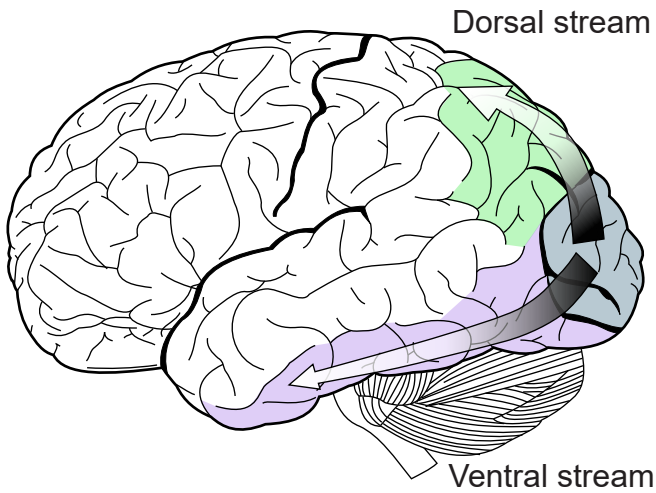


Figure 7.21 The dual stream hypothesis proposes that visual information is processed by two related neural pathways after leaving V2, a dorsal stream (green) and a ventral stream (purple).

Clinical connection: Dysfunction of the dual-stream

Deficits in the circuitry of visual processing up until this point in the chapter result in largely predictable outcomes. However, once that information enters the cortex of brain, the circuitry involved becomes much more complex, and unusual visual perception disorders may result from atypical neural communication in the brain.

Akinetopsia, or **motion blindness**, is a very rare visual condition resulting in the inability to perceive objects in motion. Instead of seeing movement, these people see the world as a series of "freeze frames" that refresh occasionally, or only whenever the object stops moving. People with akinetopsia have severe difficulty crossing the street or pouring a cup of tea. Usually, akinetopsia results from brain injury like trauma or a stroke.

Prosopagnosia, or **face blindness**, is another unusual visual condition that results in people who have tremendous difficulty with identifying faces. People with prosopagnosia recognize that they are looking at a face. They see the eyes, nose, and mouth, but they often cannot identify who that face belongs to. Some cases are so severe, that they cannot even recognize themselves in a mirror. Prosopagnosia is far more common than akinetopsia; an estimated 1% of people have some degree of face blindness.

Saccadic eye movements

Take a look at the objects around you right now. You don't notice it, but as you shifted your focus from one object to the next, your eyes made tiny, rapid, jerking motions over the different objects instead of moving smoothly across your field of vision. These rapid movements, driven by the actions of the extraocular muscles, are called saccades (pronounced "suh-KAHD"). Saccades are among the fastest muscle movements in the body.

These saccades help you accurately capture visual information from your surroundings. The fovea, which has the highest acuity vision, only collects information from a tiny portion of your visual field. By performing these saccades, you can scan over several objects with the highest acuity, thus allowing your brain to build a representation of the visual field.

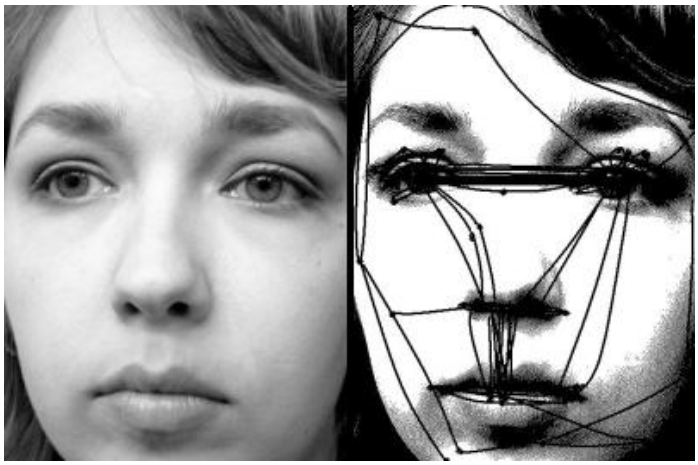


Figure 7.22 Eye tracking (right) allows us to observe the saccades that may happen when we scan over this facial stimulus (eye).

The Open Neuroscience Initiative is funded by a grant from the Vincentian Endowment Fund of DePaul University.

Cover: Niagara66 / CC BY-SA (https://creativecommons.org/licenses/by-sa/4.0) https://upload.wikimedia.org/wikipedia/commons/1/19/Mammoth_Cave_National_Park_01.jpg
7.1 Image by Anja from Pixabay (https://pixabay.com/photos/key-keychain-door-key-house-keys-453480/)
Image by Ann San from Pixabay (https://pixabay.com/photos/key-door-opening-door-lock-2732312/)
7.2 https://upload.wikimedia.org/wikipedia/commons/3/30/EM_spectrumrevised.png modified by Austin
7.3 https://upload.wikimedia.org/wikipedia/commons/d/dd/The_outlines_of_anatomy%2C_physiology%2C_and_hygiene._Being_an_edition_of_The_essentials_of_anatomy%2C_physiology%2C_and_hygiene%2C_rev._to_conform_to_the_legislation_making_the_effects_of_alcohol_and_other_%2814596873080%29.jpg
7.4 Image by Karen Arnold from Pixabay (https://pixabay.com/illustrations/vintage-flowers-vase-1772583/)
7.5https://upload.wikimedia.org/wikipedia/commons/0/06/PSM_V45_D217_Section_of_the_lens_showing_the_mechanism_of_accommodation.jpg modified by Austin Lim
7.6 https://upload.wikimedia.org/wikipedia/commons/7/74/1107_The_Extrinsic_Eye_Muscles_right_eye_lat.png modified by Austin
7.7 https://www.flickr.com/photos/internetarchivebookimages/14592617690/
7.8 https://upload.wikimedia.org/wikipedia/commons/e/e3/Rods_Cones_Synapse.svg Modified by Austin Lim
7.9 https://upload.wikimedia.org/wikipedia/commons/0/02/Red_geranium_Purkinje.jpg
7.10 https://commons.wikimedia.org/wiki/File:Cone-response-en.png#/media/File:Cone-response-en.png modified by Austin Lim
7.11 https://commons.wikimedia.org/wiki/File:Distribution_of_Cones_and_Rods_on_Human_Retina.png modified by Austin Lim
7.12 https://upload.wikimedia.org/wikipedia/commons/a/af/Color_blindness.png
https://upload.wikimedia.org/wikipedia/commons/e/e0/Ishihara_9.png
https://upload.wikimedia.org/wikipedia/commons/e/e0/Ishihara_9.png
https://commons.wikimedia.org/wiki/File:Ishihara_3.jpg
7.13 https://upload.wikimedia.org/wikipedia/commons/3/34/11-cis-Retinal_Vitamin_A-Aldehyd.svg
7.14 https://upload.wikimedia.org/wikipedia/commons/6/6f/Zludzenie_Hermana.png
7.16 https://upload.wikimedia.org/wikipedia/commons/5/5c/1543%2CVisalius%27OpticChiasma.jpg
7.17 https://upload.wikimedia.org/wikipedia/commons/4/4d/Neural_pathway_diagram.svg modified by Austin Lim
7.18 Reprinted with permission from Dr. Pasko Rakic (Early divergence of magnocellular and parvocellular functional subsystems in the embryonic primate visual system)
7.19 https://upload.wikimedia.org/wikipedia/commons/e/e5/Hemia.PNG modifed by Austin Lim
7.20 https://commons.wikimedia.org/wiki/File:Retinotopic_organization.png modified by Austin Lim
7.21 https://commons.wikimedia.org/wiki/File:Ventral-dorsal_streams.svg modified by Austin Lim
7.22 https://upload.wikimedia.org/wikipedia/commons/d/d3/Szakkad.jpg

Chapter 8:

Sensation and Perception: Other Physical Senses

Austin Lim, PhD (DePaul University)
Editor: Alexander Rajan, PhD

Our nervous system is equipped with a variety of specialized biological "tools" that can detect much more than just photons of light. We can detect the shape of air waves, and interpreting those signals give us sound information and the perception of music. We can detect stimuli with our skin, such as temperature, pressure, and textures. We can also detect physical information about our own bodies, as in the way our head is tilted or the position of our limbs. In this chapter, we describe these other forms of physical stimuli, and how that information is conveyed to and represented in the brain.

All sensory systems, including vision (Chapter 7) and the systems described below, follow the same general path of communication into our nervous systems and awareness. First, the incoming signal must reach a cell (generally called a **transducer**) that can change its electrical properties in response to the stimulus (such as a retinal photoreceptor cell responding to photons). Then, that information initiates a series of signals into the CNS, reaching structures such as the thalamus (in most cases; olfaction being an exception), primary sensory cortical areas, and finally, higher order perception. Although there are several sensory components throughout our body that detect these signals, there are no sensory receptors in our central nervous system.

As we discuss sensory and perception further, it's good to keep in mind that not all sensations are perceived. For example, your body can detect changes in blood pressure, and your brain responds to these changes activating compensatory neural circuits. However, this response does not reach consciousness and are not perceived.

Chapter 8 outline

8.1 The Auditory System

Unlike photons of light, sound waves are compressions and rarefactions of a medium. For us land animals, that medium is usually air, but sound waves can propagate very well in water or through solids. Without a medium, sound fails to propagate: As the tagline from the thriller Alien reads, "In space, no one can hear you scream."

8.1.1 Sound waves

Before we get to the anatomical structures involved in sound perception, it is important to first understand the physical nature of sound waves. All sounds, from the clattering of a dropped metal pan to the melodies of a Mozart violin concerto, are contained in their corresponding sound waves. There are three components of sound waves.

Frequency, or "How often do the sound waves compress?"

The more often they repeat, the higher the pitch. The highest notes humans are able to hear is around 20,000 Hz, a painfully-shrill sound for those who can hear it. People often tend to lose their high frequency hearing as they age. On the opposite end of the spectrum, low frequency sounds are the deep rumbles of bass, and the human ear can hear sounds down in the 20 Hz range. Doubling the frequency changes the tone by an octave: concert A is 440 Hz, and the A an octave higher is 880 Hz.

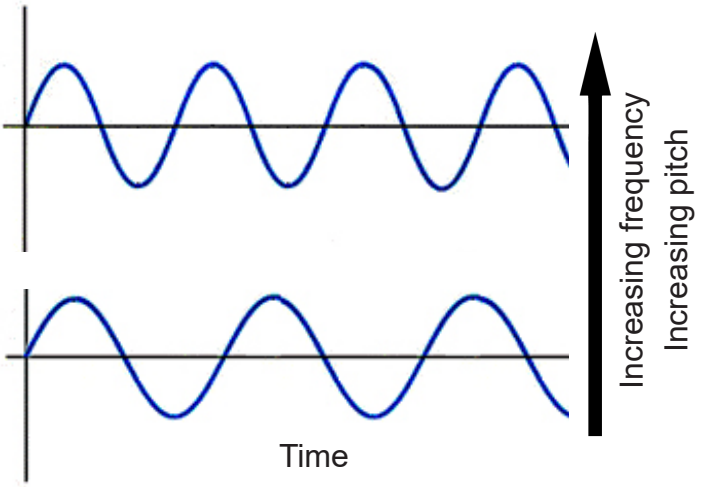


Figure 8.1 Pitch is a result of increased frequency of displacement.

Amplitude, or "How much do the waves displace the medium from baseline?"

The larger the amplitude of the wave, or the greater distance between the peak and the trough of the signal, the louder the sound is. Loudness is measured in decibels (dB). To give you an idea of approximate sound intensities, the background noise of a quiet library is about 40 dB, and a typical conversation is close to 60 dB. A rock concert or lawnmower is between 100 and 110 dB, which is right around the pain threshold. Prolonged exposure to these high amplitude sound waves can lead to permanent damage to the auditory system resulting in hearing loss or tinnitus (a ringing in the ear, even in the absence of a sound stimulus).

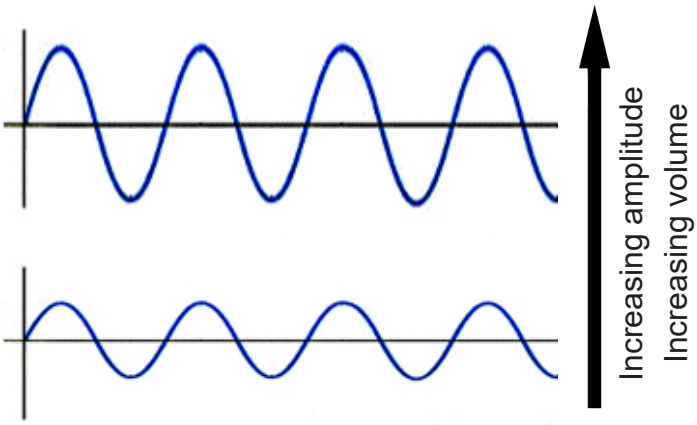


Figure 8.2 Volume is a result of increased amplitude of displacement.

Timbre, or "How complex is the wave form?"

Timbre (TAM-bur) refers to the "color" or the "character" of the sound. Both a piano and a guitar could play the exact same note, yet these two instruments would produce very different sounds. This happens because the sound waves created by each instrument differ in complexity. Both notes contain the same **fundamental**

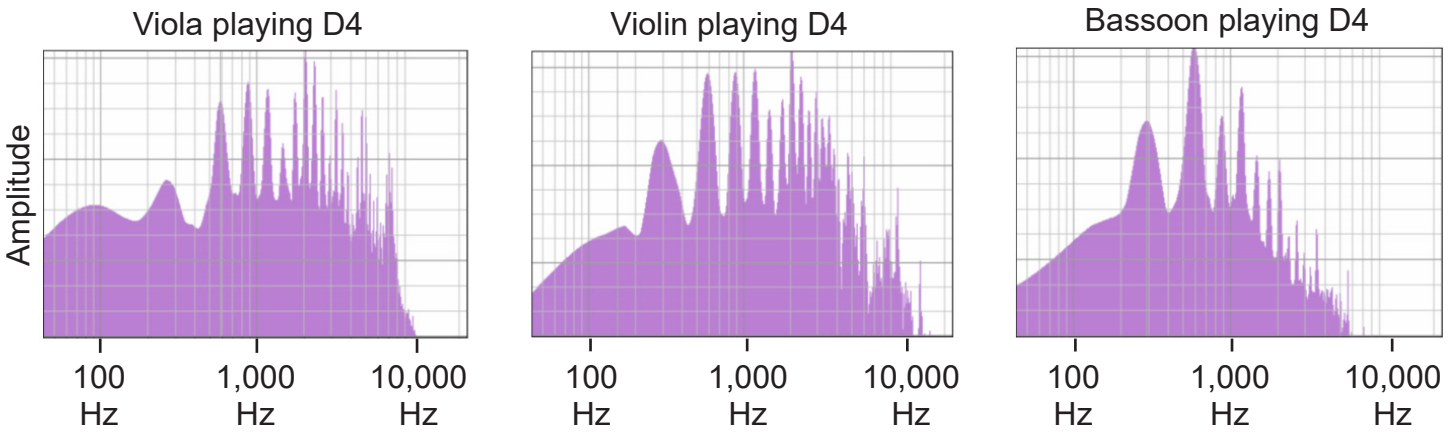


Figure 8.3 Timbre refers to the complexity of the waveform. Instruments sound different because the they produce peaks at different frequencies related to the fundamental frequency.

frequency which gives us the pitch of that note. But each wave differs in the number of other high frequency components called **overtones** that are contained within that note. These overtones oscillate at a frequency that are multiples of the fundamental frequency (twice as often, for example). A sound wave without any overtone frequencies is called a **pure tone**, or a **sine wave tone**.

8.1.2 Physical structures of the auditory system

Our auditory system is a series of physical structures and nervous system components that are responsible for conveying sound waves into meaning and context.

The external component of the auditory system begins with the **pinna** or **auricle**, or more simply, the ear. Its shape functions as a funnel, capturing and channeling sound waves into the **auditory canal**. The pinna and the auditory canal are parts of the outer ear. Also, because the pinna is asymmetrical, its shape helps us determine where a sound is coming from. In some nonhumans, the pinna serves these functions and more: some animals are able to disperse excess heat through their ears (elephants), and some even use them to display emotion (dogs, horses).

At the end of the auditory canal is the **tympanic membrane**, or **ear drum**. This membrane is a very delicate piece of tissue at only 0.1 mm thin, and is subject to damage by physical

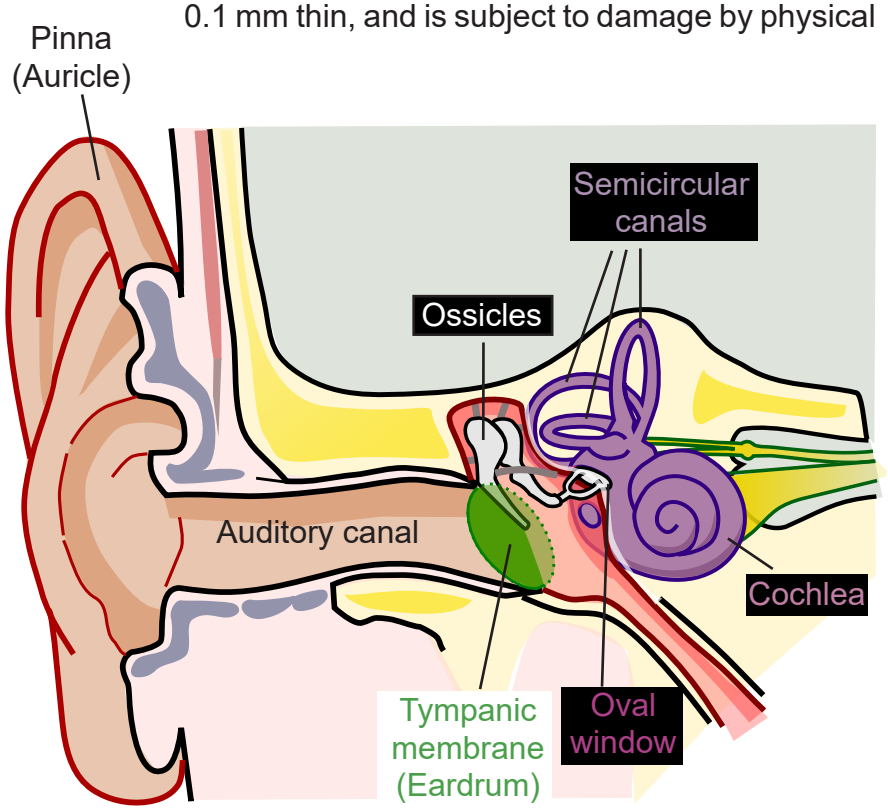


Figure 8.4 Anatomy of the auditory system.

injury such as head trauma, nearby explosions, or even changes in air pressure during scuba diving. When incoming sound waves reach the tympanic membrane, it vibrates at a matching frequency, amplitude, and timbre. The tympanic membrane also represents the boundary between the outer ear and the **middle ear**.

Physically attached to the tympanic membrane are the **ossicles**, a series of three bones that convey that vibrational sound information. These bones in order, called the **malleus**, **incus**, and **stapes**, conduct vibrations of the tympanic membrane through the air-filled middle ear. The tympanic membrane and the ossicles function to amplify incoming sounds, generally by a tenfold difference. This amplification is important because the inner ear is filled with liquid rather than air, and sound waves do not travel very well when moving from air into a denser medium - think about how muffled sounds are when you submerge your head underwater.

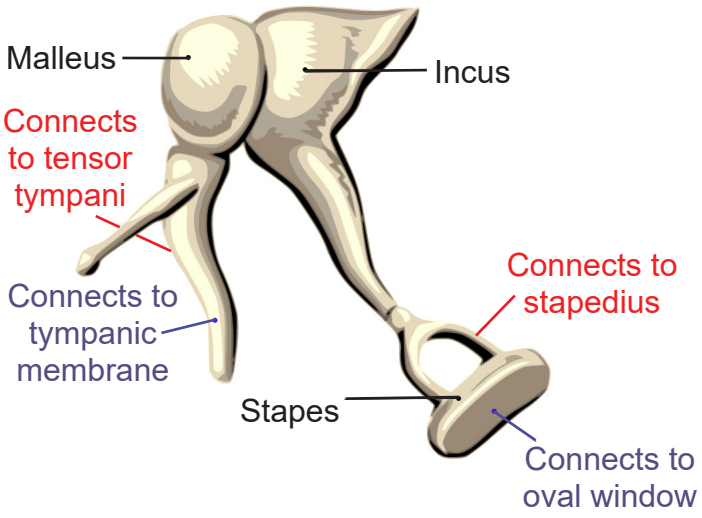


Figure 8.5 Anatomy of the ossicles, found in the middle ear. Bones are labeled in black, connections to membrane labeled in blue, and connections to muscle labeled in red.

The movement of the ossicles are partially regulated by two different muscles, the **tensor tympani** muscle which connects with the malleus, and the **stapedius** muscle which connects to the stapes. When these muscles contract, the ossicles move less, which decreases the intensity of loud sounds. This response, called the **acoustic reflex**, dampens incoming sound by about 15 dB. (This is why we talk much louder than normal when we first leave a concert: we have lessened auditory feedback from our ears, so we tend to talk louder to compensate.) Crickets use a similar reflex so that they don't deafen themselves while chirping!

The stapes, the third of the ossicles, is physically connected to the **oval window**, a thin membrane that is at the entrance to the main auditory structure of the inner ear, a hollow bone called the **vestibular labyrinth**. The auditory part of this structure is the spiral-shaped **cochlea** (*cochlea* is named for the Ancient Greek word "snail shell".) Think of the cochlea as a rolled-up cone that makes about 2 and 3/4 turns. If this cone was theoretically unrolled, the widest diameter portion (called the **base**) would be closest to the oval window, while the narrowest portion (called the **apex**) would be at the center of the spiral.

This change in shape from the base to the apex is important: objects with different stiffness vibrate at different frequencies. The base of the cochlea is stiffer than the apex, and so it vibrates are higher frequencies than the apex. This makes different points along the spiral of the cochlea responsive to different frequencies.

Inside the cochlea is a specialized epithelial membrane called the **Organ of Corti**. The Organ of Corti is the first nervous system structure that is responsible for processing physical vibrations and converting them into signals that the nervous system can interpret.

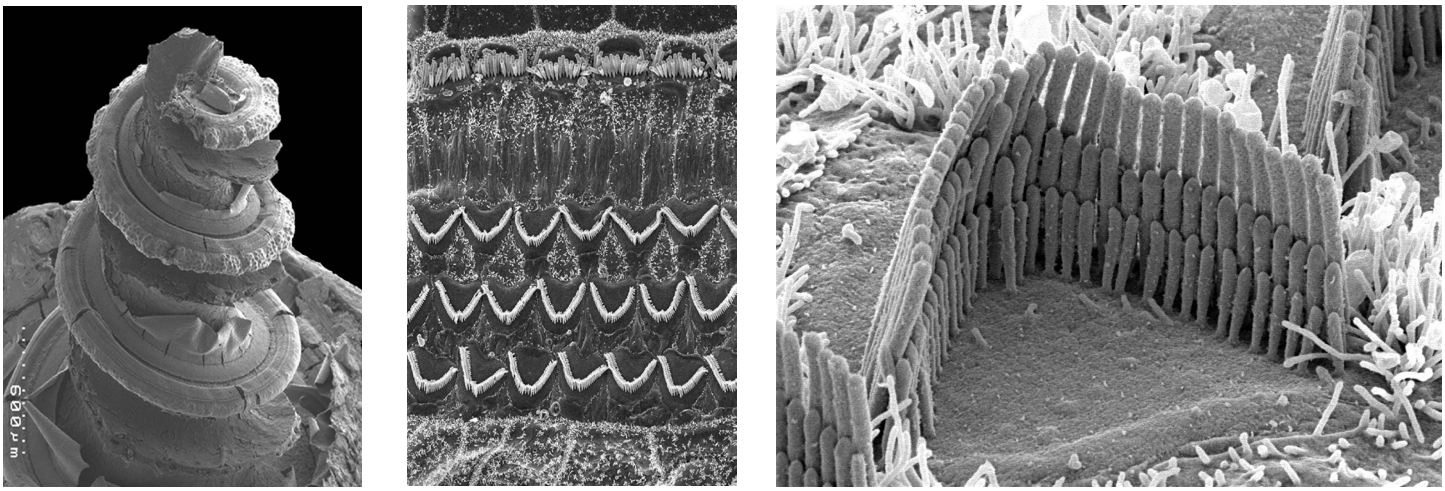


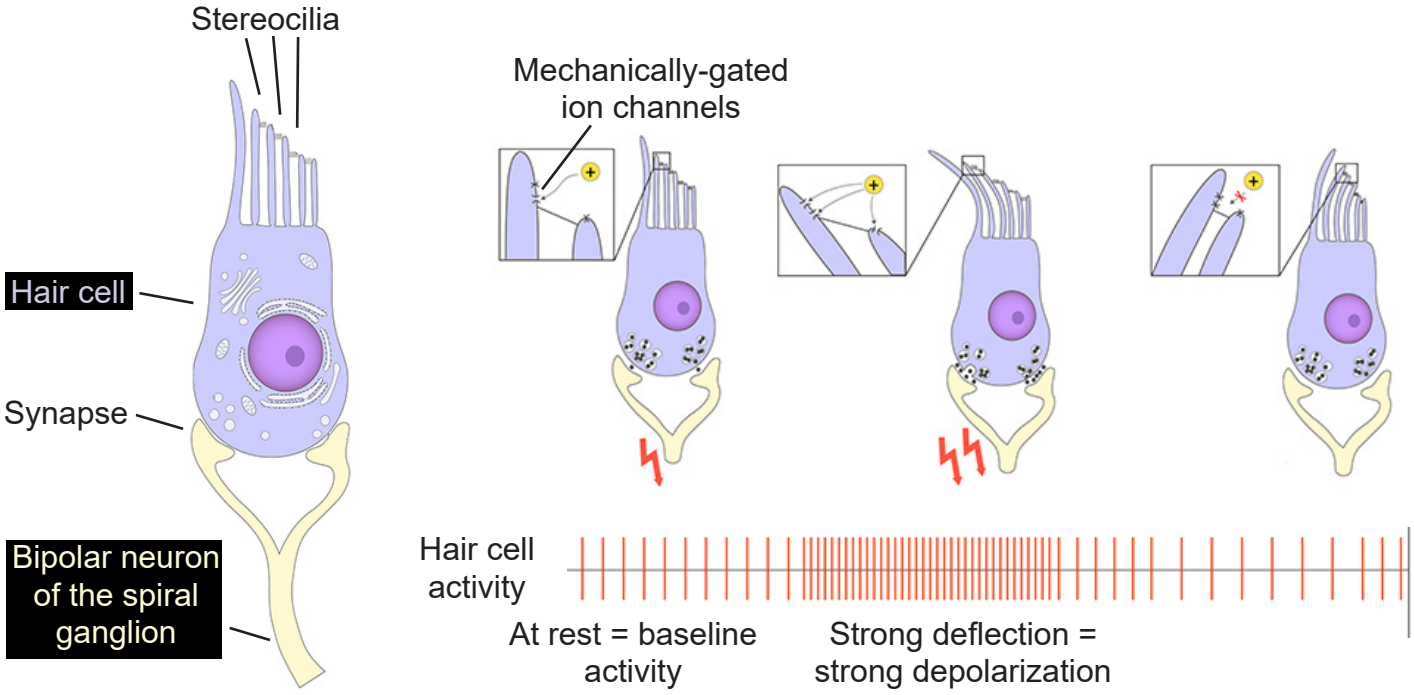
Figure 8.6 Anatomy of the Organ of Corti and hair cells under progressively higher magnifications in an electron microscope. The Organ of Corti is a spiral structure (left). The three rows of the outer hair cells and the single row of inner hair cells can be seen by zooming in (middle). At highest magnification, the stereocilia can be seen (right).

8.1.3 Neural components of the auditory system

The Organ of Corti contains the components necessary for converting sound waves into action potentials. Adjacent to the Organ of Corti in the cochlea is a liquid called **endolymph**. The endolymph is a high potassium, low sodium solution that is similar to CSF. It makes up the extracellular solution in the inner ear.

Embedded along the interior surface of the Organ of Corti are the somata of **hair cells**, the

Figure 8.7 Physiology of hair cells. Physical deflection causes the opening of the mechanically-gated ion channels, which allows movement of K^+ into the hair cells, causing excitation.



primary sensory neurons that interpret physical movement. They are named “hair cells” because of their cellular structure; each hair cell has somewhere between 30 and a few hundred hair-shaped **stereocilia** that protrude away from the Organ of Corti, reaching into the endolymph. We have two different populations of hair cells, the **inner hair cells** and **outer hair cells**.

Movement of the inner hair cells is important for the detection of sounds. When a vibration reaches the oval window, the endolymph also vibrates, causing a physical movement of the membranes within the cochlea, which displaces the stereocilia. The physical movement causes **mechanically-gated ion channels** to open, allowing K^+ to enter the hair cells, resulting in depolarization and neurotransmitter release. This process of physical motion that leads to neural signaling is called **mechanotransduction**. Hair cells are very sensitive to stereocilia movement: a deflection on the order of 100s of picometers (a picometer is one trillionths of a meter) is sufficient to cause changes in electrical signaling in a hair cell!

The length of the stereocilia differs between hair cells, and these differences allow them to respond to different frequencies of vibrations. Shorter hair cells are more responsive to high pitch sounds, and are located closer to the base of the cochlea. On the other hand, longer hair cells are more sensitive to lower pitched sounds, and are located near the apex, in the center of the spiral. Accordingly, there are two “tuning” mechanisms in the cochlea that make specific inner hair cells sensitive to specific frequencies: the position along the spiral, and the length of the stereocilia.

The outer hair cells function as an amplifier to increase the intensity of vibrations. It is estimated that the outer hair cells increase sound by anywhere between 20 and 80 dB. We have three rows of outer hair cells.

Clinical connection: Hearing loss

Many people experience permanent hearing loss, a decrease in volume by 25 dB or more. Hearing loss is divided into two categories. **Conductive hearing loss** is a result of changes to the auditory system up to the oval window, such as a tumor in the ear canal, a perforation of the tympanic membrane, or changes in middle ear pressure (such as how everything sounds muffled while changing altitudes when an airplane takes off, for example). **Sensorineural hearing loss** results from changes at the level of the inner ear or further up in the neural pathway, such as hair cell damage, a brain tumor, bacterial or viral infections, or exposure to various toxins or drugs.

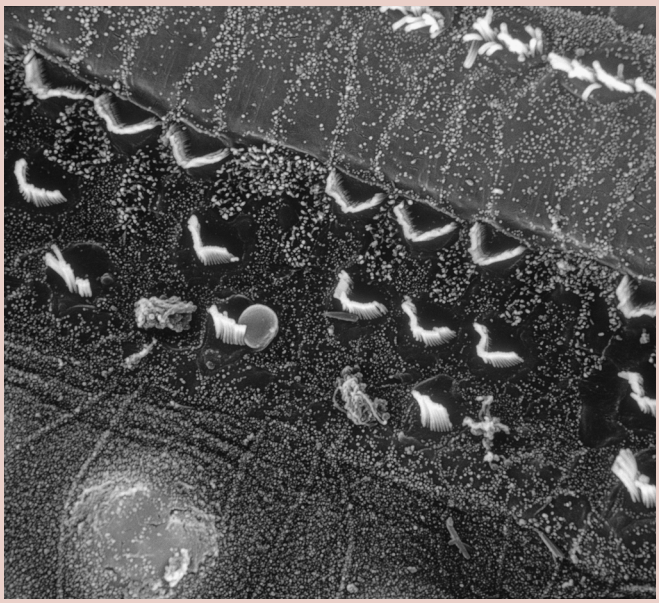


Figure 8.8 Damaged hair cells under an electron microscope.

The most common cause of hearing loss is excessive noise exposure. Although the acoustic reflex is capable of dampening the intensity of the incoming vibrations, prolonged exposure to high amplitude sound waves can still cause damage. Motorcycles, the maximum volume on headphones, or loud venues like concerts and clubs can produce sounds in the 95-110 dB range, which can cause some permanent hearing loss. Additionally, the acoustic reflex is not fast enough to minimize damage from sudden, loud sounds in excess of 120 dB, such as a gunshot. All of these sources of acoustic trauma are preventable by wearing appropriate hearing protection, which can decrease the intensity of sounds by up to 30 dB.

Old age is another common cause of hearing loss, likely because older people have had more accumulated exposure to noise. An estimated 1 in 3 people older than 65 have hearing deficits. We are born with about 15,000 hair cells, but throughout the course of our life, many get damaged irreparably. The shorter hair cells are more sensitive to injury, so it is common for people to lose sensitivity to high-frequency sounds. The loss of these hair cells can begin as early as a person's 20s - some places intentionally play high frequency sounds to prevent teenagers from loitering.

Partial hearing loss can be reversed with the help of medical devices. A **hearing aid** is a processor that helps to filter out background noise, decrease pitch, and amplify incoming sounds. A **cochlear implant** is a surgically-implanted device that receives incoming sound information and directly stimulates the auditory nerve via electrodes, bypassing the external components of the auditory system.

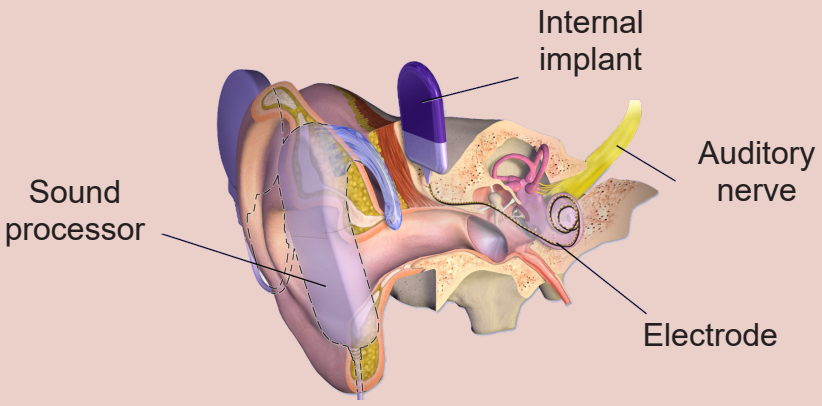


Figure 8.9 A cochlear implant directly stimulates the neural components of the auditory system.

Hair cells form glutamatergic synapses onto bipolar neurons of the **spiral ganglion**. The axons of these neurons make up the **vestibulocochlear nerve** (**Cranial nerve VIII**, historically called the **auditory nerve**. This nerve also carries vestibular information, which is also processed in the inner ear. See section 8.2.2 for more details.) Some neurons project into a pontine area called the **superior olive**, an integrative center that receives bilateral inputs from both ears. Other neurons project into an area of the rostral medulla called the **cochlear nuclear complex**, which carries out some auditory processing functions.

One of the primary functions of the superior olive is to help us figure out if a sound originates from the left or the right side of our head. To do this, the superior olive performs two different calculations. First, sounds are louder the closer you are to them. A difference in volume between what one ear and the other ear perceives, called the **interaural level difference**, is evaluated by neurons in the **lateral superior olive**. The

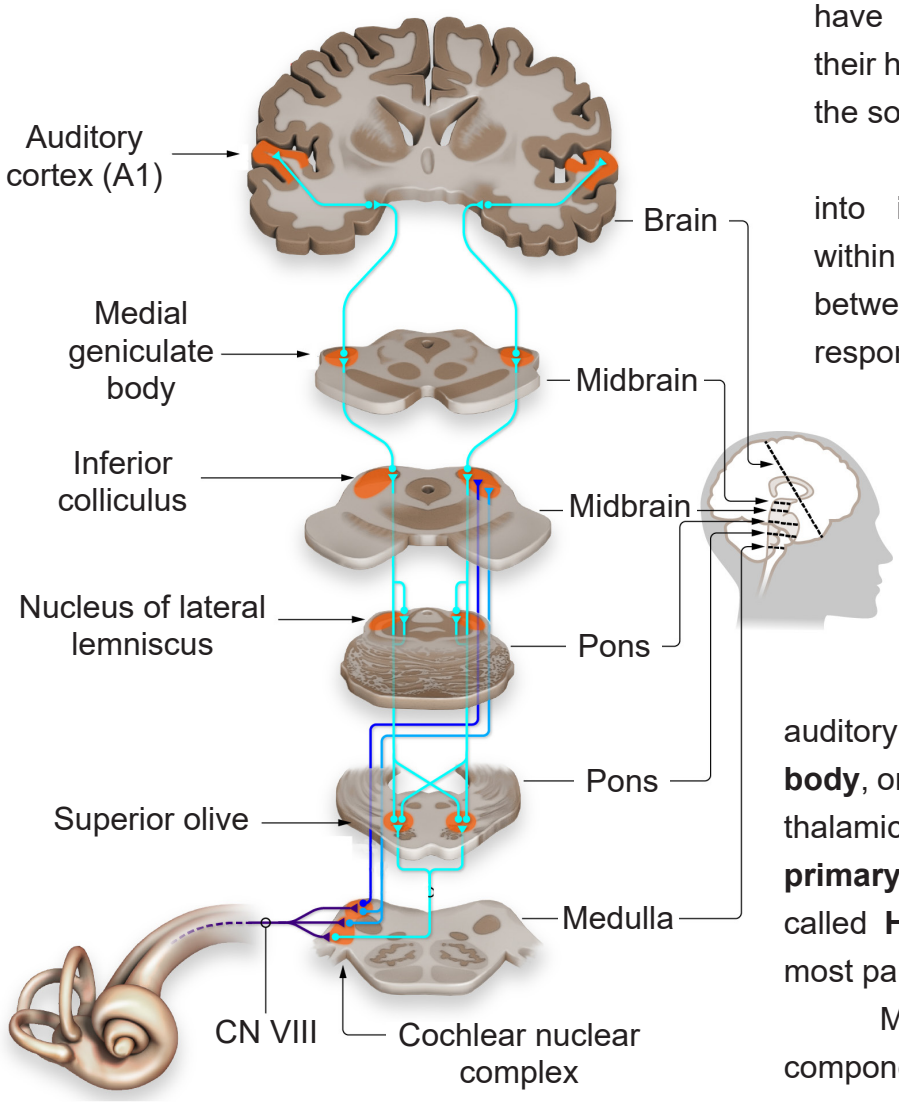


Figure 8.10 The neural pathways involved in the auditory system.

second calculation is related to the speed of sound. Sound travels roughly a kilometer in three seconds. Because of this delay, sounds reach one ear sooner than it reaches the other. This **interaural time difference** is assessed by neurons of the **medial superior olive**. The ability for these neurons to properly localize sound is the result of years of training.

Because both ears (in humans and most non-humans) are at the same height, sounds above and below can be difficult to localize. This is why sometimes people and animals tilt their heads to try to hear better. Some species of owls have ear openings at different heights on their heads, increasing their ability to localize the source of a sound on the vertical axis.

Auditory information is then passed into **inferior colliculus (IC).** Signaling within the IC is important for interactions between multiple sensory inputs and a motor response. These IC neurons are particularly responsive to biologically-relevant sounds, such as unexpected noises, which may signal an approaching predator. Processing in the IC helps the animal focus their attention towards these stimuli.

The IC then conveys that auditory information into the **medial geniculate body**, one of the nuclei of the thalamus. These thalamic neurons then send projections into the **primary auditory cortex**, or **A1** (historically called **Herschel's gyrus**.) A1 is the dorsal-most part of the temporal lobe.

Many of the neural processing components of sound, such as the Organ of Corti, the spiral ganglion, and A1, are **tonotopically organized**, meaning that adjacent physical areas are responsible for conveying information from adjacent frequencies. For example, the hair cells that respond most to 440 Hz vibrations are right next to to cells that respond maximally to 441 Hz, but far away from cells that respond most to 14,000 Hz. Likewise, in the cortex at A1, the cells that best process 440 Hz are adjacent to those that best process 441 Hz, but far away from those that maximally respond to 14,000 Hz.

After signal processing in A1, that sound information is then passed through the secondary and tertiary auditory cortices, which likely carry out higher order auditory functions. One of the earliest studied cortical areas involved in auditory

Clinical connection: Tinnitus

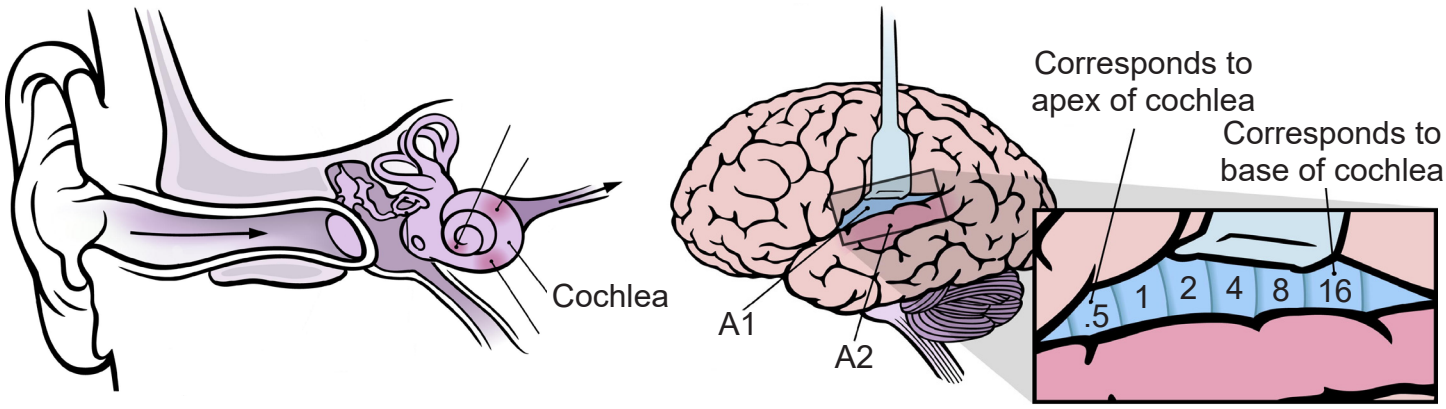
Tinnitus is a condition characterized by the occasional perception of a ringing, whistling, buzzing, or clicking sound in the absence of a genuine stimulus. Tinnitus is a symptom rather than a disease. It often appears after some degree of hearing loss or injury; alternatively, it can come and go without any particular underlying cause. An estimated 15% of people experience some amount of tinnitus, but about 2% of people have clinically significant tinnitus.

It is still unknown what causes tinnitus, however a common predictor of tinnitus is hearing loss. This correlation has led researchers to our current leading theory about tinnitus, which suggests that the perception of phantom sounds originates in the brain as a result of faulty plasticity: the "understimulated" brain, which doesn't receive expected auditory inputs, produces these abnormal signals.

processing is Wernicke's area, which is critical to the understanding of speech and language (See chapter 14 for more details).

There is evidence to support a dual stream organization to auditory perception, just like in the visual system in cortex. The dorsal auditory stream helps identify the location of sounds (analogous to the *where* pathway), speech production, and language related memory, while the ventral auditory stream contributes to sound recognition (similar to the *what* pathway) and sentence comprehension.

Figure 8.11 Tonotopy is maintained across many of the auditory neural structures, from the cochlea to the auditory cortex.



8.2 The Vestibular System

When we tilt our head to the side, or look up and down, that movement information is conveyed to our brain using the **vestibular system**. The vestibular system is a sort of three-dimensional compass that can detect head movement, and that information helps us figure out how our head is oriented and how to balance ourselves in changing conditions.

The vestibular system is made up of two structures that are intimately tied in with the anatomical features of the inner ear.

8.2.1 The Otolith Organs

Next to the cochlea and within the vestibular labyrinth are two membranous sacs, the **saccule** and the **utricle**. Collectively, these structures are responsible for determining changes in inertia. The saccule is more sensitive to vertical movements, like when you are standing in a moving elevator. The utricle is more responsive to horizontal movements, such as when driving around a corner too quickly.

Within the saccule and utricle are **otoliths**, are a series of small calcium carbonate crystals

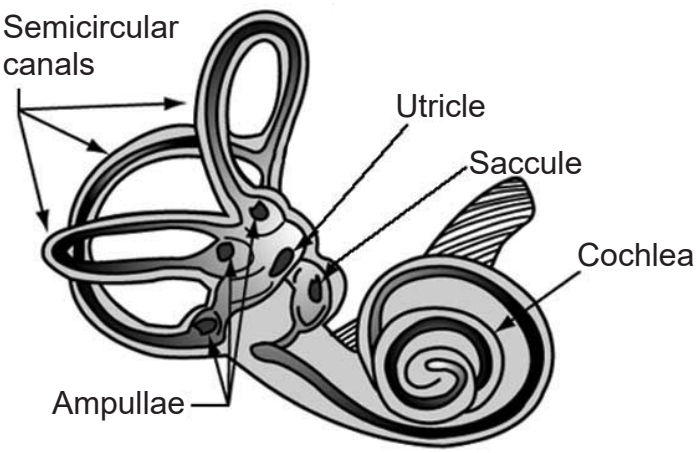


Figure 8.12 Anatomy of the vestibular system.

(the prefix *oto-* meaning "ear", and the suffix *-lith* meaning "rock") embedded in a 50 uM thick gelatinous membrane. Also embedded within this membrane are the stereocilia of a different population of hair cells. These otolith hair cells are biologically similar to the hair cells found in the cochlea: deflection of the stereocilia allows for K^+ in the surrounding endolymph to enter the cells through mechanically-gated ion channels. The main difference is the nature of the stimulus that cause the sterocilia to bend. In the cochlea, vibrations in the surrounding tissue cause hair cell movement. In the saccule and utricle, however, it is a shifting of the physical weight of the otoconia that result in hair cell movement.

The information encoded by the hair cells is passed into the brain via a branch of the **vestibulocochlear nerve (CN VIII).** These axons send projections to several brain areas, notably the cerebellum, which is a structure critically important for balance. CN VIII also projects into the reticular formation of the brain stem, the spinal cord, and the thalamus.

8.2.2 The Semicircular Canals

The **semicircular canals** are the structures that are responsible for detecting head rotation. Anatomically, they are a series of three arch-shaped membranous tubes within the vestibular labyrinth, each one oriented at a right angle to each other. Because of this shape, the semicircular canals sense and convey information about any direction of head movement: roll, pitch, and yaw.

These semicircular canals are filled with endolymph, the same potassium-rich solution that is in the cochlea that is important for auditory

sensation. At the end of each of the three canals is a small swelling called the **ampulla**. Contained in the ampulla is a gelatinous membrane called the **cupula**. Here, hair cells extend stereocilia, as well as one cellular protrusion called a **kinocilium**, into the cupula. When we tilt our head, the endolymph in the semicircular canals flows in the ampulla, which physically displaces the cilia. As in the auditory system, these hair cells have mechanically-gated ion channels which work on a "push-pull" system: when the stereocilia are deflected in one direction, the hair cells depolarize, while deflection in the opposite direction causes hyperpolarization.

Clinical connection: Vertigo

Vertigo is the sensation of spinning or movement while standing still. Vertigo often leads to dizziness, imbalance, ear pain, nausea, or vomiting. It can indirectly lead to injuries, especially as a person stands up or if they experience vertigo while driving. A symptom rather than a disease itself, an estimated 7% of people experience vertigo in their lifetimes, affecting women about 3 times as often as men.

There are a variety of conditions that could cause a person to experience vertigo. **Benign paroxysmal positional vertigo** can come and go spontaneously, and is generally not a sign of an underlying health condition. People with **Meniere's disease**, a chronic condition diagnosed in early adulthood, often experience vertigo along with other ear-related symptoms such as tinnitus or hearing loss. Bacterial or viral infections can cause inflammation of the inner ear, resulting in abnormal vestibular signaling. Excessive alcohol intoxication decreases the density of the endolymph and potentiates inhibitory signaling, resulting in exaggerated motor responses following head movement. Severe head trauma that damages the inner ear may also cause vertigo.

Vestibular reflexes

Many axons from the pontine vestibular areas send projections into the cerebellum, and these signals are important for reflexes related to balance. For example, imagine you are standing facing forward on a crowded bus when it stops abruptly. The sudden change in inertia causes your body to reflexively pushing your toes into the ground, preventing your body from toppling forward. This behavior is driven by neural signaling in the vestibular organs and their communication with the cerebellum.

A similar postural reflex that depends on vestibular inputs is the **righting reflex**. This is a behavior that develops early in several animals from humans to *Drosophila*, and allows the animal to correct their body position if they are in an abnormal orientation (imagine a baby who can get back to crawling on all fours after they have fallen on their side). Generally, this reflex is learned during the first few weeks of life. Knowing up from down depends on the afferent vestibular signals, however performance of the motor task also requires integration of visual and somatosensory inputs.

A more subtle reflex is the **vestibulo-ocular reflex (VOR)**, which is observed as our eyes stay fixated on a target while our head moves. This allows our vision to stay focused, even as the head is moving. For example, after rotating your head to the right, the eyes reflexively move to the left, which allows for the visual field to be stabilized briefly. It can be observed as a physician performs a diagnostic assessment

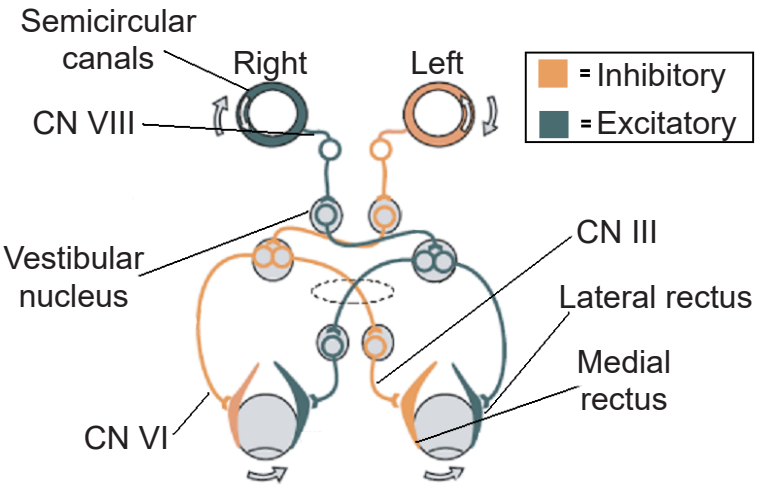


Figure 8.13 Neural circuitry underlying the vestibulo-ocular reflex (VOR) originates at the semicircular canals that detect head movement and ends with compensatory extraocular muscle activation.

called the rapid head impulse test, where they move your head quickly side to side (as if you were shaking your head “no”) while watching your gaze.

This response is driven through a series of three synapses. The first synapse is formed between the axonal projections from neurons of the vestibular system, and neurons of the vestibular brainstem nuclei. From here, these neurons send axonal projections to the contralateral hemisphere (that is to say, their axons **decussate**) and form synaptic connections with two populations of neurons in the contralateral pons. One set of motor neurons excite the extraocular muscle opposite of the eye movement: a right head turn will trigger excitation of the lateral rectus of the left eyeball, which pulls the left eyeball in the temporal direction (left). The other population are interneurons that eventually excite the medial rectus of the right eyeball, the extraocular muscle that pulls the right eyeball in the nasal direction (left, again). Simultaneously, there are inhibitory circuits that act at the opposite muscles to inhibit the eye from moving in the same direction as the head turn. Operating on the span of about 10 ms, the VOR is one of the fastest reflexes in the body.

8.3 The Somatosensory System

The nerves that receive somatosensory inputs (*soma-* referring to the body, in the same way the cell body is also called the soma) are the afferent branch of the somatic nervous system, a component of the peripheral nervous system. They give us the ability to detect information about the body. These nerves can sense a wide variety of physical stimuli, including pressure, stretch, vibration, heat, and pain.

The information about our body gets processed minimally in spinal cord circuits, ascends through the brain stem, passes through signal processing within the thalamus before reaching circuits in the somatosensory cortices. The first cortical region to receive these signals is the **primary somatosensory cortex**, or **S1**, which is the anterior-most gyrus of the parietal lobe, immediately adjacent to the central sulcus. In primates, S1 is subdivided into **Brodmann areas 3a, 3b, 1,** and **2**.

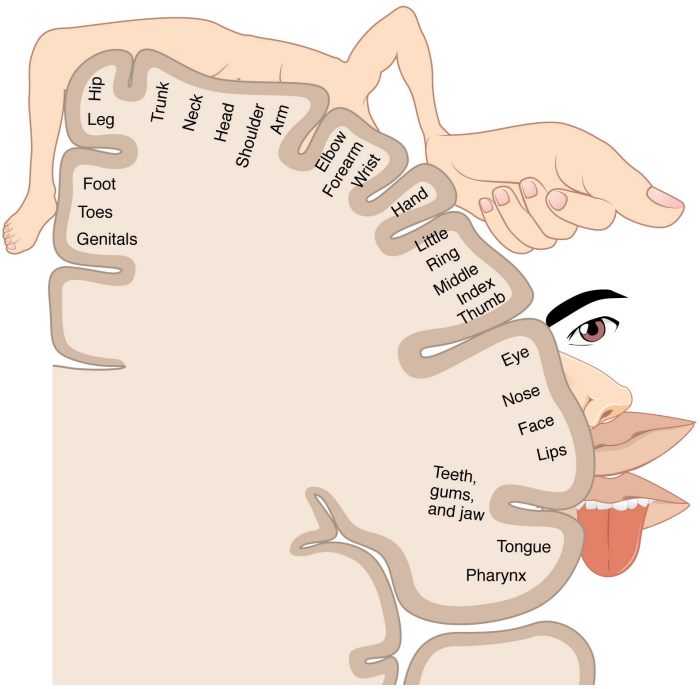


Figure 8.14 Different parts of the body ultimately send their signals through to different areas of the somatosensory cortex.

There are many different types of sensory receptors in the skin, each of which detects a very specific type of stimulus: for example, deep movement of joints, vibrations as you run a finger over a fabric, or the cold touch of an icicle in winter. For some perceptions, many of these sensory receptors are combined, but often these sources of information are carried in separate but parallel paths, a theory called the **labelled line principle**. The type of information (eg, light touch on your finger) can be tracked from the sensory receptor, through specific parts of the spinal cord, through nuclei of the thalamus, and into particular subregions of primary somatosensory cortex.

The neurons in the spinal cord, thalamus, and S1 demonstrate **somatotopic organization**, meaning that specific points on the body map to specific neural populations. For example, touching the skin on the hip will activate neurons found in the dorsal-most aspect of S1, while sensations on the tongue activate the neurons of ventral and lateral S1. Thus, there are two key bits of information kept together throughout the somatosensory pathway: what kind of stimulus, and where.

The mapping of this organization was discovered during brain surgery. In epilepsy surgery, parts of the operation are performed as the person is awake. The famed neurosurgeon Wilder Penfield electrically stimulated a small section of S1, then asked the person what they feel. Patients would respond that they feel sensations in their hand when the medial S1 was activated, the little finger when a more ventral

area was activated, and so on. In repeating this process while moving slowly across the span of S1, Penfield mapped out the organization of S1. One key observation in doing this process was the volume of brain regions corresponding to different body parts was not necessarily proportional to the volume of skin: for example, a large portion of the cortex was dedicated to sensory information from the lips or the fingertips, while very little was dedicated to the shoulders or trunk. This observation indicates that certain areas of the body are more densely packed (lips, fingertips) with sensory neurons compared to others (shoulders, trunk). A representation of a person drawn to the scale of how large each area is represented in S1 is called the **sensory homunculus**.

Figure 8.15 The somatosensory homunculus shows which parts of the body are heavily represented in somatosensory cortex, like the hands or mouth.

Some of the outputs of S1 then project into the **secondary somatosensory cortex**, or **S2**, located posteriorly within the parietal lobe, where higher order processing takes place.

8.3.1 Cutaneous receptors

Imagine reaching your hand into your pocket. In doing so, you may feel something circular with tiny ridges all along the perimeter, roughly an inch in diameter. A series of highly specialized sensory receptors just under the surface of our skin (the "*cutis*" in cutaneous refers to the skin) allows us to sense this stimulus. That information travels through ascending, afferent nerves into S1 and S2, and this is where we perceive that we have a quarter in our pocket.

It helps to divide the variety of cutaneous receptors into three categories, roughly based on the types of sensations they detect and convey.

Mechanoreceptors

Cutaneous (meaning found in the skin) mechanoreceptors are responsible for sensing mechanical changes to the skin, such as pressure or stretch. On a molecular scale, these mechanoreceptors detect changes at the skin using **mechanically-gated ion channels**. These transmembrane proteins are specialized for detecting the physical distortion of the channel, similar to the channels found in the stereocilia of the hair cells or vestibular cells. When pressure is applied to these proteins, the cation channel opens and Na^+ moves down its electrochemical gradient into the neuron, causing depolarization.

The class of mechanoreceptors can be divided into two categories based on their electrical properties in response to stimuli: **slowly-adapting** and **rapidly-adapting**. Slowly-adapting mechanoreceptors change their action potential firing rate as long as the stimulus is present,

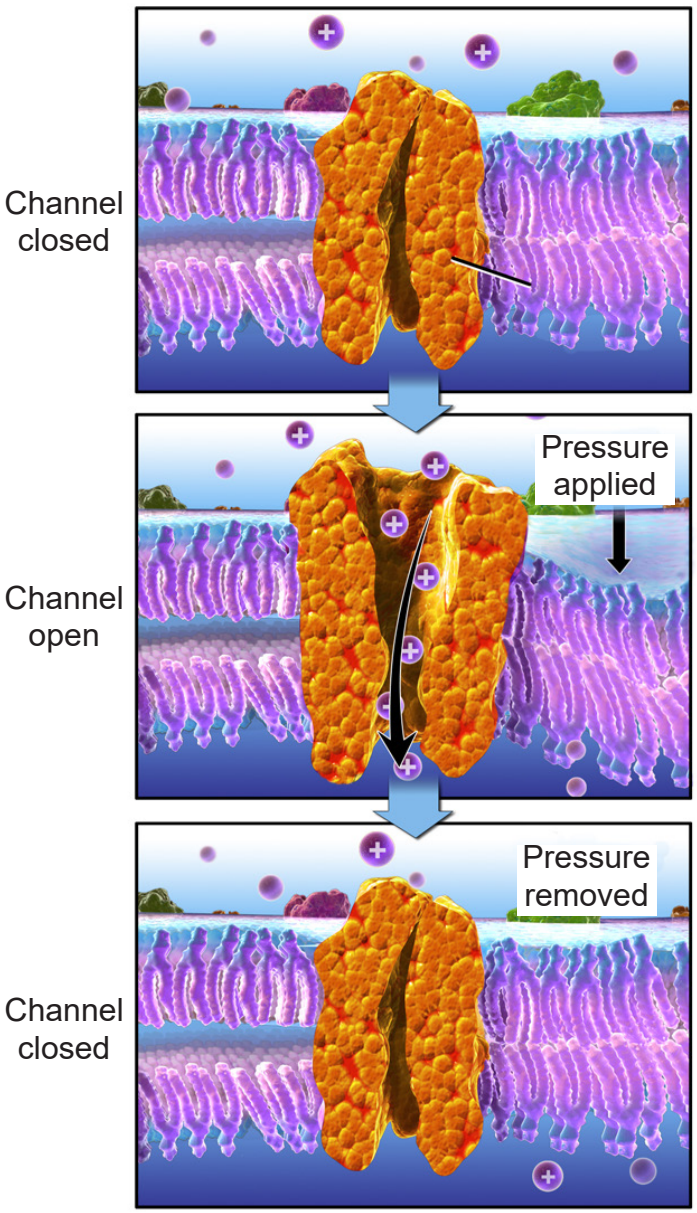


Figure 8.16 Mechanoreceptors respond to a physical distortion of the channels.

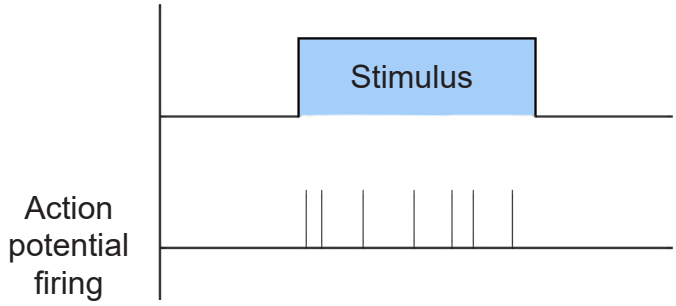
while rapidly-adapting mechanoreceptors only change activity at the moment there is a change in stimulus. For example, imagine the stimulus of a coin that is sitting in the palm of your hand. The slowly-adapting sensory neurons may increase their firing rate as long as the coin is in your palm. When the quarter is removed, they return to their baseline firing rate. These are sometimes also called **tonic receptors**. However, a rapidly-adapting sensory neuron will change its action potential firing properties only at two timepoints: the moment the coin lands in your hand, and the moment that coin is removed. These neurons signal a change in status, and are sometimes also called **phasic receptors**.

There are four different classes of cutaneous mechanoreceptors, two of which are slowly-adapting and two which are rapidly-adapting. These different mechanoreceptors have slightly different shapes, and are therefore capable of detecting different types of stimuli.

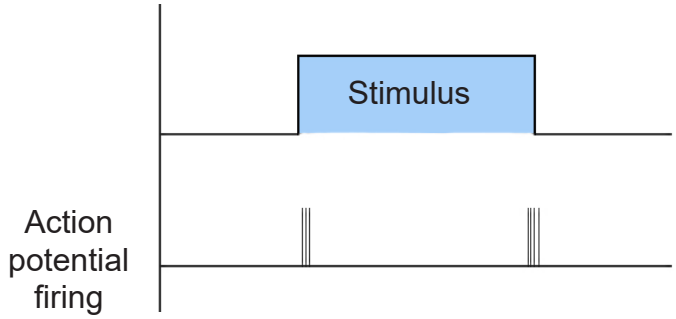
1. **Tactile epithelial cells** (also called **Merkel's discs**) are located in the superficial skin layers, and are very densely populated underneath our fingertips. They sense pressure and help us perceive edges, points, and corners. These are the most precise mechanoreceptors

Figure 8.17 Slowly-adapting mechanoreceptors modify their action potential firing as long as a stimulus is present, while rapidly-adapting mechanoreceptors do so only at the moments when a stimulus is changed.

Slowly-adapting mechanoreceptor



Rapidly-adapting mechanoreceptor



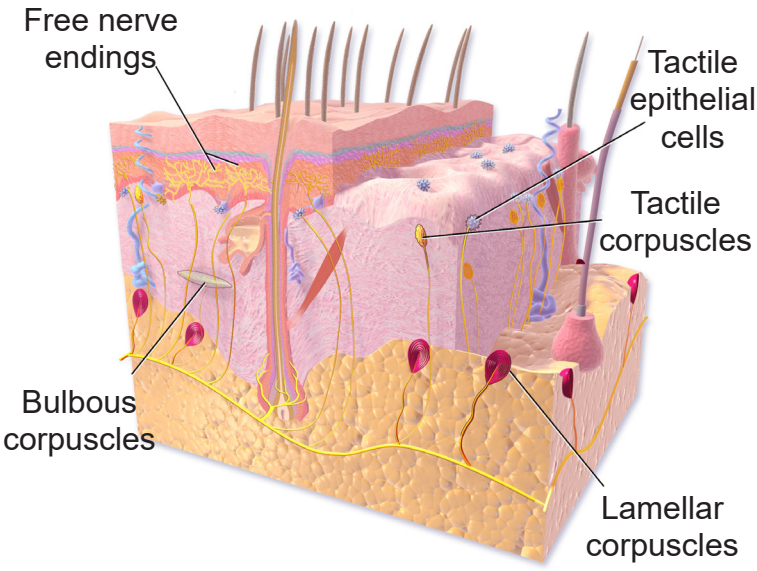


Figure 8.18 Cutaneous mechanoreceptors in the skin have different physiological properties and respond to different types of stimuli.

with resolution at the level of 0.5 mm, which allow us to detect different distances in bumps while running your fingers across Braille. These tactile epithelial cells are slowly-adapting receptors, and they release serotonin at their synapses.

2. **Lamellar corpuscles** (also called **Pacinian corpuscles**) are mechanoreceptors wrapped in several layers of connective tissue. They mostly respond to high-frequency vibrations and deep pressure. They are rapidly-adapting receptors, and are the deepest of the cutaneous receptors.

3. **Tactile corpuscles** (also called **Meissner's corpuscles**) are highly sensitive to light touch, skin movement, and low-frequency vibration. Like the tactile epithelial cells, tactile corpuscles are also found in the superficial skin layers, concentrated heavily at the fingertips. These receptors are rapidly-adapting receptors

4. **Bulbous corpuscles** (or **Ruffini endings**) respond to stretching of the skin, such as the sensation of an object slipping out of a closed hand, for example. These bulbous corpuscles are slowly-adapting receptors.

Each cutaneous mechanoreceptor is able to detect somatosensory information in an area of skin. The particular patch of skin that responds to particular cutaneous stimuli is a **receptive field**, and the smaller the receptive field, the better the brain can distinguish between two different adjacent tactile stimuli. Receptive fields differ in size throughout the body. For example, the receptive fields at the fingertips and the lips are very small (roughly 10 mm2), while the receptive fields on the back are much larger. Additionally, different classes of mechanoreceptors have different sizes of receptive fields: deeper structures, like lamellar corpuscles, have larger receptive fields compared to receptors found closer to the skin surface.

The approximate size of a receptive field can be assessed in people using a **two-point discrimination task**. Here, the experimenter will place two adjacent stimuli simultaneously onto the skin of a blindfolded person at different places. The patient is then asked if they feel one or two stimuli. In areas with large receptive fields, like

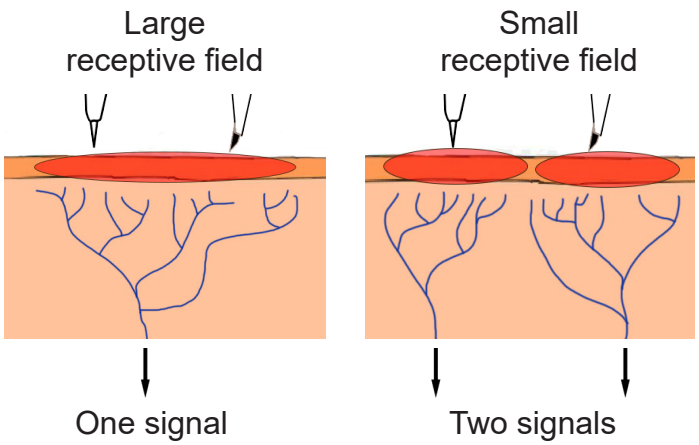


Figure 8.19 In skin areas with a large receptive field (left), two adjacent stimuli may feel like one. Skin areas with small receptive fields (right) are better able to distinguish two similarly-spaced points.

the top of the thigh for instance, a pair of points 2.5 cm apart (1 inch) may feel like one single point of contact. The distance between the two points is increased incrementally until the patient reports they first feel 2 points.

The somatosensory afferents from these receptors have their cell bodies located in the **dorsal root ganglion (DRG),** a clump of somata that are close to the dorsal side of the spinal cord. These touch receptors are the primary neurons, and are classified as **pseudounipolar** neurons based on their morphology. Pseudounipolar neurons don't have true dendrites; instead each soma sends a single protrusion which branches into two directions, one towards the skin surface where it terminates as one of the mechanoreceptors, and the other direction towards the spinal cord. Mechanosensory information travels via **Aβ axons**, which are large diameter (~10 μM) axons that can send signals on the order of 50 meters per second.

The next neurons in the ascending signaling pathway are the secondary neurons of the spinal cord, which have their cell bodies in the **dorsal horn**. These neurons send ascending axonal projections through the two different signaling pathways. The **dorsal-column medial lemniscus (DCML) tract** sends projections through the ipsilateral white matter. These DCML signals carry crude touch, pressure, vibration, and two-point discrimination information. The other ascending pathway is the **spinothalamic tract**, which projects upwards through the contralateral white matter. Spinothalamic tract carries light touch and pressure information, as well as heat and pain (described below).

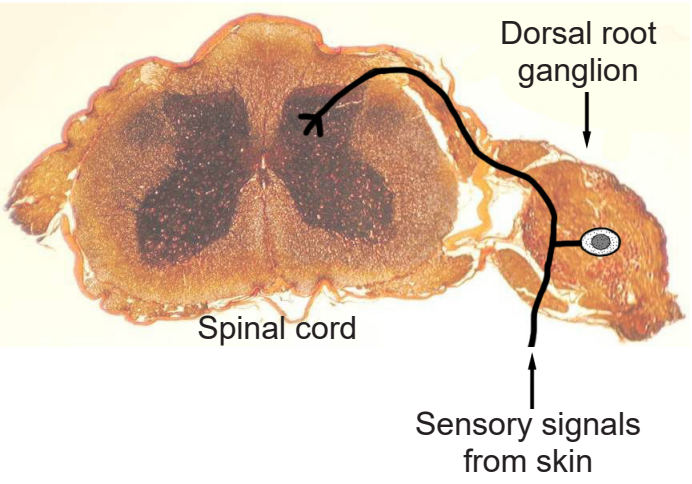


Figure 8.20 Somatosensory neurons are pseudounipolar, and have their cell bodies in the dorsal root ganglion.

The DCML neurons form synapses onto neurons found in the medulla. At this point, these neurons decussate (cross the midline), then ascend further into the **ventral posterior lateral (VPL) nucleus** of the thalamus. Spinothalamic tract neurons project directly into the thalamus.

The thalamic neurons are the tertiary neurons of the somatosensory system. They project into S1, in particular Brodmann areas 3b and 1.

Thermoreceptors

Thermoreceptors are the cutaneous receptors that sense temperature. There are two classes of thermoreceptors. The **low-threshold thermoreceptors** detect innocuous, non-harmful temperatures in the range of 15 to 45 degrees C. **High-threshold thermoreceptors** detect painful and potentially damaging temperatures hotter than 45 C or colder than 15 C.

Thermoreceptors are located in a class of cutaneous receptors called **free nerve endings**. Compared to the mechanoreceptors, free nerve endings are found closest to the surface of the skin. Like mechanoreceptors, the free nerve endings are the sensory end of the pseudounipolar neurons which have their cell somata in the dorsal root ganglion.

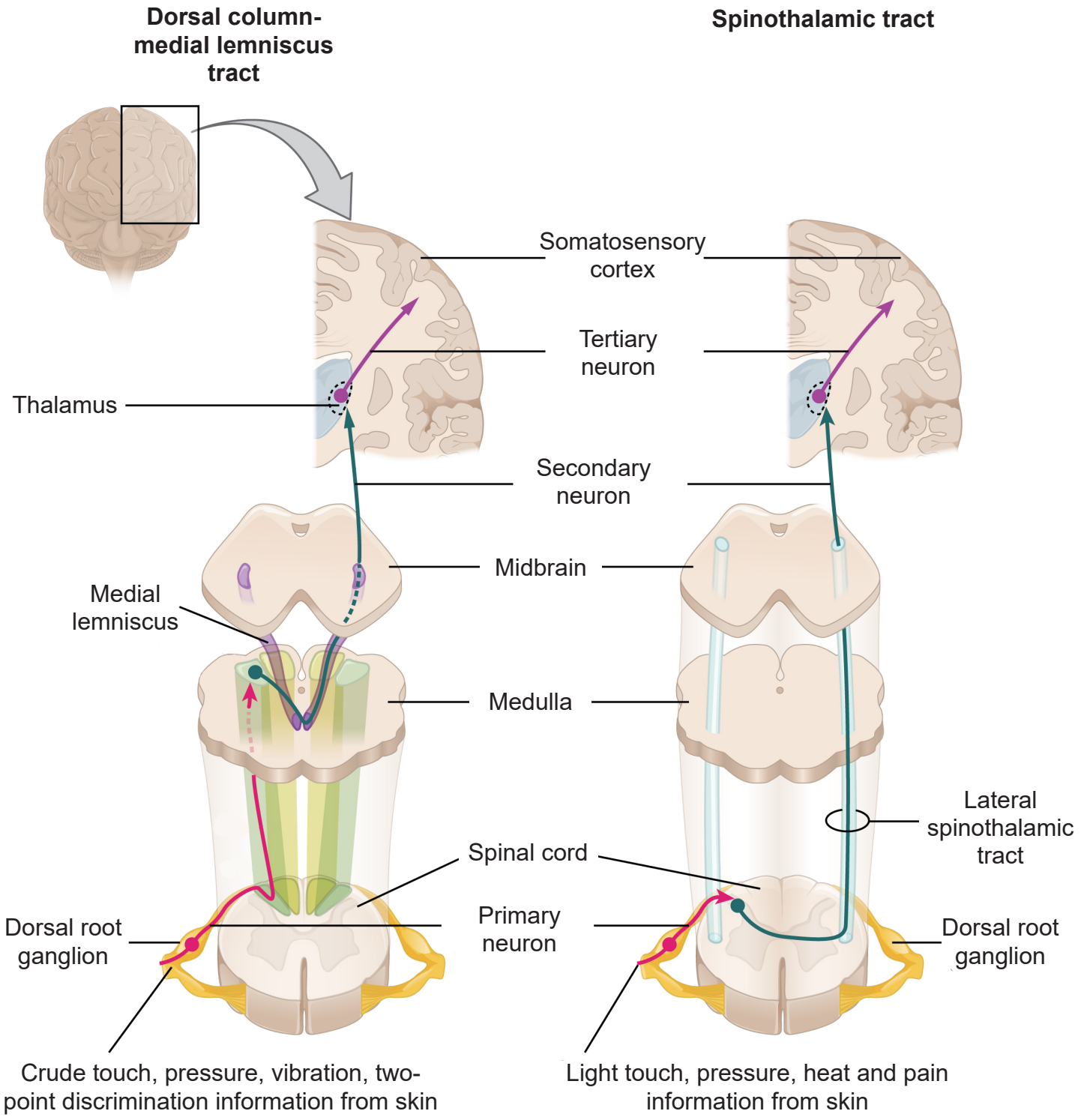


Figure 8.21 Summary of the two major ascending somatosensory pathways, the dorsal column-medial lemniscus (left) and the spinothalamic tract (right).

Expressed on these free nerve endings are a population of receptors called **transient receptor potential receptors (TRP channels).** All proteins are sensitive to changes in temperature, but these TRP channels change shape more dramatically than other proteins, making them ideal for sensing temperature. TRP channels are nonselective cation channels. Cold sensations are detected by the action of the **TRPM8** receptor, which is also activated by menthol, a chemical

isolated from peppermint that produces a cooling sensation. Warm sensations are sensed by the TRPV1 receptor, which is also activated by the chemicals capsaicin, the compound that makes spicy peppers feel hot, and allyl isothiocyanate, which is found in wasabi.

Temperature information does not get passed along the same neurons that carry mechanosensitivity signals. Cool and cold temperature sensation is passed through thinly-myelinated **Aδ fibers**, which are smaller in diameter (~5 μM) and transmit signals slower (~25 m/s) compared to the Abeta fibers. Warm and hot temperature sensation are passed through unmyelinated **C fibers**, which are even smaller (~1 μm) with slower conduction velocity (~1 m/s).

The secondary neurons of the thermoreceptors are within the dorsal horn of the spinal cord, and the ascending pathway runs through the white matter of the lateral aspect of the spinothalamic tract.

Nociceptors

Detection and avoidance of pain is a highly-adaptive behavior that can improve the odds of survival chances of an animal. Pain detection is carried out by **nociceptors**. Nociceptors can detect a variety of noxious stimuli, ranging from crush to acid and high heat. They are expressed on **free nerve endings**. Many pain-detecting neurons respond to more than one type of noxious stimulus, and these are called **polymodal nociceptors**. Nociceptors also send their signals through Aδ and C fibers. Adelta fibers detect sharp, highly-localized pain, while C fibers carry a more dull, throbbing pain that is difficult to pinpoint.

Clinical connection: Pain disorders

In healthy individuals, nociceptors that detect pain send those signals into the spinal cord. But in people experiencing **allodynia**, noninjurious tactile stimuli, such as strands of hair brushing across your forehead or the feeling of a shirt resting on your shoulders, cause the sensation of pain. Diabetes, physical trauma, or postherpetic neuralgia may cause allodynia. Drugs that modify the action potential firing properties of the afferent somatosensory pathway, such as topical anesthetics (voltage-gated sodium channel inhibitors) or calcium channel antagonists, can help relieve the pain.

A related disorder is **hyperalgesia**, an abnormally heightened perception of pain. Hyperalgesia can be the result of sensitization or as a withdrawal symptom of opioid use. Both allodynia and hyperalgesia can be symptoms of **neuropathic pain**, a broad category of pain conditions resulting from damage to the nervous system.

People experiencing pain even in the absence of pain-producing stimuli may be diagnosed with **chronic pain**. Chronic pain may be localized or widespread, affecting large areas of the body at once. Up to 10% of the global population may experience chronic pain, and a variety of risk factors contribute, from advanced age, being male, low socioeconomic status, unhealthy lifestyles, and a history of surgical intervention. Chronic pain can lead to significantly shorter lifespans, and may contribute to the opioid misuse epidemic.

A philosophically-opposite condition is called **congenital insensitivity to pain (CIP).** People with CIP are incapable of perceiving pain, no matter how severe. CIP is a rare genetic condition that is observed from birth. People with CIP have significantly shorter lifespans since they do not sense that their bodies may be injured, and may not learn to avoid exposure to damaging stimuli. Some patients may walk for days on a compound fracture of the shin bone, or never even notice that their back is broken. CIP is associated with mutations in the SCN9A gene, which codes for a component of the voltage-gated sodium channels that are expressed in nociceptors.

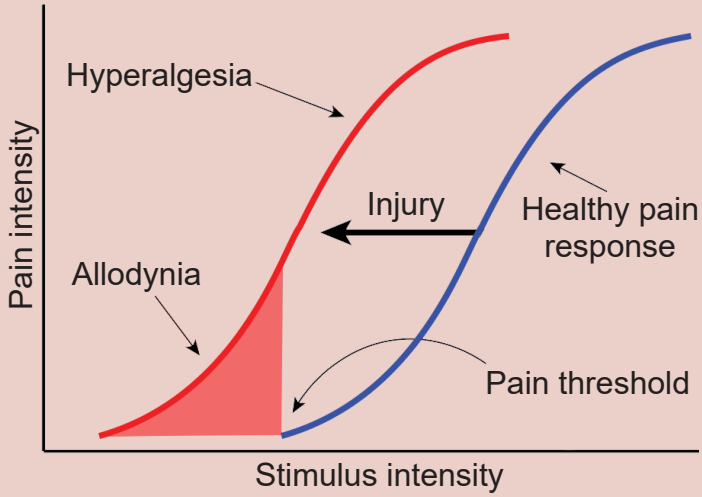


Figure 8.22 Allodynia and hyperalgesia are pain disorders characterized by altered responses to somatosensory stimulation.

There are several molecular components that make up the pain detection systems in the skin.

Some mechanoreceptors are likely expressed on free nerve endings. Nociceptive mechanoreceptors can sense crushing, shearing, or cutting of the skin. These may have a high threshold of activity, which prevents them from being activated under harmless circumstances.

Acid-sensing ion channels are cation channels that respond to low pH conditions in the dermis, which is usually seen in inflammation, which is often downstream of tissue injury. The inflammation causes a release of various cellular signaling molecules (such as prostaglandins and cytokines) which cause sensitization, an enhancement of future incoming pain stimuli.

High-threshold thermoreceptors are also involved in pain detection. On a molecular level, the same TRP channels that sense cool and hot (TRPM8 and TRPV1) are also implicated in detecting painfully cold or painfully hot temperatures.

Imagine you are reaching into your bookbag, and accidentally stick your fingertip with the sharp end of a pencil. Within half a second, the **withdrawal reflex** causes a series of muscular changes which moves your hand away from the pointy end. This motor response is driven by a circuit of neurons in the spinal cord, and is mediated completely independently from descending motor control from the brain. The nociceptive input from the sensory neurons enters the spinal cord through the dorsal horn where it forms a synapse onto an excitatory interneuron. This interneuron then signals to two other populations. The first is the motor neuron that innervates the **flexor**, a generic term used to describe muscles that withdraw your hand when contracted. The other connection is formed with an inhibitory interneuron, which innervates the motor neuron that controls the **extensor**, a

generic term used to describe the muscle that functions opposite the flexor. In total, simultaneous activation of an extensor and inhibition of the extensor result in the rapid withdrawal of the hand.

In addition to communicating with the neurons involved in the withdrawal reflex, the nociceptive signals also send ascending projections through the white matter of the lateral spinothalamic tract.

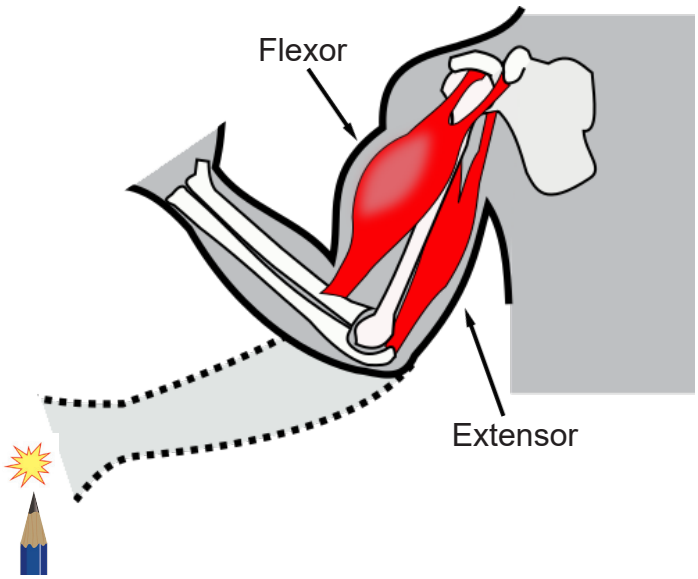


Figure 8.23 A painful stimulus produces a withdrawal reflex, which is driven by the action of skeletal muscles.

Clinical Connection: Referred pain

One of the warning signs that a person is having a heart attack is perception of pain in the shoulders or medial aspect of the left arm. This sensation is **referred pain**: the feeling of pain at a site separate from where an injury is located. Referred pain happens when the nervous system is unclear about how to process signals from an internal organ, and the brain interprets those afferent signals as bodily pain. Injury at different internal organs cause different patterns of somatic pain: liver injury may present as pain in the shoulder blades, lung cancer sometimes causes shoulder pain, and kidney stones can cause pain in the lower back, abdomen, and sides. Even the headache-like pain you get from a "brain freeze" from drinking cold liquids too quickly is a form of referred pain.

It is unclear what causes referred pain. The **convergence projection theory** suggests that the afferent projections from the internal organs and the nociceptive somatosensory neurons of the skin form

8.3.2 Proprioception

Raise your arms above your head. Even without seeing your arms, your nervous system has mechanisms that inform you about the location and position of your body parts, including how much your joints are bent. This sense is called **proprioception**, and is critically important for coordinated movement and motor reflexes that contribute to those tiny, rapid adjustments that are made while maintaining balance.

Proprioceptive information ascends through the spinal cord and into the brain via the dorsal column-medial lemniscus tract. Proprioception is also processed in S1, more specifically Brodmann areas 3a and 2.

There are two main neural systems that work together to give us our sense of proprioception.

Muscle spindles

Wrapped around the intrafusal skeletal muscle fibers is a series of nervous structures called **muscle spindles** which detect the status of the muscle. Each muscle spindle is

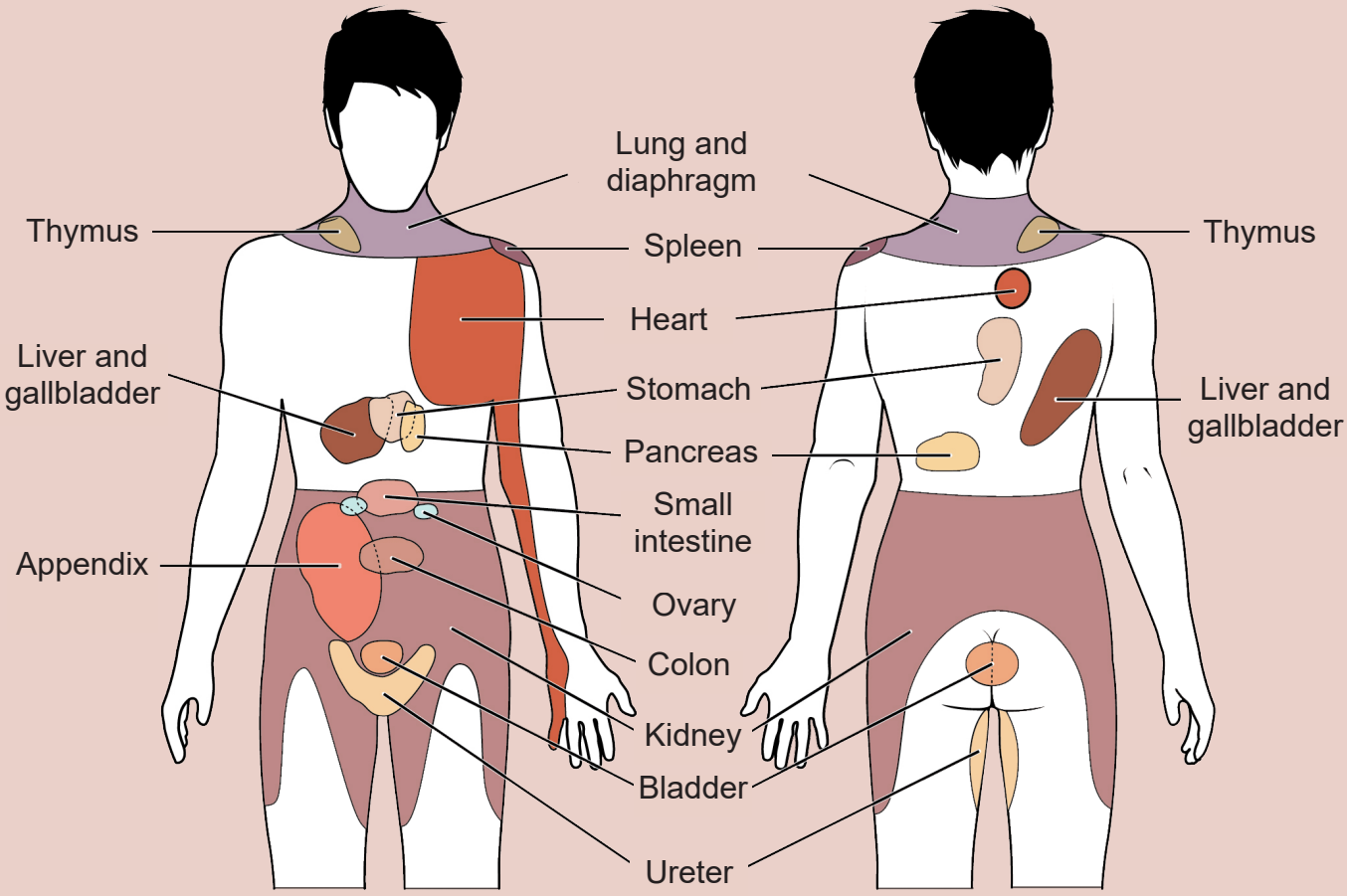


Figure 8.24 A map of referred pain sensations on the body and the corresponding visceral organ that may be triggering the pain.

synapses onto the same population of spinothalamic tract secondary neurons. When these secondary neurons project towards the brain, there is no ability to differentiate the origin of the signal coming from the primary neuron. The **central sensitization theory** suggests that the neurons of the spinal cord change in their excitability threshold with prolonged exposure to injurious stimuli. So, when an injured internal organ repeatedly sends signals into the secondary neurons, a host of neurotransmitters cause sensitization of the nociceptive signaling, which are now activated under non-noxious conditions.

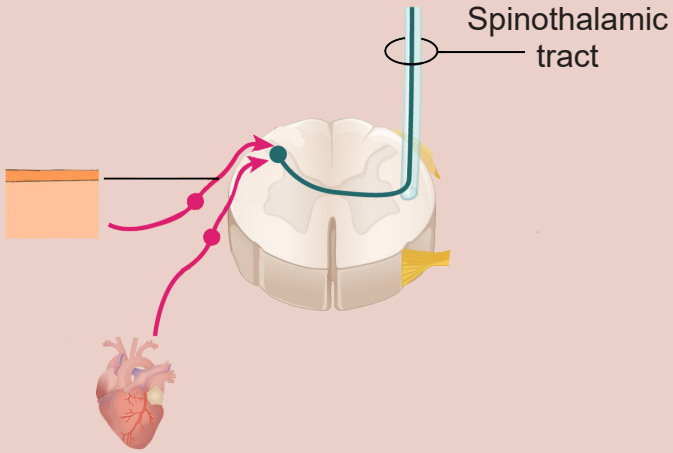


Figure 8.25 The convergence theory suggests that the signals from the internal organs and pain sensory inputs share the secondary neuron.

between 6 and 10 mm long, and spirals around the thickest part of the muscle fiber. When a skeletal muscle is flexed, it widens in diameter, becoming shorter and thicker. As it does, the muscle spindle wrapped around it also changes in shape. The muscle spindles communicate this information to the nervous system through the action of nonselective-cation mechanoreceptors that respond to physical distortion - muscle stretching increases firing rate, while muscle flexion decreases firing rate. Muscle spindle density is much greater among skeletal muscles that are used for very precise movements (like in the hand) compared to those used for coarse movements (like in the thighs).

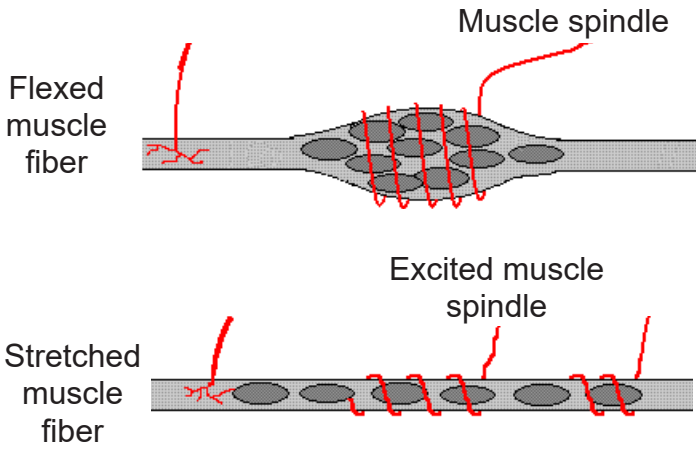


Figure 8.26 Muscle spindles wrap around the center of the inner muscle fibers and detect flexion.

There is also a motor function of these muscle spindles. Through a circuit of neurons in the spinal cord, they communicate with the **gamma motoneurons** that terminate at these intrafusal muscle fibers. The combination of the sensory and motor components allow the muscle spindles to function in the **stretch reflex** (also called the **myotatic reflex**), a spinal cord-mediated response to muscle stretch that causes flexion to prevent excess stretching, which can cause damage to the muscles.

The most-well known example of the stretch reflex is the **knee-jerk reflex** (or the **patellar tendon reflex**), which may be tested in a standard medical physical. Here, the patient sits on the end of the exam table with their knee bent and their lower leg dangling freely over the edge. The examiner uses a pointed rubber hammer to gently tap on the tendon that connects the kneecap (patella) to the muscle in front of the shin bone. When tapped, it causes a stretch of the **quadriceps**, the muscle on the top of the thigh. The muscle spindles detect this extension of the muscle, they communicate this information with an excitatory motor neuron of the spinal cord that causes a flexion of the quadriceps, which causes the foot to kick forward. The circuitry of this reflex is described as a monosynaptic reflex arc. (There is also a spinal cord inhibitory interneuron that receives muscle spindle inputs that decreases the activity of the hamstring, the muscle on the bottom of the thigh. In this reflex, the quadriceps is the flexor while the **hamstring** is the extensor.)

The neural circuitry responsible for the knee-jerk reflex is located in the L2, L3, and L4 regions of the spinal cord. If the knee-jerk reflex is weak or absent, there may be some degeneration or death of the nerves that communicate with the muscles, or possibly an injury to these specific areas of the spinal cord.

Golgi tendon organs (GTOs)

Made up of collagen fibers, GTOs are found at the insertion site between muscles and tendons. Signals from the GTO convey information about the amount of the tension that each set of skeletal muscles is experiencing as we move around. They also contribute to our detection of weight, as we lift something heavy

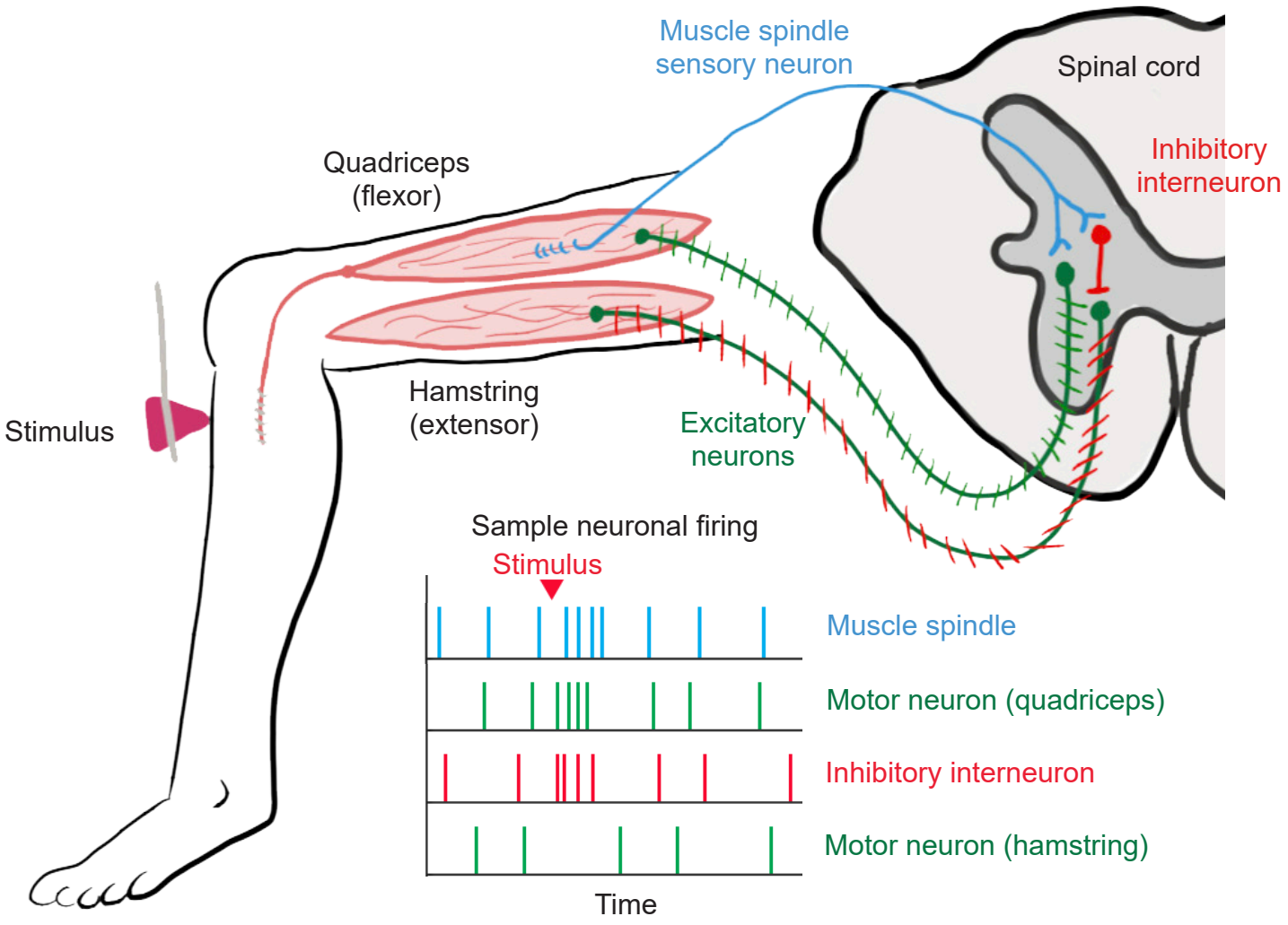


Figure 8.27 The knee jerk reflex is downstream of a series of neural circuits that start with changes in muscle spindle activity.

for example. Each GTO connects to about 20 muscle fibers and is about 0.5 mm long.

During muscle contraction, the tension in the GTO increases. Contained within the GTO are nonselective cation mechanoreceptors that open during this physical deformation (similar to those found in other aspects of the somatosensory system) causing changes in the excitable properties of the GTO. The outputs of the GTO communicate with interneurons found in the spinal cord. These interneurons in turn inhibit the motor neurons that innervate the muscle that is "pulling" on the tendons.

The GTO communicates via **Aα fibers**. These large diameter axons (15 μM) are heavily myelinated, and can transmit action potentials as fast as 100 meters per second, which are the fastest projections in the somatosensory system.

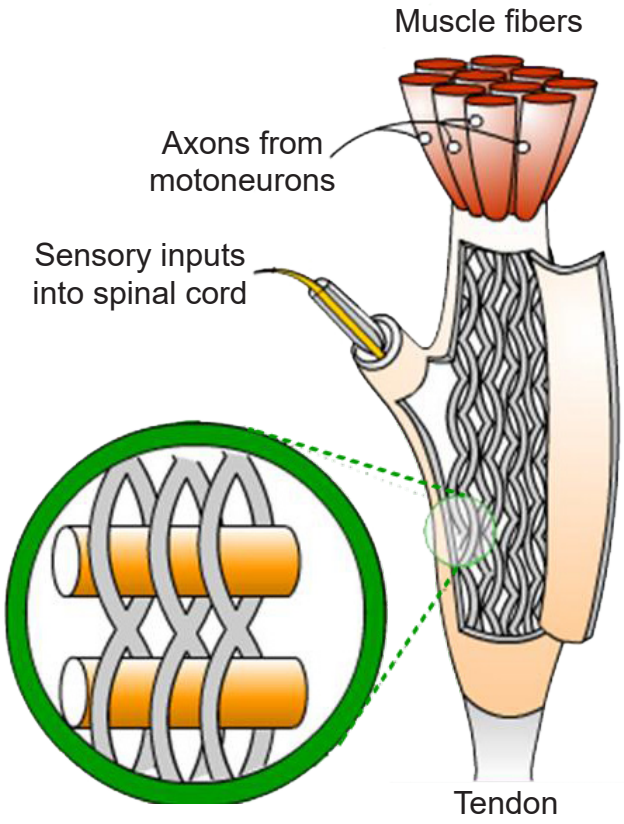


Figure 8.28 Anatomy of the Golgi tendon organ.

The Open Neuroscience Initiative is funded by a grant from the Vincentian Endowment Fund of DePaul University.

Image credits
Cover: https://pixabay.com/photos/guitar-ropes-instrument-music-wood-56916/
8.1 https://upload.wikimedia.org/wikipedia/commons/1/18/Sound_wave_physics.gif Modified by Austin
8.3 https://commons.wikimedia.org/wiki/File:Frame-level_Instrument_Recognition_by_Timbre_and_Pitch.pdf
8.4 https://upload.wikimedia.org/wikipedia/commons/e/ee/Anatomy_of_the_Human_Ear_blank.svg
8.5 https://commons.wikimedia.org/wiki/File:Auditory_ossicles-en.svg
8.6 https://wellcomecollection.org/works/fsg372tb
https://wellcomecollection.org/works/jfsbzy4g
8.7 https://upload.wikimedia.org/wikipedia/commons/2/2c/Hair_cell_action.png
8.8 https://wellcomecollection.org/works/hrvtw6fx
8.9 https://commons.wikimedia.org/wiki/File:Cochlear_Implant.png modified by Austin Lim
8.10 https://commons.wikimedia.org/wiki/File:Auditory_Pathway.png modified by Austin Lim
8.11https://commons.wikimedia.org/wiki/File:Frequency_mapping_in_human_ear_and_brain_-_10.1371_journal.pbio.0030137.g001-L.jpg modified by Austin Lim
8.12 https://commons.wikimedia.org/wiki/File:Vestibular_organs-_canals,_otolith,_cochlea.jpg modified by Austin Lim
8.13 https://commons.wikimedia.org/wiki/File:Vestibulo-ocular_reflex.PNG modified by Austin Lim
8.14 https://commons.wikimedia.org/wiki/File:1421_Sensory_Homunculus.jpg
8.15 https://commons.wikimedia.org/wiki/File:Front_of_Sensory_Homunculus.gif
8.16 https://commons.wikimedia.org/wiki/File:Mechanically_Gated_Channel.png modified by Austin Lim
8.18 https://commons.wikimedia.org/wiki/File:Blausen_0809_Skin_TactileReceptors.png modified by Austin Lim
8.19 https://commons.wikimedia.org/wiki/File:Compass_tool.svg modified by Austin Lim
https://commons.wikimedia.org/wiki/File:Skin_numbers.jpg modified by Austin Lim
8.20 https://commons.wikimedia.org/wiki/File:Histological_preparation_of_DRG.jpg
https://commons.wikimedia.org/wiki/File:Structure_of_sensory_system_(4_models)_E.PNG modified by Austin Lim
8.21 https://commons.wikimedia.org/wiki/File:1417_Ascending_Pathways_of_Spinal_Cord.jpg modified by Austin Lim
8.22 https://commons.wikimedia.org/wiki/File:Hyperalgesia_and_allodynia.tif modified by Austin Lim
8.23 https://commons.wikimedia.org/w/index.php?curid=53328677
https://pixabay.com/vectors/pencil-education-office-stationery-23648/ modified by Austin Lim
8.24 https://commons.wikimedia.org/wiki/File:1506_Referred_Pain_Chart.jpg
8.25 https://commons.wikimedia.org/wiki/File:1417_Ascending_Pathways_of_Spinal_Cord.jpg modified by Austin Lim
https://commons.wikimedia.org/wiki/File:Human_Heart.png
8.26 https://commons.wikimedia.org/wiki/File:Spindle.GIF
8.27 https://commons.wikimedia.org/wiki/File:Knee_jerk_reflex.png
8.28 https://commons.wikimedia.org/wiki/File:Tendon_organ_model.jpg

Chapter 9:

Sensation and Perception: The Chemical Senses

Ben Marcus, PhD (University of Chicago)
Austin Lim, PhD (DePaul University)

In the previous two chapters, we examined some of the ways we detect and perceive our physical environment, including through phenomena such as light, sound, and tactile sensations.

In this chapter, we will complete our tour of the senses by exploring how we sense and perceive chemical compounds. In particular, we will describe the neural processes underlying our senses of smell, taste, and the chemical status of our internal homeostatic state.

Our **olfactory system** (smell) is activated by compounds called **odorants**, while our **gustatory system** (taste) is activated by compounds called **flavorants**. These two chemical senses are closely intertwined; odorants affect how we perceive taste, and bad tasting foods often smell bad. Try eating a fancy meal with a stuffed-up nose, and you'll quickly notice that you lose the subtle complexities (and most of the joy!) of those foods.

Chemicals in our body, such as the amount of carbon dioxide in the blood or the presence of toxins in our digestive tract, are detected with our internal chemosensory systems. Many of these sensory systems are intimately tied with our autonomic nervous system (Chapter 2). Changes in chemical balance here results in unconscious or involuntary physiological changes to restore homeostasis.

Chapter 9 outline

9.1 Olfactory System

Olfaction is the ability to sense and perceive volatile chemicals that are suspended in the air. The typical human can distinguish up to 10,000 distinct odors, ranging from the sweet aroma of esters produced by apples and oranges, to the putrid smells of sulfurous compounds produced by skunks and rotten eggs. Because odors can drift along in the air, some chemicals can be detected long before the source is within eyesight: think about smelling a burning bonfire from miles away.

Smells affect our conscious behavior. They can motivate us to approach freshly-baked bread or avoid a rotting animal carcass. These chemicals serve as survival cues: bread gives us energy-rich carbohydrates while a decaying carcass can expose us to disease. Odorants can even affect our behaviors unconsciously:

Figure 9.1 A threatened skunk produces mercaptans, volatile chemicals that serve as a warning signal against to deter predators.

sleeping amidst the scent of a romantic partner increases the efficiency of our sleep, and subliminal exposure to citrus smells can lead people to clean their space more completely after eating a crumbly biscuit (an effect described in a study called “Smells Like Clean Spirit,” a nod to 90s punk band Nirvana).

Olfaction is one of the oldest functions we possess as animals. While our sense of smell has seemed to take a backseat to other senses relative to other animals (about 2% of the total mouse brain mediates smell, while only a scant 0.01% of the volume of the human brain is dedicated to this function), scientists now appreciate that our olfactory system is simply more specialized. For example, humans use the smells of sweat to clue us into the emotional state of others, and we can

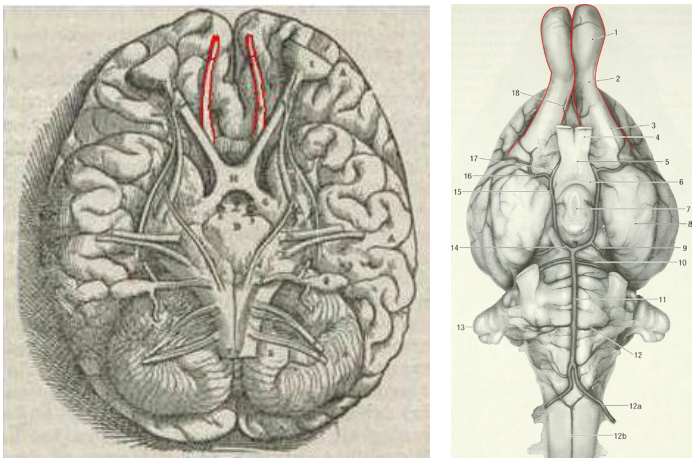


Figure 9.2 Ventral view of a human brain (left) and a woodchuck (right), showing that proportionally less brain tissue is dedicated to olfactory systems (red) in humans compared to other mammals.

even subconsciously detect sickness through body odor.

9.1.1. Anatomy of the Olfactory System

Your sensation of smell begins when an odorant traverses your nostrils and passes through the **nasal cavity**, an empty, air-filled space just behind the front of the skull. The dorsal-most portion of the nasal cavity is covered in a mucus-covered patch of tissue called the olfactory epithelium.

Embedded within the olfactory epithelium are the **olfactory receptor neurons** (**ORNs**, also called **olfactory sensory neurons**, or **OSNs**), neurons which begin processing smell. ORNs are bipolar neurons, with long dendritic projections protruding into the epithelium. Here, they peek out and contact the air, making them the only neurons that are directly exposed to the outside world. This, unfortunately, causes them to encounter all sorts of dangers such as toxins, particulates, and microbes. They are one of the few known populations of neurons where adult

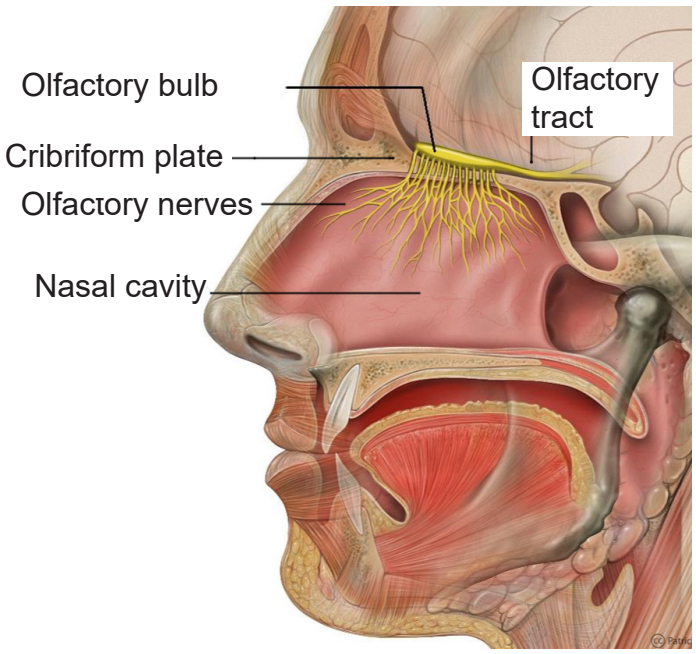


Figure 9.3 Cross section showing the major anatomical structures of the olfactory system.

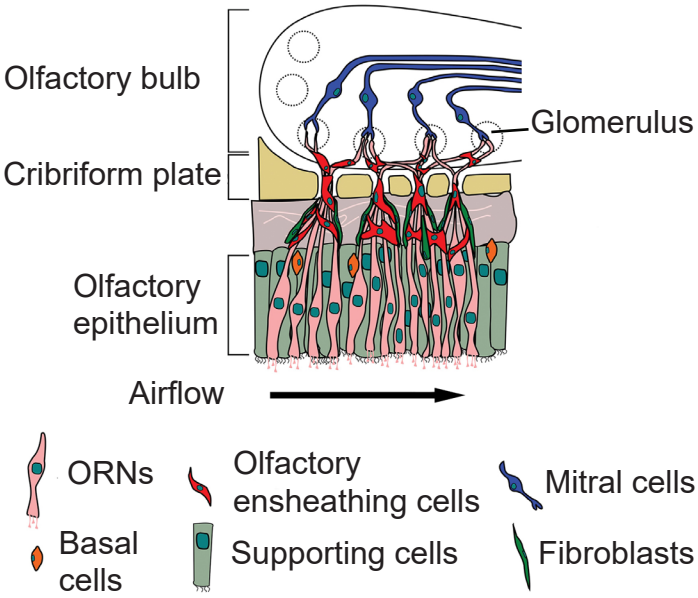


Figure 9.4 The cellular anatomy of the olfactory epithelium and olfactory bulb.

neurogenesis occurs, each having a lifespan in the range of 30 days to a year. The average human olfactory system has somewhere between 6-20 million ORNs.

Within the olfactory epithelium are a population of **supporting cells**, which function much like glial cells. They help dispose of dead and dying cells, metabolize pollutants, and may also help to physically maintain the epithelium.

On the dendrites of ORNs are **olfactory receptors**. It is estimated there are about 1,000 different genes (about 3% of the total human genome) that code for roughly 400 different olfactory receptors. Each ORN expresses only one type of olfactory receptor, and each receptor is believed to respond to activation by a different chemical. For example, the receptors expressed on ORN293s are highly responsive to cadaverine, the smell characteristic of death and decay. Even though scientists have identified the odorants that activate some receptors, most of the olfactory receptors are still "orphaned," and have not yet been matched with their corresponding odorant. The initial research into the genetics underlying these neurons earned Drs. Linda Buck and Richard Axel a Nobel prize in 2004, but there is still much work to be done.

Olfactory receptors are transmembrane G protein-coupled receptors that signal downstream effectors using the intracellular transduction molecule $\mathbf{G_{\alpha olf}}$. This protein complex is 90% similar to the stimulatory G-protein G_s, and likewise triggers activation of adenylate cyclase, elevating the intracellular concentration of cyclic AMP (see section 5.3 for a refresher on this signaling pathway). Activation of $G_{\alpha olf}$ causes depolarization, causing the ORN to fire action potentials.

ORNs encode the intensity of smells through the frequency of action potential firing, which changes in accordance with the concentration of odorant molecules. Imagine standing over a fresh-baked pizza and inhaling deeply. Due to the high concentration of odorant molecules in the air, several receptors will be

activated, leading to frequent neuronal firing. Now, imagine you are down the block from a pizza restaurant, getting only a slight whiff of those same scents. Here, the concentration of odorants is low, meaning that the ORNs fire less frequently.

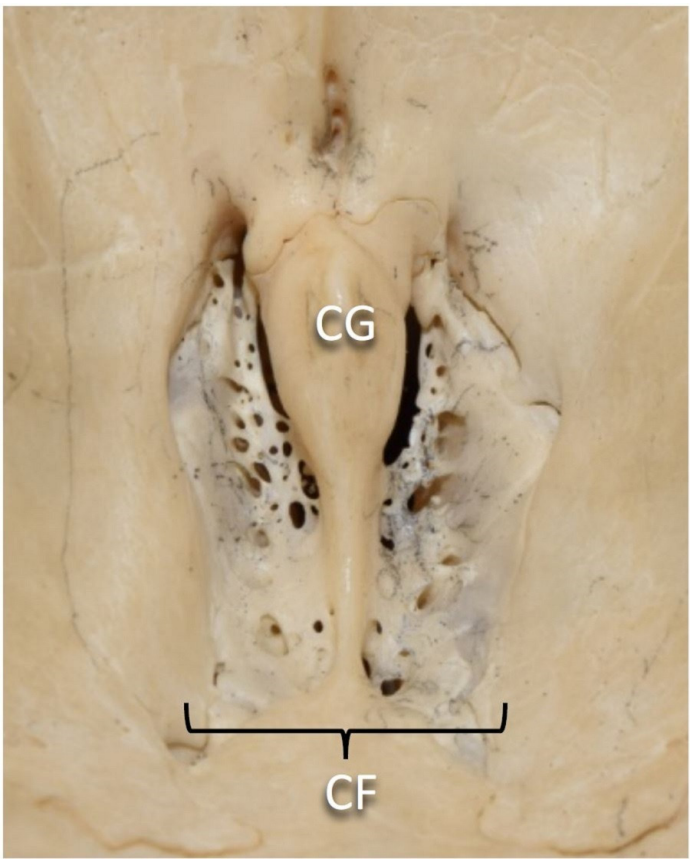


Figure 9.5 The cribriform plate at the base of the skull allows for the axons of the ORNs to pass into the brain.

Following activation of the receptors, the axons of the ORNs pass through the skull through a tiny series of holes at the **cribriform plate**, a sieve-like section of the ethmoid bone. These primary neurons form synaptic connections onto neurons in the **olfactory bulb**, the beginning of the **olfactory nerve (CN I).** Like the optic nerve, the olfactory nerve runs along the ventral surface of the brain.

The site of synaptic connectivity between the ORNs and the secondary neurons in the olfactory bulb is a highly specialized clump of tissue called a **glomerulus**. The typical human has a little under 2,000 **glomeruli**, and each glomerulus only receives inputs from ORNs that express the same type of olfactory receptors.

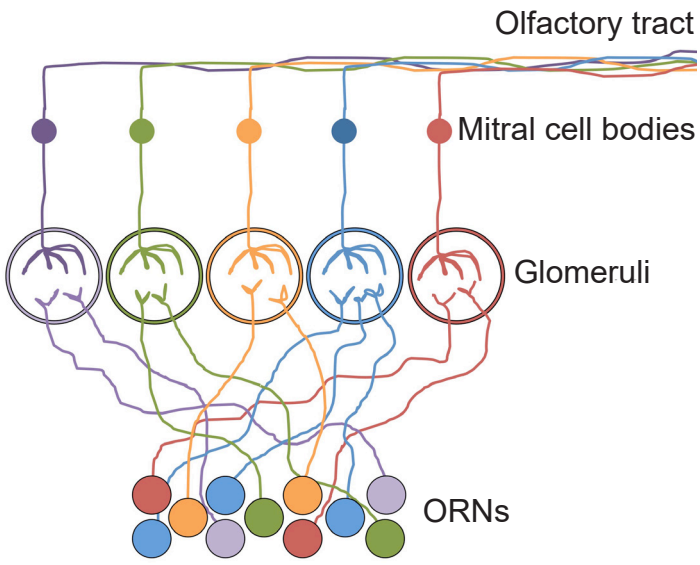


Figure 9.6 Communication between the ORNs and downstream neurons occur at glomeruli. Different colors indicate distinct populations of ORNs, which express only one type of olfactory receptor.

Within each glomerulus are the dendrites of the secondary neurons, which exist in two types: The **mitral cells** and the **tufted cells**. These two cell populations project axons directly into the olfactory cortex. This makes the olfactory system the only sensory system that does not pass signals through the thalamus before cortical processing.

There are also two types of inhibitory neurons that regulate this pathway: **granule cells** found within glomeruli, and **periglomerular cells**, which send axonal projections into the glomeruli, both help refine the synaptic processing of scent information using lateral inhibition, a system similar to that found in the retina of the visual system (Chapter 7).

Collectively, the olfactory cortex is made up of several different regions.

1. Piriform cortex

The **piriform cortex** is the main cortical input site for axonal projections from the olfactory bulb. In some species of mammals that rely very heavily on their sense of smell for survival, up to 10% of total cortical volume is piriform cortex. The outputs of the piriform cortex project into other structures of the olfactory cortex, as well as the **mediodorsal thalamus**, a relay center that contributes to learning and decision making.

2. Cortical amygdala

The **amygdala** is a part of the brain that helps mediate complex emotional states (chapter 15). Generally, it is subdivided into a few different subnuclei. One of them in particular, the **cortical amygdala**, receives strong inputs from the olfactory nerve. The cortical amygdala sends projections into the hippocampus, a brain structure critically important for the formation of new memories.

The strong neural connections between the olfactory nerve and amygdala are believed to be the reason why smell and memory are strongly linked. Think about a time when you caught a slightest whiff of a scent, and that scent mentally transported you to a distant place from a long time ago (this experience was also described by author Marcel Proust in his highly-regarded novel, *In Search of Lost Time*.) We form associations between specific smells and emotionally salient memories. Some odorants, like the volatile Maillard reaction products that are made as a product of cooking, remind us of the "good ole days" watching a parent bake or grill. Many bakeries take advantage of these positive memories and point their exhaust out into the street to entice passersby.

Unfortunately, the memories associated with smells can also have negative valence. For example, the smell of petroleum and oil fumes can trigger a sickness-like conditioned response in Gulf War veterans.

3. Entorhinal cortex (EC)

The **entorhinal cortex** is a small section of the medial temporal lobe. As with the cortical amygdala, the EC sends strong connections

Figure 9.7 The close connections between the olfactory nerve and amygdala + hippocampus may account for why smells are strong triggers of emotional memories.

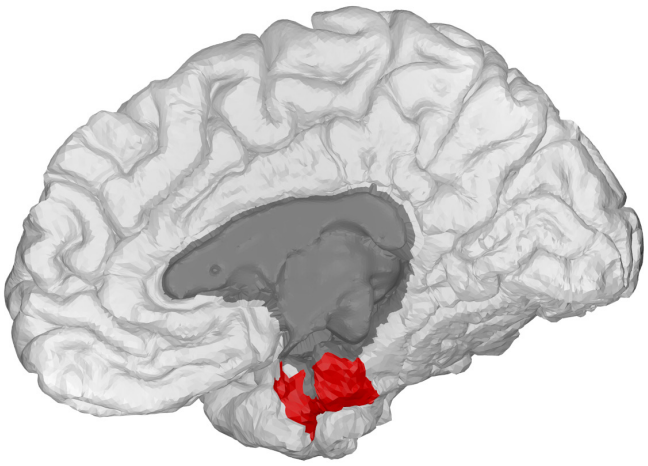


Figure 9.8 The entorhinal cortex (red) is found along the ventral surface of the medial temporal lobe.

into the hippocampus, indicating that olfactory signals contribute to the strong associations formed between smell and memory through EC, as well.

Other than receiving and processing olfactory signals, the entorhinal cortex is also involved in spatial navigational tasks.

4. Orbitofrontal cortex (OFC)

As the name implies, the **orbitofrontal cortex (OFC)** is found just behind the orbit, the bony socket of the skull where the eyes sit. OFC is found at the ventral surface of the frontal lobe. Circuits in this brain area function as an integration site for sensory inputs, since it also receives projections from visual, taste, and somatosensory cortices.

The full extent of the OFC is still under examination, but it is also implicated in decision making and social behaviors.

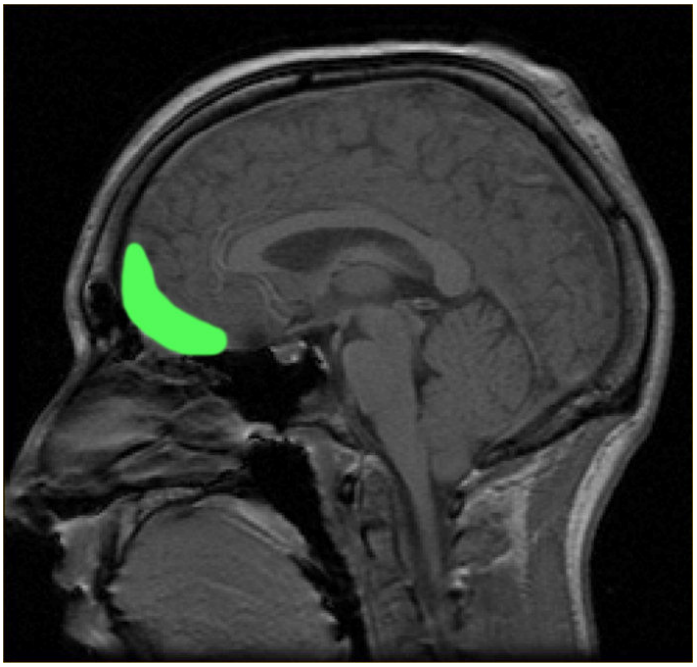


Figure 9.9 MRI image of a parasagittal section of the brain showing the orbitofrontal cortex (green).

Do humans respond to pheromones?

Some species secrete chemicals which influence the behavior of other members of the same species. These **pheromones** trigger complex responses in others, such as motivating social insects like ants to follow meandering trails towards faraway food sources or causing a group of bees to swarm and attack a predator.

There is still debate over whether or not humans produce and release pheromones, and whether or not other humans are sensitive to these signals. Many nonhumans, notably snakes and dogs, have a specialized structure in their nasal cavity called the **vomeronasal organ**, which detects organic compounds produced by predators and reproduction-related hormones produced by the opposite sex. These signals are sent into the brain via **cranial nerve 0**, which can then trigger behavioral changes. Humans have these anatomical structures, but they are not believed to carry any functional information. However, other evidence suggests that smelling certain compounds influences mood, alters neuroendocrine signaling, and even affects mate selection.

9.1.2 Disorders of the Olfactory System

Like other sensory systems, the structures involved in olfaction can be injured. An injury to the olfactory system can result in **hyposmia**, a reduced ability to smell, or **anosmia**, a complete loss of smell.

The most common insult to the olfactory system is simple nasal congestion, a temporary, physical blockage of the entry to the nasal cavity

that decreases airflow and, therefore, the number of particles that reach the olfactory epithelium. Congestion can be caused by allergies, the common cold, upper respiratory bacterial or viral infections, or sinus infections. Hyposmia is also one of the main neurological symptoms of COVID-19.

Hyposmia is common among healthy, older adults, affecting about half of the population between 65 and 80 years old. As a person ages, spontaneous calcification causes the holes in the cribriform plate to shrink, which can impinge on and damage ORN axons.

Hyposmia can also be caused by abrupt head injuries. The ORN axons that project through the holes of the cribriform plate are particularly sensitive to blows to the head.

Neurodegenerative disorders, such as Parkinson's disease and Alzheimer's disease, contribute to smell deficiency. Usually, hyposmia precedes the major clinically observed symptoms of these disorders, hinting that smell deficiency may serve as an early diagnostic biomarker.

Another olfactory deficit, **phantosmia**, is when a person perceives "phantom" scents, or in other words, experiences an olfactory hallucination. Phantosmia may be triggered by a temporal lobe seizure or a stroke. It can also be caused by a brain tumor affecting the olfactory nerve (CN I), or the subsequent surgical removal of the tumor, leading to injury. Schizophrenia, a psychiatric condition characterized by auditory hallucinations, may also cause phantosmia.

9.2 Gustatory System

Your **gustatory system**, which mediates your sense of taste, helps you walk the line between health and illness. It guides you towards foods that are energy rich, and keeps you away from food that could make you sick. The specific taste modalities, salt, sweet, sour, bitter, and umami, all support this balance. Sweet foods taste good because foods rich in sugar, such as fruit and wheat, contain large amounts of usable energy, and humans have evolved to find these foods appetizing. In contrast, toxic compounds are often bitter, causing you to respond with feelings of disgust.

To further refine this balance, a tastant's appeal varies depending on its concentration. For example, we find a small amount of sourness desirable in fruit and candy, but aversive in milk. And while bitterness is often a sign of toxicity in food, mild bitterness contributes positively to the taste of coffee and chocolate. Sugar can make candy, pastries, and drinks very appealing, but too much can make the food barely palatable. The same applies to salt: the right amount makes potato chips and pretzels delicious, but it also causes sea water to make you vomit. All of these reactions are adaptations humans have developed to maintain homeostasis in our body.

9.2.1 Anatomy of the gustatory system

Lingual papillae (singular; **papilla**) are the large anatomical structures that give the tongue its characteristic rough surface. These structures can be seen with the unaided eye. Each papilla contains up to a hundred **taste buds**, which are onion-shaped taste receptors. Taste buds can also be found on the palate and in the throat. In total, a person has about 10,000 of these receptors, but the number varies by age: Taste bud concentration in the mouth peaks in childhood and decreases throughout adulthood.

Within each taste bud are approximately 100 **taste receptor cells**. Technically, these cells are not neurons, but are instead derived from specialized epithelial cells. These cells reach toward the apical tip of the taste bud and sprout thin projections called **taste hairs**, which extend into taste pores. These tiny pockets at the apical tip of the taste bud are where taste hairs meet saliva. Taste buds also contain **basal cells**, which reproduce to form **supporting cells**, and over time, these mature further into taste receptor cells. Taste receptor cells turnover every 8–22 days.

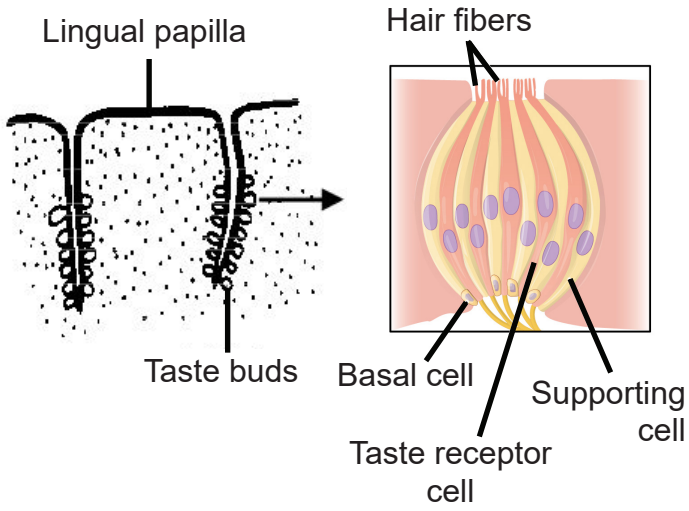


Figure 9.10 On each lingual papilla are taste buds (left). A magnified view of a taste bud (right).

Taste receptors cells are responsible for sensing and conveying information about taste in accordance with the main taste **modalities** (salty, sweet, sour, etc). Like how the olfactory system is organized, taste receptor cells do not express receptors for more than one taste modality - a salty

taste receptor cell only carries salty information, a sweet taste receptor cell only pass on sweet information, and so on.

Taste receptor cells communicate with afferent gustatory neurons. These gustatory nerve fibers originate from three of the twelve cranial nerves.

- The nerve fibers from the anterior two-thirds of the tongue are part of the facial nerve (CN VII).
- The posterior third of tongue sends information through the glossopharyngeal nerve (CN IX).
- The back of the palate and the throat can send taste-related signals through the vagus nerve (CN X).

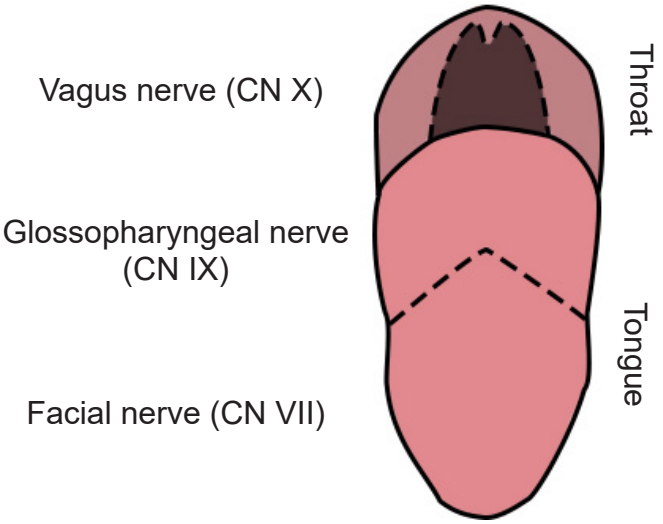


Figure 9.11 Different cranial nerves communicate information from different areas of the gustatory system.

These neurons then form synapses on second order neurons in the rostral medulla in an area called the **solitary nucleus** (or gustatory nucleus) in the medulla oblongata. Unlike most other sensorimotor systems, the gustatory system sends ipsilateral projections into the CNS; that is to say, taste information from the left half of the tongue gets represented in the left hemisphere of the brain, and visa versa.

From the medulla, these neurons send axonal projections into the **ventral posteromedial (VPM) nucleus** of the thalamus. (These represent the third-order neurons of the gustatory system.) These neurons send projections widely across several areas of the cortex.

The gustatory cortex is the beginning of the processing of taste perception. It is made up of two different parts: the anterior end of the insular cortex and the frontal operculum of the frontal lobe. These neurons convey information such as the taste modality and intensity.

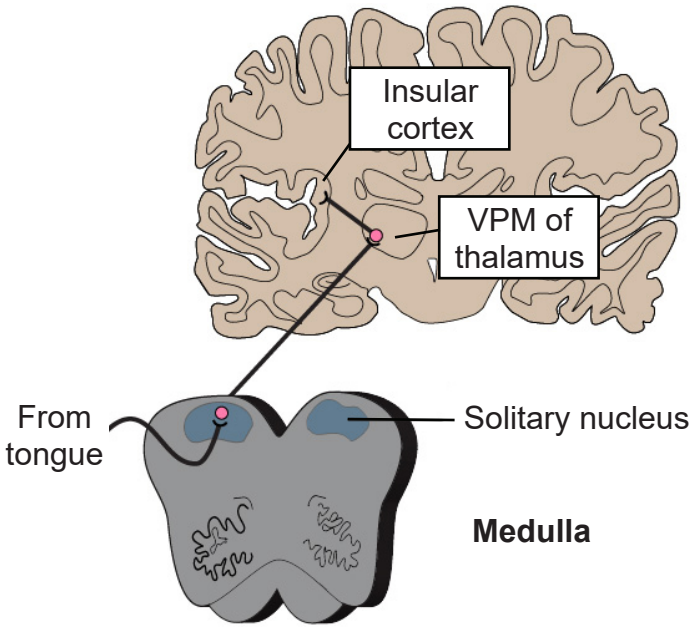


Figure 9.12 Afferent signaling from the tongue passes through medulla, then thalamus, then cortex.

9.2.2 Taste modalities of the gustatory system

Currently, it is believed that humans sense five basic tastes: Salty, sour, sweet, bitter, and umami. Salty and sour taste are both mediated by ionotropic taste receptors, while sweet, bitter, and umami taste are mediated by metabotropic

taste receptors. Besides varying by type, these receptors are also not distributed evenly across the surface of the tongue. For example, the tip of the tongue is most sensitive to sugars, while the back of the tongue is most sensitive to bitter compounds.

Salt

Sensation of salt taste is primarily driven by Na^+ ions. When you sprinkle a little table salt (NaCl) onto your tongue, it dissolves in saliva and the free sodium ions can passively influx into salt taste receptor cells through **epithelial sodium channels (ENaCs).** This movement of positively charged Na^+ ions causes depolarization of the taste receptor cell, just like in neurons. This depolarization activates voltage-gated calcium channels, prompting neurotransmitter release that, in turn, activates the afferent gustatory nerve fibers.

Salty foods elicit a biphasic response depending on concentration. Foods cooked with a low concentration of salt taste bland and are not very appetizing; however, high salt concentrations elicit a strong aversive reaction - imagine how disgusted you were when you first tried to cook and were overly generous with the salt! The taste of salt is typically desirable at salt concentrations lower than 100 mM.

Additionally, the appeal of salt at a given moment depends on our body's need for salt at the time. Several hormones such as the appetite-stimulating hormone **ghrelin** contribute to regulating the concentration of salt in the body by mediating Na^+ absorption. Current salt levels can also impact appetite for salt. For example, in chronically-sodium deprived animals, high salt solutions are highly rewarding.

Why are we so sensitive to the taste of salt? As it turns out, both Na^+ and Cl^- are essential nutrients. They are critical for maintaining blood volume and pressure, for regulating body water, for maintaining muscle contractions, and mediating action potentials. Cl^-, in particular, helps maintain a healthy pH balance. But for these functions to be performed optimally, salt must be present at a specific range of concentrations in the body.

Sour

Sensation of sour taste is mediated by proton-selective channels, which is why acids are sour (acids are low pH, which means a high concentration of H^+). Like neurons, taste cells can depolarize with the entry of positively-charged ions, which influences signaling passed through the downstream nerve fibers.

The purpose for our ability to sense acids in our food is under debate. Sour tastants do not inherently provide any nutritional value, except in the case of Vitamin C. Humans and other higher primates cannot synthesize Vitamin C on their own, so it's possible that we evolved to find a combination of sourness and sweetness attractive enough to consistently consume Vitamin C-rich fruits. However, sour can be aversive, motivating us to avoid spoiled or unripe foods that might contain pathogens.

Sweet

Sweet taste transduction is carried out by the activation of G-protein coupled receptors expressed on the sweet taste receptor cells. The most likely candidate for the primary sweet-sensitive GPCR is a heterodimeric receptor with the subunits **taste receptor type 1 member 2** (**TAS1R2** or **T1R2**) and **taste receptor type 1 member 3** (**TAS1R3** or **T1R3**). All sweet substances, including carbohydrates and artificial sweeteners, activate the taste pathway through these T1R2/T1R3 receptors. Interestingly, these

receptors are also found in other parts of the body, including in the brain, pancreas, gastrointestinal tract, and fat tissue. While its function in these areas is not known, it's theorized that it plays a role in glucose homeostasis.

Sugars like glucose and sucrose are essential for the survival of a species, since they are the main source of cellular energy. Therefore, our ability to detect sweetness plays a central role in regulating how much energy we take into our bodies. Of all the taste modalities, sweet is the strongest driver of food selection.

Bitter

Bitter taste is sensed via the **T2R receptor**. From here, the molecular signaling cascade branches off into two distinct pathways. The first pathway is similar to the sweet taste pathway; it recruits a similar set of signaling proteins en route to the afferent gustatory nerves, and knocking out any of the receptors along the sweet taste transduction pathway also attenuates bitter taste transduction. The second pathway uses a taste-specific phosphodiesterase, an enzyme that lowers the intracellular concentration of the signaling molecule cAMP. Some bitter compounds bypass the T2R receptor altogether, permeating the cell membrane and directly activating the GPCR. The bitter compound quinine, found in tonic water, does this. In high concentrations, quinine can cause disgust and vomiting.

Bitter taste prompts avoidance because bitter tastes often indicate toxicity. For example, cucurbitacins, a group of bitter-tasting compounds found naturally in gourds, including cucumbers and pumpkins, are believed to have evolved as an antiherbivory defense mechanism to prevent predation. The exact mechanisms that connect toxic compounds to disgust and vomiting are unknown; it's possible that there is not a causative relationship among these reactions to toxins, but rather that humans evolved to respond to toxic compounds at multiple levels in parallel: in the mouth, through bitter taste, in the stomach with vomiting, and in the colon, with diarrhea.

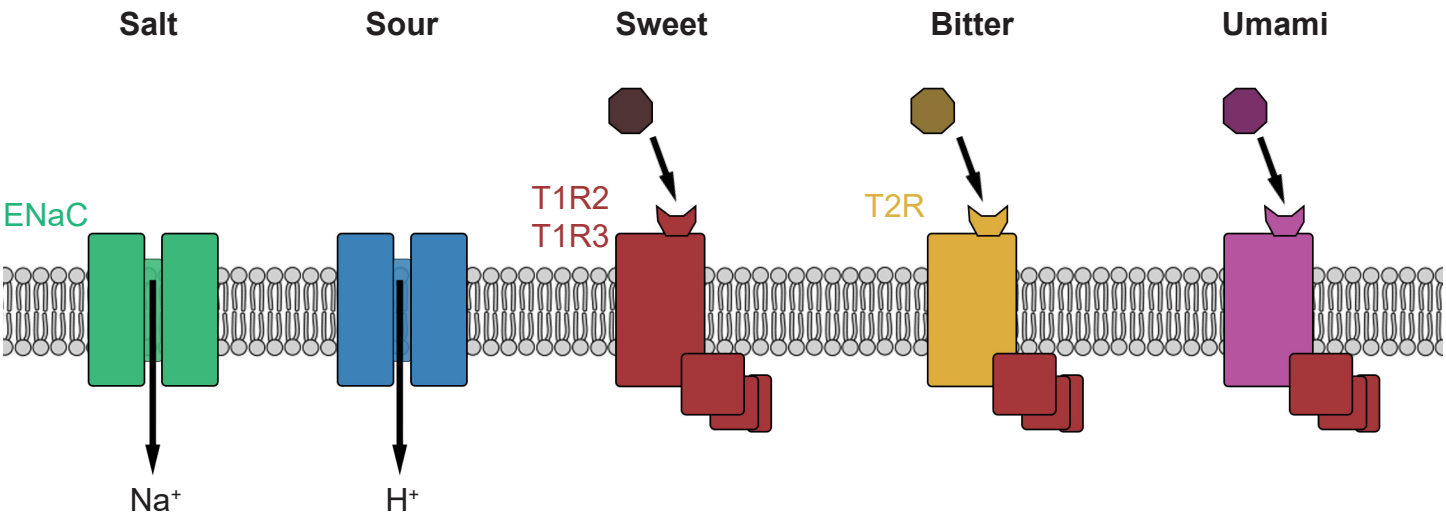


Figure 9.13 Taste modalities signal using different populations of receptors. Salt and sour are detected with ionotropic receptors, while sweet, bitter, and umami are sensed with metabotropic receptors. Many of the metabotropic receptors signal inside the cell using the α-gustducin molecule.

Umami

Umami is the taste of savory deliciousness, such as the taste of rich chicken broth, a perfect medium-rare steak, or aged cheese. The word is derived from the Japanese word *umai*, whicih means "delicious." Like sweet flavor, umami taste is mediated by a heterodimeric metabotropic receptor. Umami is signaled when a molecule of **glutamate** (chemically the same as the neurotransmitter!) binds to T1R1/T1R3 receptors.

The intracellular signaling transduction process is similar for sweet, bitter, and umami sensation. Tastants of these modalities all activate GPCRs that use the G protein **α-gustducin**. This activation increases the activity of phospholipase C-β2 (PLC-β2), which, in turn, activates the inositol 1,4,5-triphosphate (IP3) receptor, causing the release of calcium into the cytoplasm. The calcium opens the transient receptor potential cation channel subfamily M member 5 (TRPMP5), which causes the taste cell membrane to depolarize, generating an action potential. This causes the release of ATP into the synapse, which activates afferent nerve fibers to signal the presence of these tastants.

Through evolution, we have come to prefer the taste of umami because glutamate is a byproduct of cooking food. Cooking foods changes their chemical properties, thereby improving digestion, reducing toxicity, and increasing absorption of nutrients. Glutamate is one of the byproducts in the process of heating a food, and so it benefits us to appreciate these flavors.

Fat

A triglyceride is a chemical made up of three fatty acids bonded to a glycerol molecule. They are a large part of our diet, and are commonly found in animal fat and butter. Triglycerides contribute to the "mouthfeel" of foods, giving foods a creamy, rich quality. However, triglycerides themselves don't have a taste.

In 2015, researchers found that humans might be able to sense fatty acid chain molecules in the mouth. It turns out that in their study participants could distinguish between a control drink, a bitter drink, a sour drink, and drinks containing fat. Participants initially grouped the fatty drinks with the bitter and sour tastes, suggesting that fat by itself is an aversive flavor, despite being high in energy content.

The research on this taste modality, called **oleogustus**, is still in its infancy compared to that of the other taste modalities. Researchers have not elucidated the mechanisms underlying this pathway.

Spicy

It's no accident that the word "hot" is used to describe both the temperature and the spiciness of a food. After eating that delicious plate of curry, your body may exhibit a strong somatic response: intense salvation, a flushing of the skin, sweating, and sometimes even crying. As it turns out, the pain of eating a hot pepper is similar to experiencing other forms of physical injury. Taste receptor cells send signals to the brain via C fibers, the same neuronal pathway that carries afferent painful information (chapter 8).

Our ability to detect spicy flavors originates at the **TRPV1 receptor**, a nonselective cation channel. Upon activation, these receptors cause depolarization of the taste cell. Peppers contain the compound **capsaicin**, a potent activator of the TRPV1 receptor. These TRPV1 receptors are also temperature sensitive, opening at around 43C. Because these receptors can be activated by either chemical ligands or high temperatures,

biting into a ghost pepper causes a similar sensation as if your tongue was literally being burned - the “heat” of spicy foods is more than just a colloquialism.

TRPV1 receptors are not only activated by capsaicin. Ethanol, for example, can also activate these receptors, which is why a shot of hard liquor causes a painful, burning sensation. Since the TRPV1 receptors permit proton movement into the cell upon activation, acidic conditions potentiate receptor activation, which also accounts for why acids often taste “hot.”

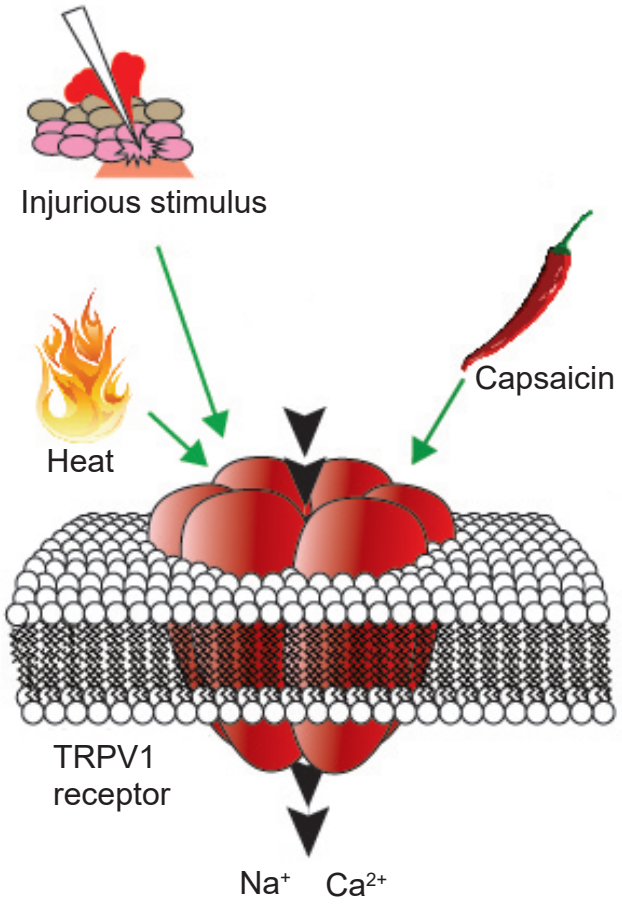


Figure 9.14 TRPV1 recptors are activated by different stimuli, ranging from physical heat to the chemical capsaicin.

9.3 Internal Chemosensory Systems

In addition to detecting chemicals with our noses or mouths, we also host a variety of chemosensory systems that sense various conditions about our internal environment. These systems contribute to the maintenance of homeostasis. Whenever the body is pushed out of it's ideal operating range, these chemosensory systems respond by reflexively adjusting chemical absorption or behavior.

9.3.1 Respiration

Neural control of the respiratory system originates in several circuits within the hindbrain. Of particular relevance is the **medulla**, the inferior-most segment of the brain stem. These complex circuits communicate with descending motor signals that are critically important for respiration through the action of two main nerves. The main driver of respiration is the **phrenic nerve**, which is the only nerve that innervates the diaphragm. The other drivers of respiration are the **intercostal nerves**, which innervate the intercostal muscles, the set of accessory respiratory muscles found between the ribs that expand the chest cavity during inhalation. People with spinal cord injury at the level of C5 or higher, which results in damage to the phrenic nerve, may need to be put on a ventilator. These circuits express opioid receptors, which is why opioid overdose can lead to fatal respiratory depression.

Regular respiration is an autonomic function. When CO2 levels rise (a condition called **hypercapnia**), these hindbrain neurons drive increased respiratory rate, which helps the body expel excess CO_2.

Respiratory patterns are also regulated homeostatically to restore a healthy level of pH

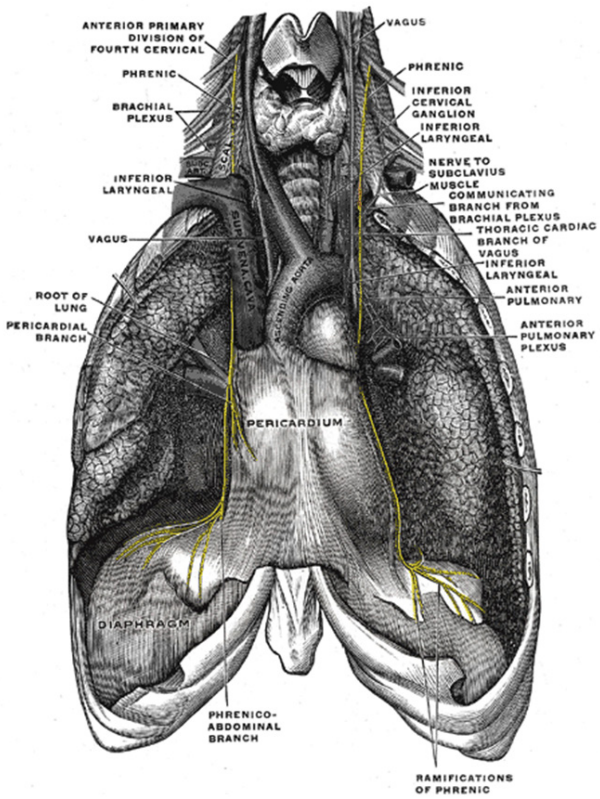


Figure 9.15 The phrenic nerve (highlighted yellow) is the main driver of respiration.

in the blood. The pH of CSF is essentially a proxy measure for CO_2 in the blood: CO_2 diffuses easily across the blood brain barrier into the CSF. Once there, CO_2 quickly reacts with H_2O to form carbonic acid, which then dissociates into a bicarbonate ion and a H^+ ion. Because of this chemical reaction, when blood CO_2 is elevated, so is the concentration of H^+ (low pH) in CSF.

Central chemoreceptors detect changes in the pH level of the CSF by sensing H^+ ions, which enter the cells through **acid-sensing ion channels (ASICs)**. When ASIC-expressing neurons detect pH levels less than 7, they send signals to the nerves that mediate diaphragm and intercostal muscle activity to increase respiration. This increases the exchange of CO_2 out of the

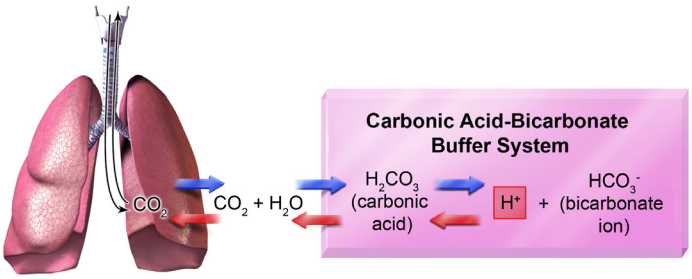


Figure 9.16 The nervous system regulates blood pH using the carbonic acid-bicarbonate buffer system.

lungs, shifting the pH of the blood back towards more physiological levels of 7.4.

9.3.2. Vomiting

Vomiting (or **emesis**) is a rapid contraction of respiratory and abdominal muscles, compressing the stomach, thereby expelling stomach contents through the esophagus. Vomiting is often preceded by **nausea**, the unpleasant sensation of stomach discomfort.

Although aversive and painful, vomiting can be a natural and healthy protective response. For example, when toxins are produced during bacterial gastroenteritis (food poisoning), it is beneficial to expel the spoiled or rotten food from the stomach to minimize further exposure to bacterial toxins.

The neural signals that lead to vomiting originate at the afferent inputs of the **vagus nerve (Cranial Nerve X),** found in the intestinal tract. These ascending inputs form connections within the **dorsal vagal complex (DVC),** a series of nuclei found in the medulla of the brain stem. A region within the DVC that mediates the vomiting response is **area postrema (AP)**, which is found on the floor of the fourth ventricle. Within AP is the **chemosensory trigger zone**, which is dense with neurons that sense the presence of various chemicals. The AP is considered to be a **circumventricular organ**, meaning that it is not isolated from the blood by a blood brain barrier. Instead, toxins and other large molecules in the blood are able to influence AP neurons directly. Additionally, because they are bathed by CSF, they can also sense the presence of toxins in CSF.

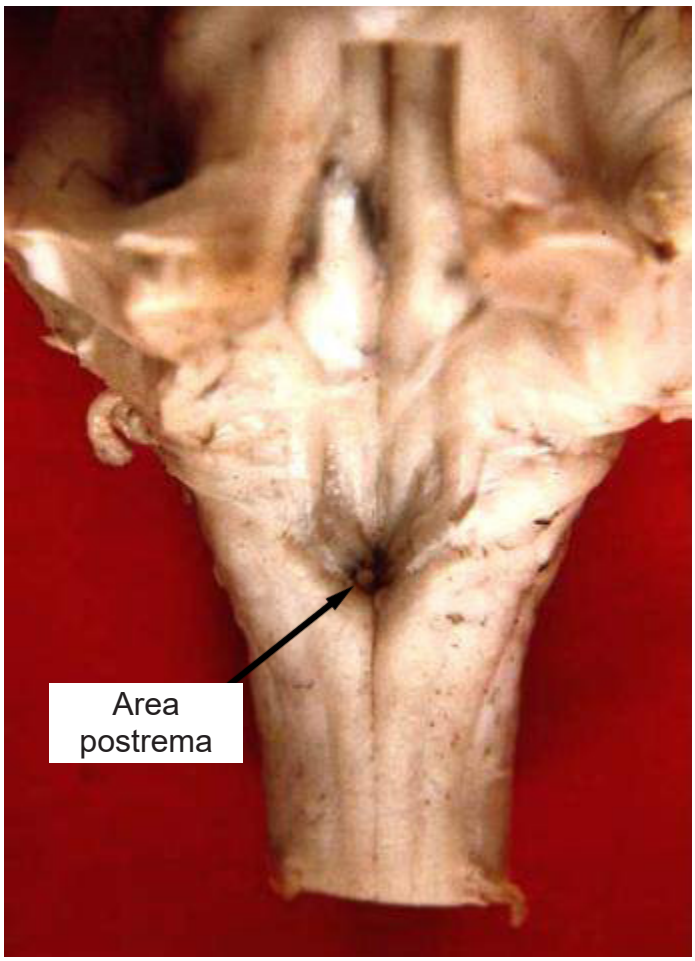


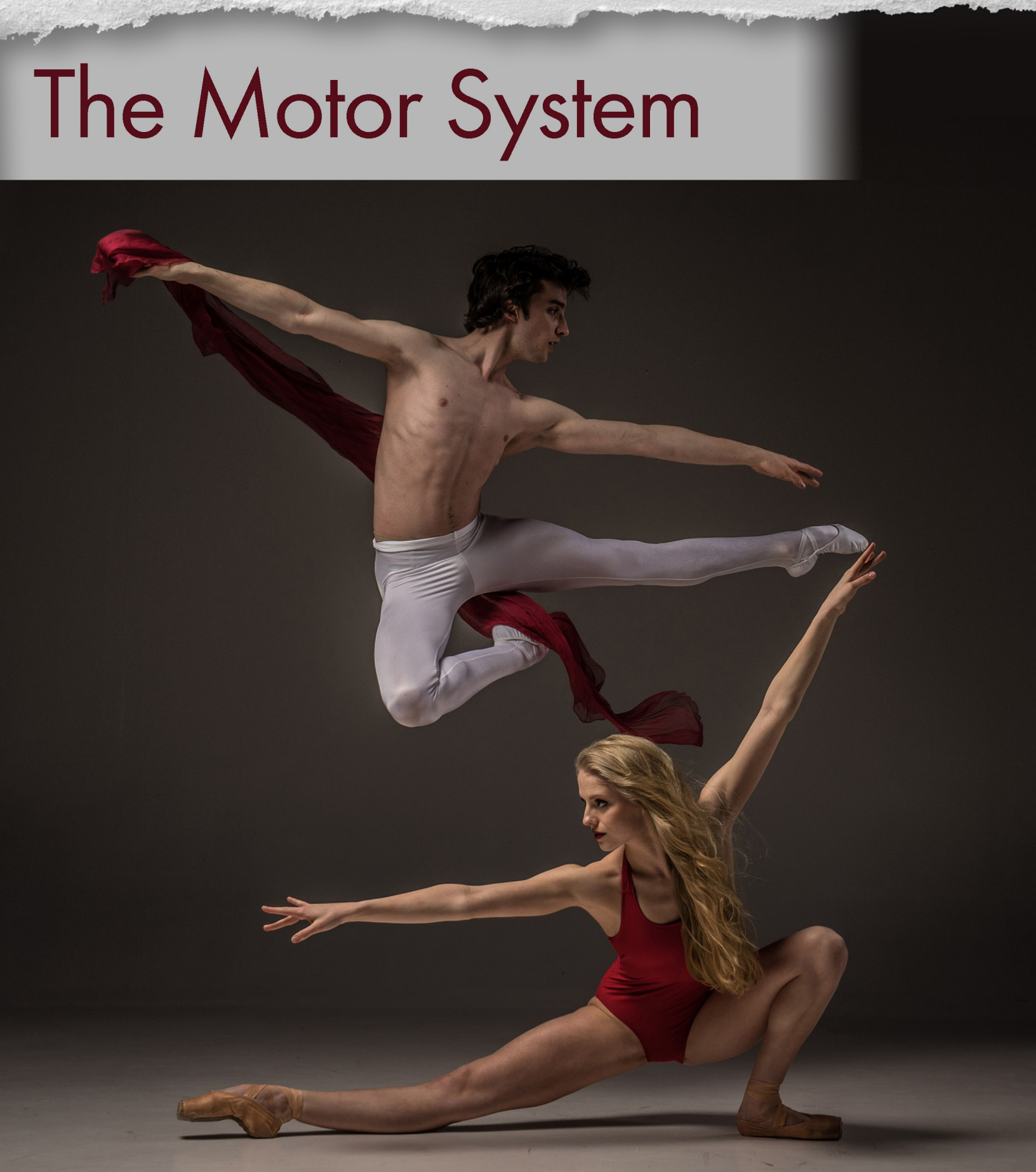
Figure 9.17 Posterior view of the brain stem showing area postrema, the emetic center of the brain.

Cover: https://pixabay.com/photos/pork-ribs-dinner-pork-food-meat-2157179/
9.1 https://commons.wikimedia.org/wiki/File:Southern_spotted_skunk.jpg
9.2 https://commons.wikimedia.org/wiki/File:1543,Vesalius%27OlfactoryBulbs.jpg
https://commons.wikimedia.org/wiki/File:Anatomy_of_the_woodchuck_(Marmota_monax)_(2005)_(18007398570).jpg
olfactory structures outlined by Austin Lim
9.3 https://commons.wikimedia.org/wiki/File:Head_olfactory_nerve.jpg
9.4 https://commons.wikimedia.org/wiki/File:Location_of_olfactory_ensheathing_cells_(OECs)_within_the_olfactory_system.png modified by Austin Lim
9.5 https://commons.wikimedia.org/wiki/File:Cribriform_plate_Close-up_view.png
9.6 https://commons.wikimedia.org/wiki/File:Olfactory_Sensory_Neurons_innervating_Olfactory_Glomeruli.jpg modified by Austin Lim
9.7 https://pixabay.com/photos/grain-bread-bread-rye-bread-cut-3135224/
9.8 https://commons.wikimedia.org/wiki/File:Entorhinal_-_DK_ATLAS.png
9.9 https://commons.wikimedia.org/wiki/File:MRI_of_orbitofrontal_cortex.jpg
9.10 https://commons.wikimedia.org/wiki/File:Anatomy_and_physiology_of_animals_Taste_buds_on_the_tongue.jpg modified by Austin Lim
https://commons.wikimedia.org/wiki/File:1402_The_Tongue.jpg modified by Austin Lim
9.12 https://openbooks.lib.msu.edu/neuroscience/chapter/taste-central-processing/ modified by Austin Lim
9.14 https://commons.wikimedia.org/wiki/File:EMA401_Mechanism_of_Action.jpg
9.15 https://commons.wikimedia.org/wiki/File:Gray806.png phrenic nerve highlighted by Austin Lim
9.16 https://commons.wikimedia.org/wiki/File:Buffer_Part_1.png modified by Austin Lim
9.17 https://commons.wikimedia.org/wiki/File:Human_caudal_brainstem_posterior_view_description.JPG modified by Austin Lim

The Open Neuroscience Initiative is funded by a grant from the Vincentian Endowment Fund of DePaul University.

Chapter 10:

The Motor System



Alexander Rajan, PhD
Austin Lim, PhD (DePaul University)
Andrea Gaede, PhD
Editor: Katie Willis, PhD (University of Oklahoma)

The **motor system** refers to the nerve cells that are used to control our body. The two key roles of the motor system are to plan, control, and execute **voluntary** (deliberate) movements, and to control **involuntary** (subconscious or automatic) functions, such as digesting food.

The motor system is sometimes described as a top-down process: in a voluntary movement, neural activity in the frontal lobe sends commands down to motor neurons located in the brainstem or spinal cord, which in turn activate muscle groups.

In reality, motor control is more of a loop, rapidly communicating between the sensory cortex and motor cortex. Sensory information about limb position, posture, and objects in contact with the skin inform the descending motor plan. Simultaneously, the motor plan provides predictions about upcoming movement. Without knowing where you are, it's difficult to plan a route to your destination, and getting feedback along the way helps your brain know if it needs to adjust the plan.

Accordingly, think about motor control as a set of nested loops, where motor and sensory processes are closely intertwined. The main loop is the descending (from brain to muscle) motor command and the ascending (from muscle to brain) sensory feedback. Examples of nested loops within this circuit could include spinal cord-mediated reflexes and communication between areas within the brain.

We will describe the process of motor control by first describing the signals that originate in the brain, then tracing that signal down through the brain stem, spinal cord, the neuromuscular junction, and finally the muscles.

Chapter 10 outline

10.1 Motor Control in the Brain

All **voluntary** (or **non-reflexive**) movements begin as signals in the brain. Specifically, the neurons involved in motor control are primarily found in the frontal lobe of the telencephalon, which includes areas such as primary motor cortex (or M1), premotor cortex, and supplemental motor areas. The posterior parietal cortex also contributes to movement.

Through this section, we will walk through the brain processes leading to voluntary motor action, beginning at the highest areas of the hierarchy.

Association cortices

Neural control of voluntary movement begins with high order thought processes which are carried out by two major associative areas, the prefrontal cortex and the posterior parietal cortex. These areas do not specifically correlate with specific muscle groups, and activation of these neurons do not necessarily cause muscle activity. Instead, these structures are important for initiation of motor control. Primates have many cortical areas dedicated to movement, which allows for fine control of small muscle groups, complicated patterns of movements, and long-term planning of motor action.

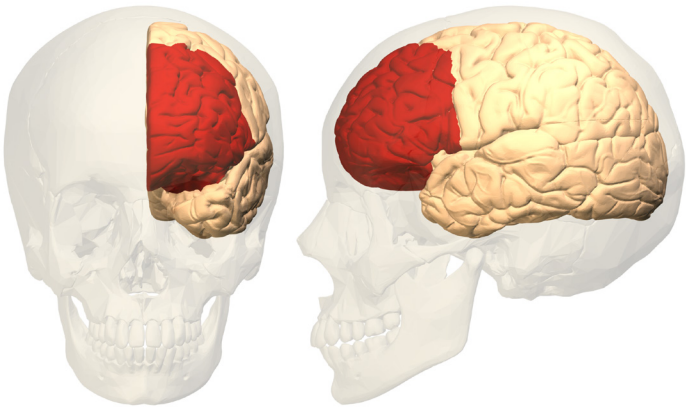


Figure 10.1 Anterior view (left) and lateral view (right) of one hemisphere showing the location of the prefrontal cortex (red).

With respect to motor control, the **prefrontal cortex (PFC)** initiates the long-term planning or cognitive aspects of movements. For example, consider the motor actions related to brushing your teeth. PFC signals are more akin to “brushing is good for my hygiene and health”, rather than “move my arm and open my mouth.”

PFC also helps determine if some motor action is appropriate for the specific situation. Think of a behavioral test where you are given two clickers, one to hold in each hand. You are told that when the experimenter shows you a green item, you should click with the right button. After repeating this behavior multiple times, you are told to switch - now, when you see a green item, you need to click the left button. In this experiment, PFC is responsible for deciding which motor pattern (left button or right button press) is appropriate in response to the stimulus.

PFC also works to weigh the consequences of motor actions, and makes updates about future motor actions in similar or different circumstances. The exact same motor action produces different results depending on the specific situation, and PFC contributes to evaluating and predicting outcomes. For example, the **Wisconsin Card Sorting Task** is a human behavioral test that is strongly dependent on PFC activity. In this task, a patient is asked to classify cards based on criteria, such as number, shape, or color. The patient is told whether they are correct or wrong; however, the criteria will occasionally change without them

being explicitly informed. The ability to shift to a new set of rules is PFC mediated.

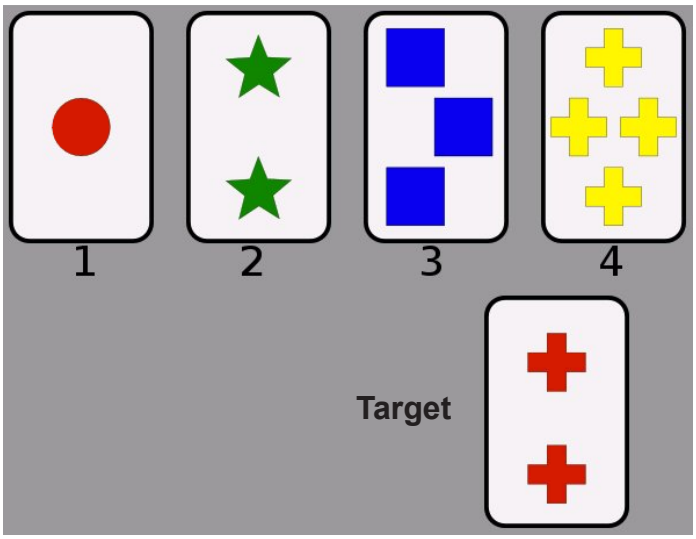


Figure 10.2 In the Wisconsin Card Sorting Task, the target card (bottom) could be placed in category 1 (matching color), category 2 (matching number of objects), or category 4 (matching shape). Ability to switch to a new set of rules uses prefrontal cortex.

The other major associative cortex contributing to motor function is the **posterior parietal cortex**, which is largely concerned integrating somatosensory with visual information and determining an appropriate motor action. For example, if you were planning to get up from your seat to walk across the room, the posterior parietal cortex would take in the somatosensory proprioceptive information about how your body is positioned, and the visual information from the objects in the room to avoid running into them (recall the dorsal stream pathway; chapter 7.4).

Motor cortex

The motor cortex is made up of three closely related brain structures that contribute to execution of movement: premotor area (PM), supplemental motor area (SMA), and primary motor cortex (or M1). These motor cortex areas

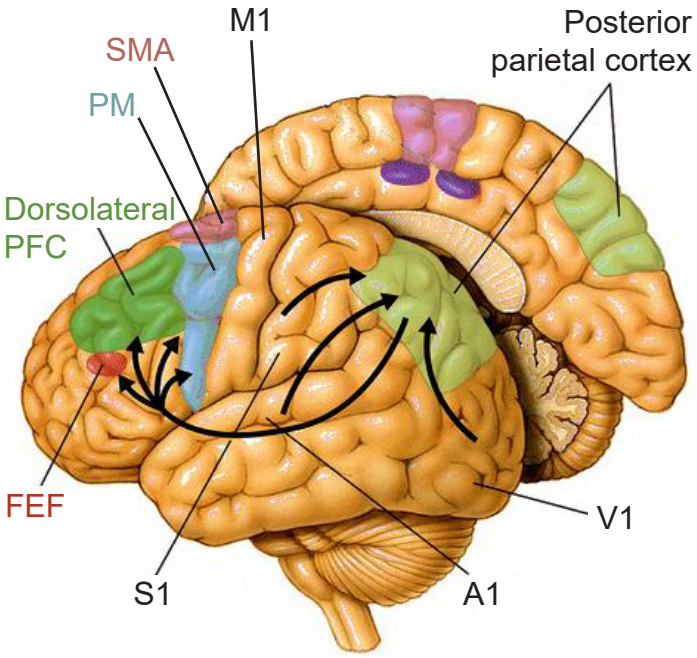


Figure 10.3 The posterior parietal cortex is part of the dorsal stream, and integrates sensory information with motor commands.

are found in the posterior aspect of the frontal lobe, directly adjacent to the central sulcus. The PM is the anterior most structure, and SMA is more dorsal to the PM along the medial aspect of the brain. M1 is posterior to that, bordering the central sulcus. Generally speaking, information from the associative cortices travel through PM, then M1, before projecting down through the brain stem and spinal cord.

These three structures are not an exhaustive list of motor cortex structures, however. For example, the **frontal eye field (FEF)** communicates with the extraocular muscles and mediates saccadic eye movements (chapter 7.4). The **inferior frontal gyrus**, or **Broca's area**, contributes to motor processes related to language (chapter 14).

The **premotor area (PM)** modulates motor output, and generally activates prior to motor activity. For example, if monkeys are trained to press on a button on a delay in response to a specific light cue, PM neurons will increase activity

upon presentation to the light cue, even before the arm or hand begins to move. Also, activity of PM increases as we imagine performing a complex series of finger patterns, such as during mental rehearsal of guitar or violin.

There is debate about the existence of a special population of cells in PM called **mirror neurons**. These are cells that are active during a movement, but also when that same movement is observed in another animal or person. Proponents argue that these neurons are involved in learning behaviors and for understanding the behaviors of others. However, given the complexity of the task of understanding others, it is unlikely to be encoded at the level of individual neurons.

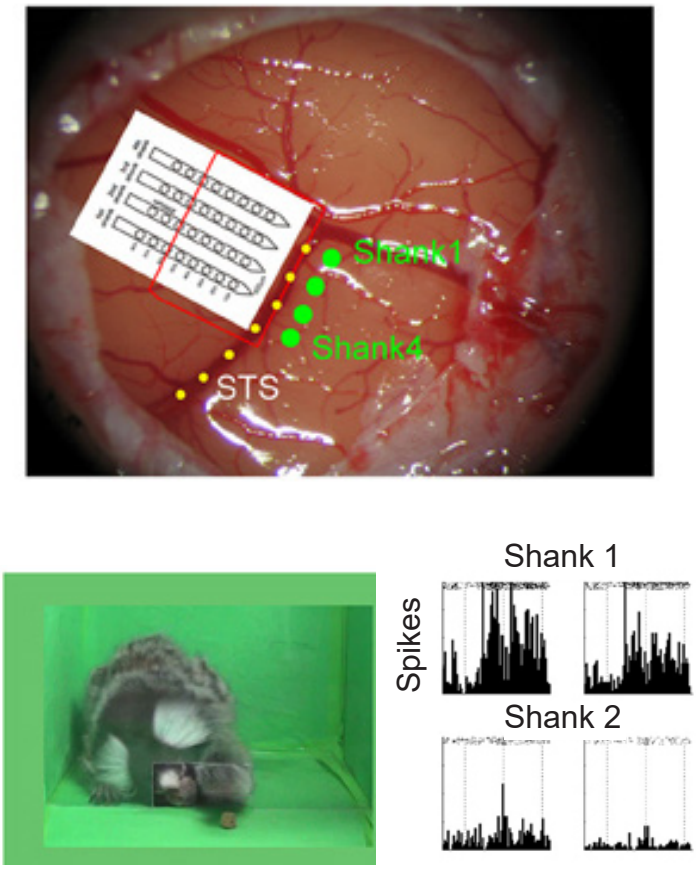


Figure 10.4 Surgically implanted electrodes (top) in a marmoset brain. While viewing a fellow marmoset reaching (bottom left), mirror neuron activity increases (bottom right).

The **supplemental motor area (SMA)** is upstream of primary motor cortex, but also sends downward projections through spinal cord. SMA communicates bilaterally, and lesioning this area causes deficits in manual coordination tasks that require both hemispheres to communicate. For example, in one monkey study, animals were presented with a table in which was a hole stuffed with a food reward. In order to obtain the food reward, the monkey needed to use one hand to push, and the other hand to catch. Lesions of SMA caused the monkey to try to push the food from both sides simultaneously. This suggests that SMA is responsible for the proper communication between motor commands in both hemispheres.

Primary motor cortex (M1) is a major motor control center, required for deliberate, voluntary movements, movements made in response to a "command." Furthermore, motor cortex cells influence **motor neurons**, neurons that communicate down into spinal cord that ultimately influence muscles or glands. This connection is so strong that the motor cortex cells are sometimes called **upper motor neurons**. In this terminology, a **lower motor neuron** is found at the brain stem or spinal cord, and fires whenever the upper motor neuron sends a signal.

In the 1930s, neurosurgeon Wilder Penfield conducted several brain operations to treat patients with severe epilepsy. However, since the brain has no pain receptors, he was able to remove a portion of the skull under local anesthesia while the patients were awake and responding to his questions. During surgery, the goal was to electrically stimulate portions of the cortex to determine the origin of the seizures, and to ensure that areas critical to speech and hand movement are left untouched so that the patient will not have major impairments following surgery. This is currently done in neuro-oncology

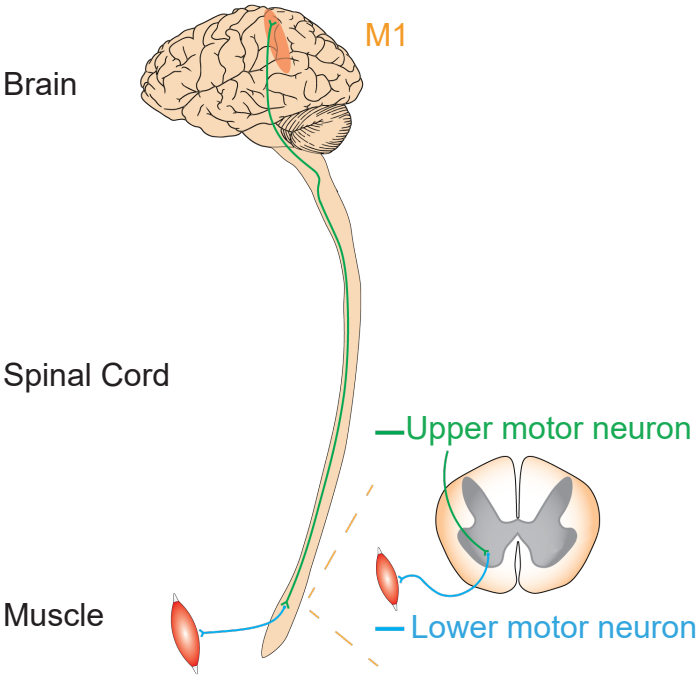


Figure 10.5 Relationship between upper motor neurons in M1 and the lower motor neurons further down of the signaling pathway.

to reduce the loss of critical motor function and overall morbidity.

Penfield progressively moved across different brain areas of M1 while using an electrode to stimulate patches of the cortex. He had two major observations. First, stimulation caused contralateral activity: that is, stimulating the left side of the brain affects the muscle activity of the right side of the body. Secondly, by systematically moving across M1, he observed that different populations of neurons are responsible for communicating with specific muscle groups. For instance, dorsal M1 activates hip and trunk muscles, while more lateral M1 activates muscles of the face.

Penfield discovered that within the motor cortex, different muscle groups are laid out in a rough topography, meaning that neurons that control the thumb are near other neurons that control the thumb, and near other populations of neurons that control the index finger. A graphical representation of this map is called the "**motor homunculus**," in which body parts with larger representations in the brain are shown with larger size (much like the sensory homunculus; chapter 8.3).

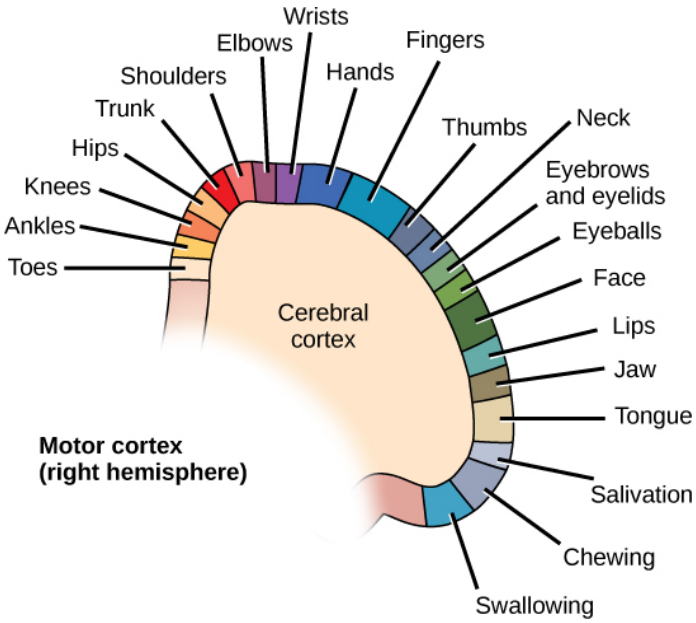


Figure 10.6 Topographical organization of the motor cortex (top), showing that neurons that control adjacent body parts are often adjacent themselves. The motor homunculus (bottom) is a representation of how much motor cortex is dedicated to control over a specific body part.

Clinical connection: Prosthetic limbs

If activation of specific neurons can produce muscle activity, then decoding these descending signals could be used to help amputees control their prosthetic limbs. This technology is called a **brain-machine interface**. In a technique called an **electrocorticogram (eCoG)**, the surgeon puts a high-density electrode array capable of sensing neural activity directly onto the surface of the cortex of M1. With this strategy, it is possible to detect electrical activity at high spatial resolution (hundreds of microns) and temporal resolution (hundreds of microseconds), which is critical for movements as precise and rapid as muscle movement. ECoG grids have already been successfully implanted in people with tetraplegia, allowing them to control external devices such as a cursor on a screen or a prosthetic limb.

In the early 2000s, researchers discovered that complex preprogrammed actions, such as opening the mouth and bringing a hand to the mouth, could be activated from a single parcel of brain regardless of where the hand was located in space - on your knee, on the table in front of you, or to the side, for example. In these experiments, the researchers delivered long stimulation trains, similar in duration to that observed in large muscle group movement - approximately a half a second or longer. Further experiments developed a gestural map, and even more complex behaviors such as reaching, climbing, and defensive postures.

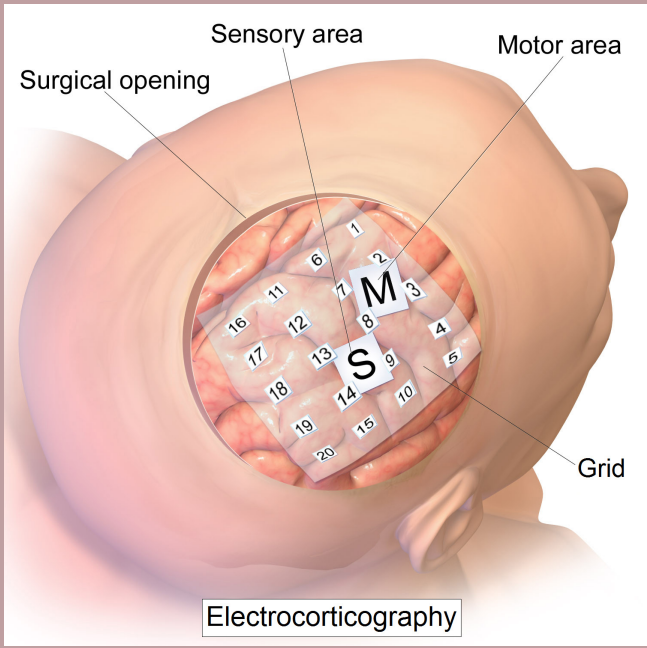


Figure 10.7 An eCoG allows a machine to read cortical electrical activity with high temporal and spatial precision, which can be used to control prosthetic limbs.

Sensory feedback

An important component of healthy movement is somatosensation, the process of sensing and perceiving the body (chapter 8.) One aspect is proprioception, the perception of our body in the world. For example, if you close your eyes, you can still feel if you are slouching and if your hands are in front of you or to your sides. Even if you were to move your arms forward, such as when reaching out to grab a nearby object, you could do it reasonably well without visual feedback. In order to make an accurate movement, the brain needs to know how well the actual movement matches the intended movement. If you are making an error, the sooner it is discovered, the sooner you can update your movement plan and adjust the movement.

Receptors in our muscles, tendons, joints, and skin detect contraction, stretch, and vibration. Sensory neurons carrying this

information synapse onto neurons in the spinal cord, sending sensory information to both motor neurons and also up to the brain, particularly the primary somatosensory cortex (or S1; chapter 8). Here, several bidirectional connections are made onto neurons in the adjacent motor cortex. Through these signaling communication routes, sensory information is used to update the motor commands and correct movements that go awry.

10.2 Modifiers of Descending Information

Although the descending signals from M1 are the major motor command signals that regulate activity, that signal is fine tuned through the action of two major brain structures, the cerebellum and a series of nuclei called the basal ganglia.

The Cerebellum

The **cerebellum** ("little brain" in Latin) is the most prominent structure of the hindbrain, located at the ventral-most part of the brain. The structure is evolutionarily ancient, and the general architecture and cell types of the cerebellum are conserved between teleost fish and mammals. The human cerebellum contains 69 billion neurons, which represents 80% of the total number of neurons in the human brain, despite being physically only 10% of the total brain mass.

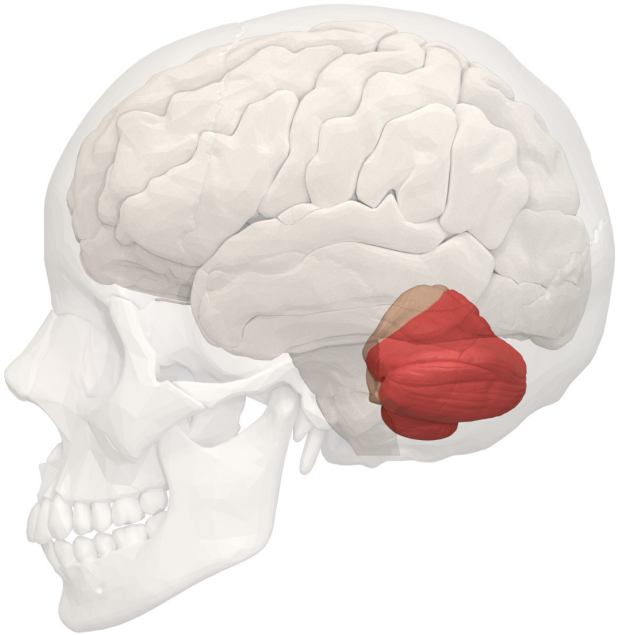


Figure 10.8 The cerebellum is relatively small, but contains more neurons than the rest of the brain.

Since the time of Galen in Ancient Greece (~200 CE), behavioral functions related to sensorimotor control have been ascribed to the cerebellum. In 1824, Marie-Jean-Pierre Flourens demonstrated that pigeons with cerebellar damage have poor wing flap coordination, resulting in a diminished ability to fly. Anatomist Jan Evangelista Purkyně (1787–1869) described the major output cell of the cerebellar cortex, and histological studies conducted by Santiago Ramon y Cajal (1909) later described cell morphology and the layers of the cerebellum.

The idea that cerebellum is key for coordinated movement remains central to cerebellar research. Cerebellum plays a role in integrating sensory information to produce coordinated movement, refining motor-related outputs to learn motor tasks, and processing cognitive and executive functions.

Nonmotor functions of cerebellar circuits have not received significant attention until recent years. Stimulation or lesion of deep cerebellar nuclei produced various autonomic and complex

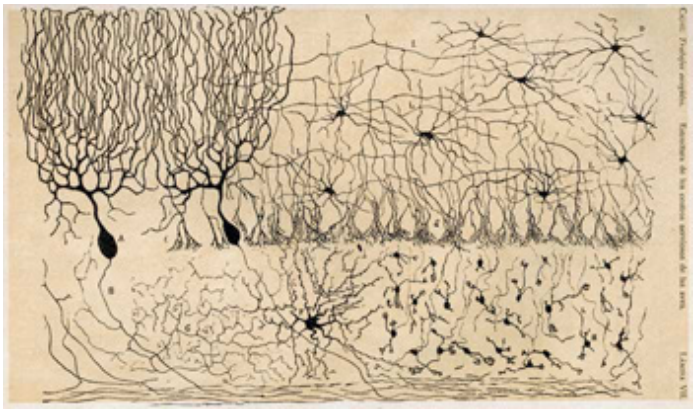


Figure 10.9 Early illustration by Ramon y Cajal of the different types of neurons found in the cerebellum.

behavioral outputs, and associative learning occurs in the cerebellum. Cerebellar abnormalities like Purkinje cell loss may be involved in autism. Furthermore, cerebellar cognitive affective syndrome (CCAS) is characterized by dysfunction in executive tasks (e.g. planning, working memory, abstract reasoning), impaired visual-spatial memory, changes to personality and emotional control, and problems with language production, demonstrating that cerebellum contributes to much more than just motor related behaviors.

Anatomy and function of the cerebellum

The cerebellum is composed of an external layer of gray matter (the cerebellar cortex), an internal core of white matter, and three pairs of deep nuclei (the fastigial nucleus, the interposed nuclei, and the dentate nucleus). The cerebellum is connected to the dorsal brain stem by three pairs of **peduncles**, or stalks of tissue: the inferior cerebellar peduncle, the middle cerebellar peduncle, and the superior cerebellar peduncle.

The cerebellum is organized in a series of regular, repeating units. It has a somatotopic organization with different regions of the cerebellum receiving afferent fibers from different sensory systems and projecting to different motor systems.

The surface of the cerebellum has a series of parallel folds called **lobules**. Two transverse fissures, the primary fissure and the posterolateral fissure, divide the cerebellum into three lobes. On the dorsal surface, the primary fissure separates the anterior and posterior lobes, while on the ventral surface, the posterolateral fissure separates the body of the cerebellum from the flocculonodular lobe. The midline region is called the vermis (Latin for "worm"), and on either side of the vermis is a cerebellar hemisphere.

Functionally, the cerebellum can be divided into three defined regions with roles in distinct types of movement.

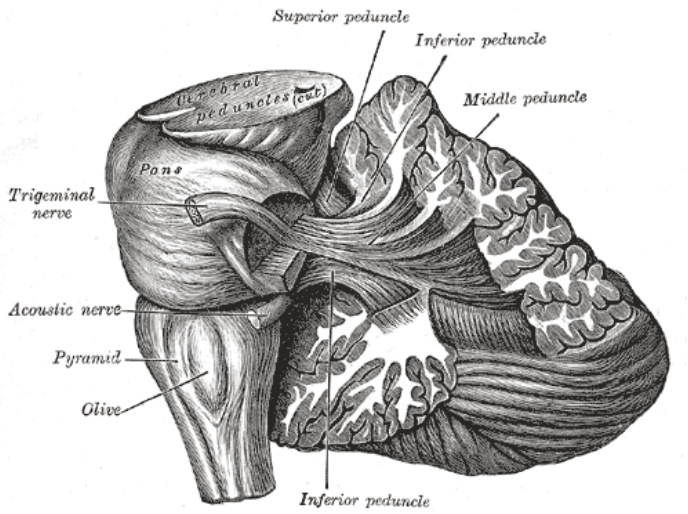


Figure 10.10 The cerebellum is attached to the brain stem via three large pairs of nerve fibers.

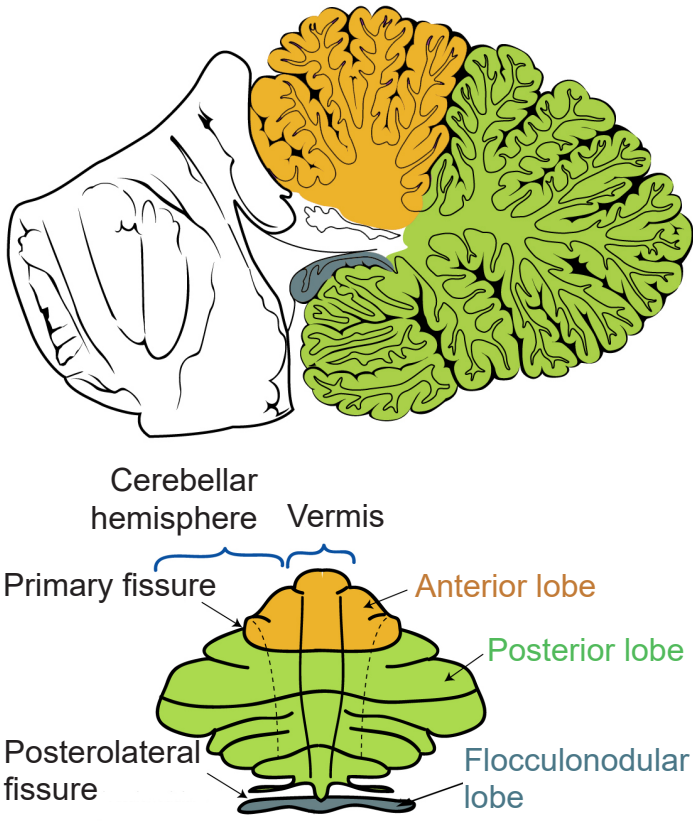


Figure 10.11 A midsagittal view of the cerebellum (top) and an "unrolled" view (bottom) labeling the major anatomical structures.

1. The **vestibulocerebellum**, composed of the flocculonodular lobe, is the most evolutionarily -conserved region of the cerebellum. It integrates visual and vestibular inputs, relaying information about head and body position in space. It can be thought of as two components, a medial aspect and a lateral aspect.

Medial vestibulocerebellum controls muscles of the trunk and head, which regulate posture. It also communicates with the limb extensor muscles, which help maintain balance at rest and during locomotion.

Lateral vestibulocerebellum controls eye movements and coordinates head and eye movements. In clinical observations, patients with lesions in the lateral vestibulocerebellum display deficits in smooth pursuit eye movement towards the side with the lesion.

2. The **spinocerebellum**, which includes the vermis and intermediate regions of the hemispheres, receives somatosensory and proprioceptive inputs from the spinal cord and is important for locomotion and extremity movement. The distal muscles of the limbs and digits are controlled by neurons in the intermediate regions of the cerebellar hemispheres that receive somatosensory inputs from the limb.

Somatosensory information (touch, pressure, proprioception; chapter 8) is passed to the spinocerebellum from the spinal cord. These connections provide feedback to the organism about its changing position and environment so that comparisons and adjustments can be made. Projections into the spinocerebellum display approximate somatotopy. There are two functionally distinct ascending pathways into spinocerebellum, both carrying information from spinal interneurons.

The **dorsal spinocerebellar tract** carries joint and muscle somatosensory information, providing the cerebellum with sensory feedback related to both voluntary and involuntary movement.

The **ventral spinocerebellar tract** is only active during voluntary movements. It sends a “copy” of spinal motor neuron activity to the cerebellum, informing the movement commands assembled at the spinal cord level.

Together, these spinocerebellar connections are important for motor learning. Sensory information from the actual movement (dorsal) is compared with the expected movement (ventral). If there is a mismatch, the motor command gets modification to achieve the desired output.

The cerebellum also initiates feed-forward control of muscle activity to regulate the timing of movements. This is anticipatory activity that causes muscle contractions to generate smooth and accurate motion.

3. The **cerebrocerebellum**, composed of the lateral portions of the cerebellar hemispheres, is the most evolutionarily recent region, and is much larger in humans and apes than other basal-order mammals. As the name implies, these structures

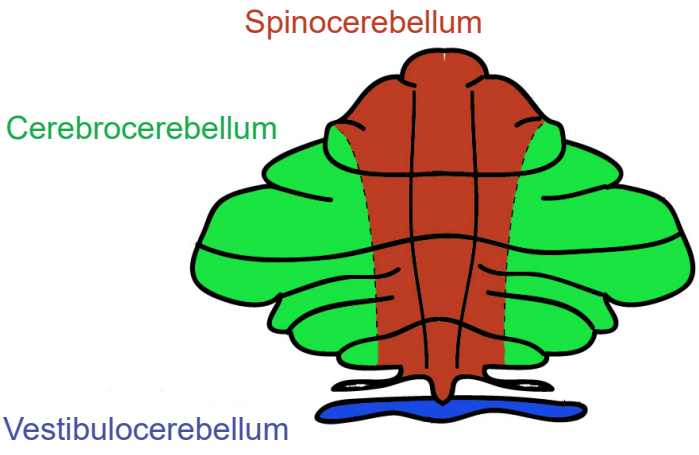


Figure 10.12 Functional divisions of the cerebellum.

Clinical connection: Ataxia

Ataxia is a condition characterized by poor coordination of voluntary movements, atypical eye movements, poor balance, or changes in gait. They may exhibit **dysarthria**, a speech disorder resulting in difficulty with articulating language. Acute alcohol intoxication produces an approximation of the symptoms of ataxia.

A wide variety of causes can result in ataxia, such as cerebellar injury from a stroke or tumor, toxin exposure, radiation poisoning, hereditary diseases, or diet-induced (gluten) autoimmune mediate cell death. Treatment of ataxia is most effective when the underlying cause is addressed, and may be completely reversible or untreatable. Medication and physical therapy may also help treat the symptoms.

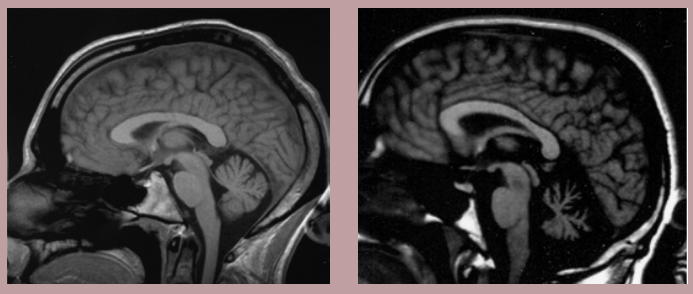


Figure 10.12 MRI showing a healthy cerebellum (left) and one experiencing degeneration from ataxia (right).

communicate with the cerebral cortex. Outputs of the cerebrocerebellum travel to motor and premotor cortices, and function in planning and executing movement. The cerebrocerebellum also targets nonmotor associative areas, such as prefrontal and posterior parietal cortex, contributing to cognitive functions. Neuroimaging and neuropsychological studies demonstrate cerebellar activation during tasks designed to evaluate attention, planning, working memory, abstract reasoning, language, pain, emotion and addiction.

Cellular organization of the cerebellum

The cerebellar cortex is arranged into three layers with distinct populations of neurons.

The deepest layer, called the **granular layer**, is the input layer of the cerebellar cortex. It contains a substantial number of small, densely packed excitatory neurons called **granule cells**, which are the primary unit of this layer. The granular layer also includes more sparsely distributed inhibitory interneurons (Golgi cells, Lugaro cells and chandelier cells) as well as excitatory interneurons (unipolar brush cells). **Mossy fibers**, one of the two major types of afferent projections into the cerebellum, terminate in the granular layer. Mossy fibers form synaptic complexes with granule cells and Golgi cells called **cerebellar glomeruli** that allow neurotransmitter spillover and crosstalk.

The middle layer is called the **Purkinje cell layer** and constitutes the output layer of the cerebellar cortex. It consists of a single layer of **Purkinje cells**, which have giant cell bodies and a broad dendritic arbor that stems from a central dendrite and fans in a single plane, extending upward into the outer layer of the cerebellar cortex. Purkinje cell axons are responsible for the output of the cerebellar cortex, projecting to deep cerebellar nuclei in the white matter or vestibular nuclei in the brain stem where GABA is released.

The external layer of the cerebellar cortex is the **molecular layer**, which contains the apical dendrites of the Purkinje cells and **parallel fibers**, the axons of granule cells. These parallel fibers

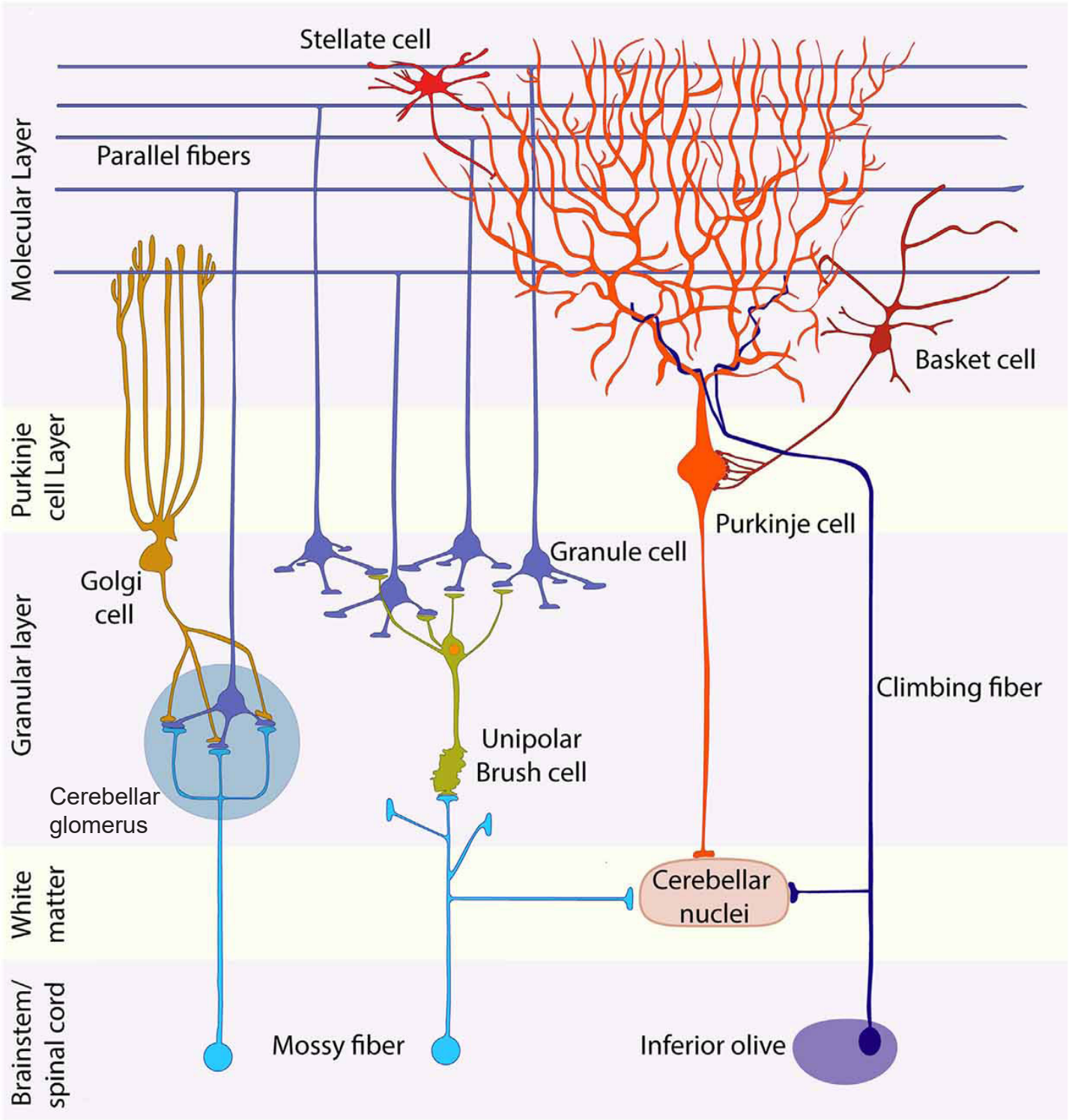


Figure 10.14 Cartoon illustrating the cellular organization of the neurons of the cerebellum.

ascend from the granular layer to the molecular layer where they bifurcate to form branches that extend mediolaterally, parallel to the long axis of the folia/lobules, thus giving them their name. Also in the molecular layer are two populations of inhibitory interneurons, the **stellate cells** and **basket cells**, which provide feed-forward inhibition to Purkinje cells.

The Basal Ganglia

The basal ganglia are a series of subcortical brain structures that are intimately involved with various aspects of movement, such as voluntary motor activity, habit learning, and the selection of actions. Despite being major structures involved in motor functions, none of the components of the basal ganglia directly send projections down the spinal cord. Rather, they communicate mostly within themselves before signaling through the motor cortex. While the connections between basal ganglia structures have been largely mapped out, the specific functions of each individual structure in isolation is not entirely clear.

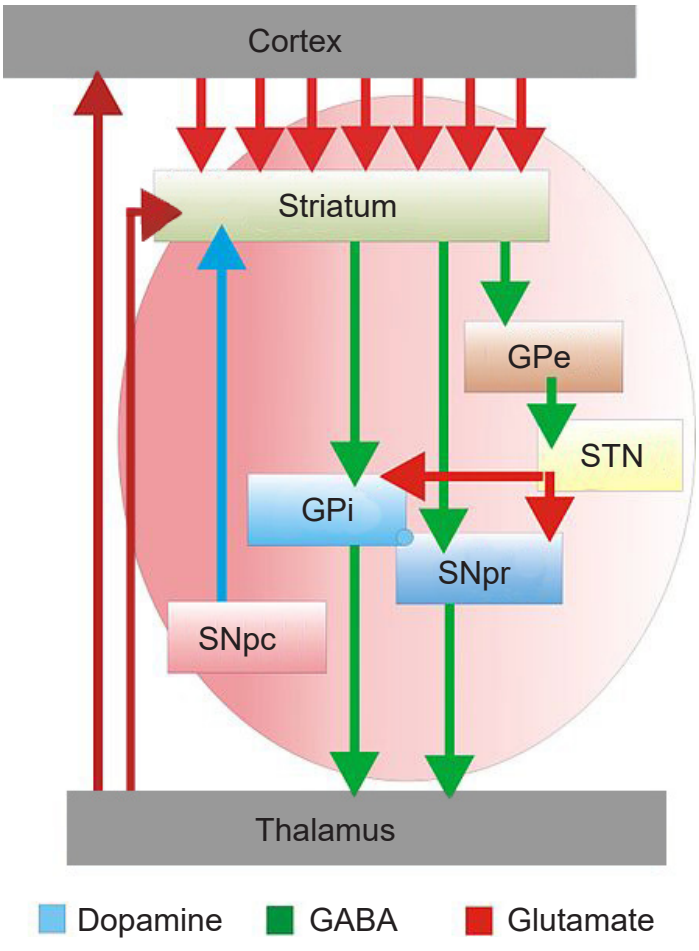


Figure 10.15 Diagram showing the connectivity and the nature of the signaling between different nuclei of the basal ganglia.

Diseases of the basal ganglia include movement disorders such as Parkinson's disease, dystonia, Huntington's disease, and Tourette's, as well as complex psychiatric disorders including addiction and obsessive-compulsive disorder (OCD).

Striatum

The striatum is the largest of the basal ganglia structures. It is the main input site of the basal ganglia, with projections coming from both cortical and limbic structures such as thalamus and amygdala.

On a cellular level, the majority of neurons in the striatum are GABAergic cells called **spiny projection neurons (SPNs)**, or **medium spiny neurons (MSNs)**. These neurons express different subtypes of dopamine receptors, and some of them are excited by dopamine while others are inhibited. Mixed within the SPNs are several populations of interneurons with various

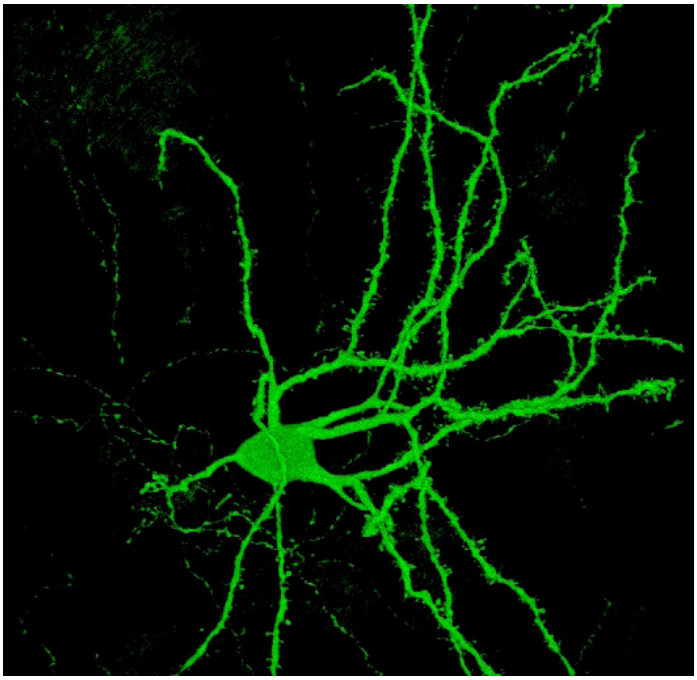


Figure 10.16 Confocal image of a striatal spiny projection neuron identified with green fluorescent protein (GFP).

electrophysiological properties that utilize different neurotransmitter systems, such as GABA, acetylcholine, and nitric oxide.

The striatum is divided into two anatomically and functionally different structures, a dorsal and ventral striatum.

Dorsal Striatum

The dorsal striatum is made up of two components, the **caudate nucleus** and **putamen**. In rodents, the two are indistinguishable.

Functionally, dorsal striatum contributes to learning of habitual behaviors. **Habits** are learned motor patterns that can be performed without full attention and generally inflexible due to repeated training. Habits can either be adaptive (such as changing physical stance in response to different environmental threats) or maladaptive (such as compulsive hair pulling or in substance use disorder).

An example of a behavioral test of habit learning is a **serial reaction time test**, where one of four visual stimuli are presented and the patient is expected to make a motor response, such as pushing on a corresponding button. If the pattern of stimuli repeat with some regularity (button 2 is often followed by button 4, for example) patients can unconsciously learn the repetition of the pattern, decreasing reaction time.

The caudate is particularly important for **goal-directed actions**, which are sets of motor behaviors that are made in response to knowledge of which actions lead to which outcomes. such as the complex motor actions made while performing an operant conditioning task. An experimental paradigm for this behavior would be a Skinner box, where an animal lever-presses or nose-pokes for a food or sugar reward (chapter 11.2).

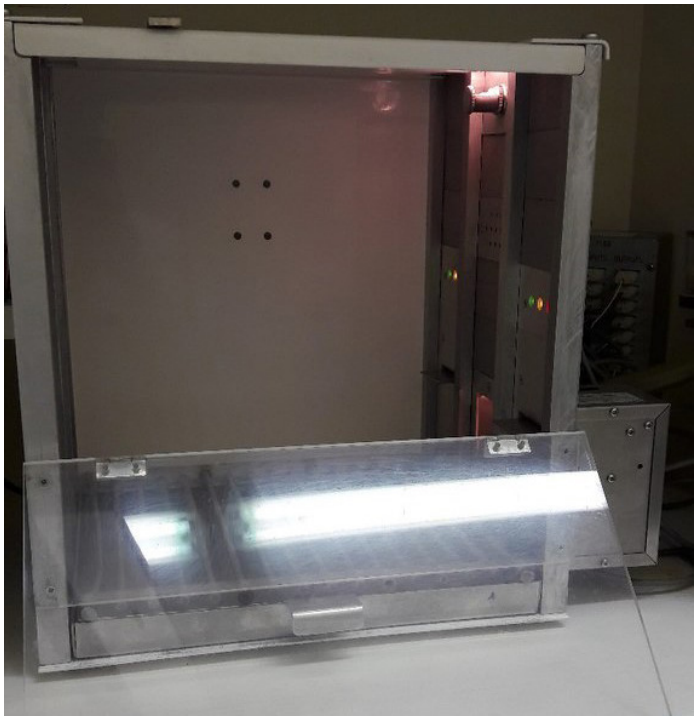


Figure 10.17 Skinner box can be used to measure operant conditioning, a dorsal striatum-mediated behavior.

Putamen contributes to motor-associated procedural learning tasks, like a **mirror tracing task** (chapter 13.1), which involves learning how to draw with a pencil while watching only a reflection of your hand. As with the serial reaction time test, improving at these behaviors happens unconsciously.

Anatomically, the dorsal striatum receives excitatory glutamatergic inputs from both motor cortex and thalamus. It also receives dopaminergic projections from another basal ganglia structure, the **substantia nigra pars compacta**, a communication route called the **nigrostriatal pathway**.

The only output of the dorsal striatum are the GABAergic spiny projection neurons (SPNs), which exist in two different types with two different targets. One population, the **direct SPNs (dSPNs),** express dopamine D1 receptors and send axons into the **internal globus**

pallidus (GPi). When this direct pathway is activated, it increases motor activity. The other output neurons, the **indirect SPNs (iSPNs)**, express dopamine D2 receptors and project into the **external globus pallidus (GPe)**. Activation of this indirect pathway decreases motor activity.

For healthy behavioral output, the dSPNs drive the intended motor action, while the iSPNs shut down competing motor actions. The balance of these systems is closely regulated, and imbalance of this signaling (such as in Parkinson's disease) leads to motor disturbances.

Ventral Striatum

The ventral striatum includes the **olfactory tubercule** and the **nucleus accumbens**, which is further subdivided into **core** and **shell**.

Importantly, these ventral striatum structures are important for reward, motivation, and aversion. As with dorsal striatum, dopamine

Clinical connection: Huntington's Disease

Huntington's disease is a rare neurodegenerative movement disorder resulting from various dysfunctional signaling pathways of the basal ganglia. Symptoms include **hyperkinesia** (uncontrolled movement), poor coordination, and cognitive and psychiatric changes eventually leading to dementia. The onset of symptoms happens when a person is in their 30s-50s, and prognosis is generally fatal within 15 years after diagnosis. There is currently no cure for the disease.

Huntington's disease passed through families in an autosomal dominant manner. A protein called **huntingtin (htt)** is implicated, and mutant huntingtin has several repeats of the amino acid glutamine. The greater the number of repeats, the more severe the symptoms and the earlier the onset of the disease. A leading theory suggests that mutant htt accumulates inside neurons, leading to neurodegeneration.

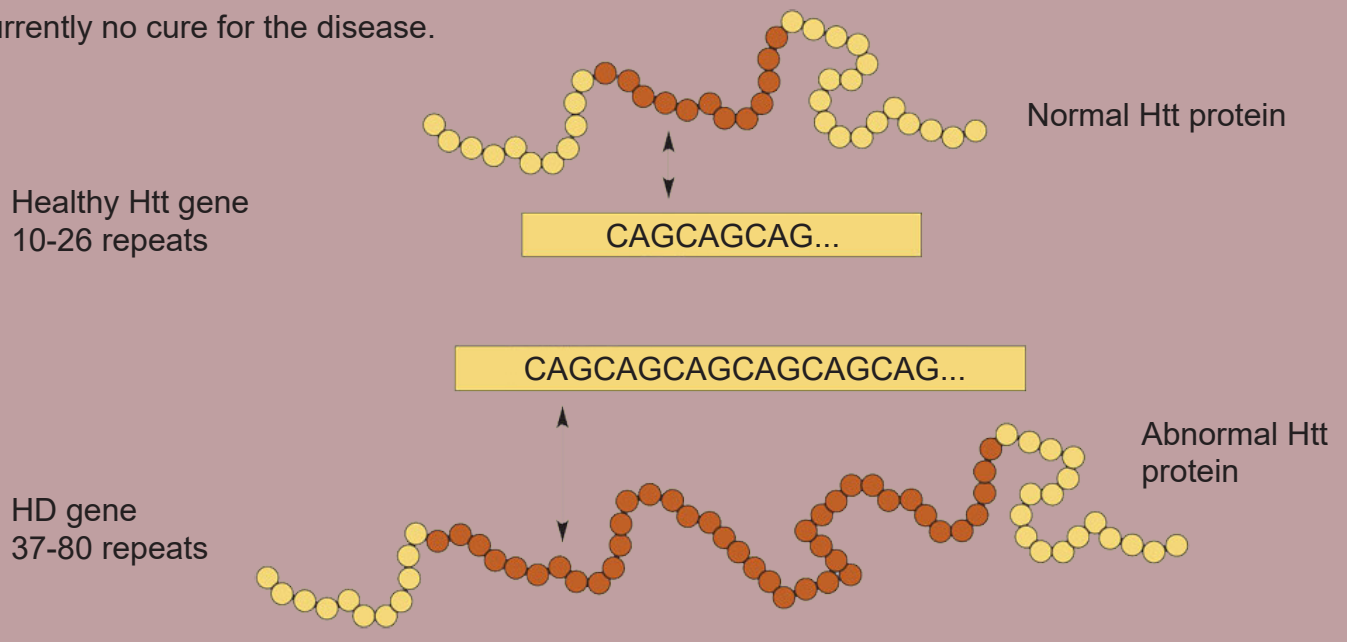


Figure 10.18 Longer CAG repeats in the Htt gene code for abnormal proteins. Greater number of repeats are associated with earlier age of onset of symptom presentation.

signaling is critical for mediating these behaviors. Increased dopamine signaling in ventral striatum is a "learning signal" that encourages the organism to repeat that behavior again. Dopamine is released in response to unexpected rewarding stimuli, such as getting a sip of a sugary drink or engaging in sexual contact. Many drugs of abuse, such as nicotine and cocaine, artificially drive up this dopamine signal, which leads to repeated use and drug seeking (chapter 11).

The major inputs into the ventral striatum are glutamatergic afferents from prefrontal cortex and limbic areas, such as hippocampus, amygdala, and thalamus. The dopaminergic inputs come from the **ventral tegmental area** in the midbrain.

Globus Pallidus (GP)

In dissection, the GP appears as a pale globe. It is subdivided into two components with functionally different connections. The **internal globus pallidus (GPi)** is movement promoting, and receives inputs from the dSPNs of the dorsal striatum. On the other hand, the **external globus pallidus (GPe)** is movement inhibiting, and receives dorsal striatal afferents from iSPNs.

Subthalamic Nucleus (STN)

The STN is part of the indirect pathway, downstream of the GPe. The GPe makes up the axonal inputs into the STN, which are inhibitory GABAergic projections. The output of the STN are glutamatergic signals into the GPi.

In a medical intervention called **deep brain stimulation (DBS),** a surgeon implants permanently indwelling electrodes directly into brain tissue. These electrodes are controlled by an external battery pack that delivers pre-programmed stimulation protocols. DBS in the STN is used to alleviate the symptoms of Parkinson's disease.

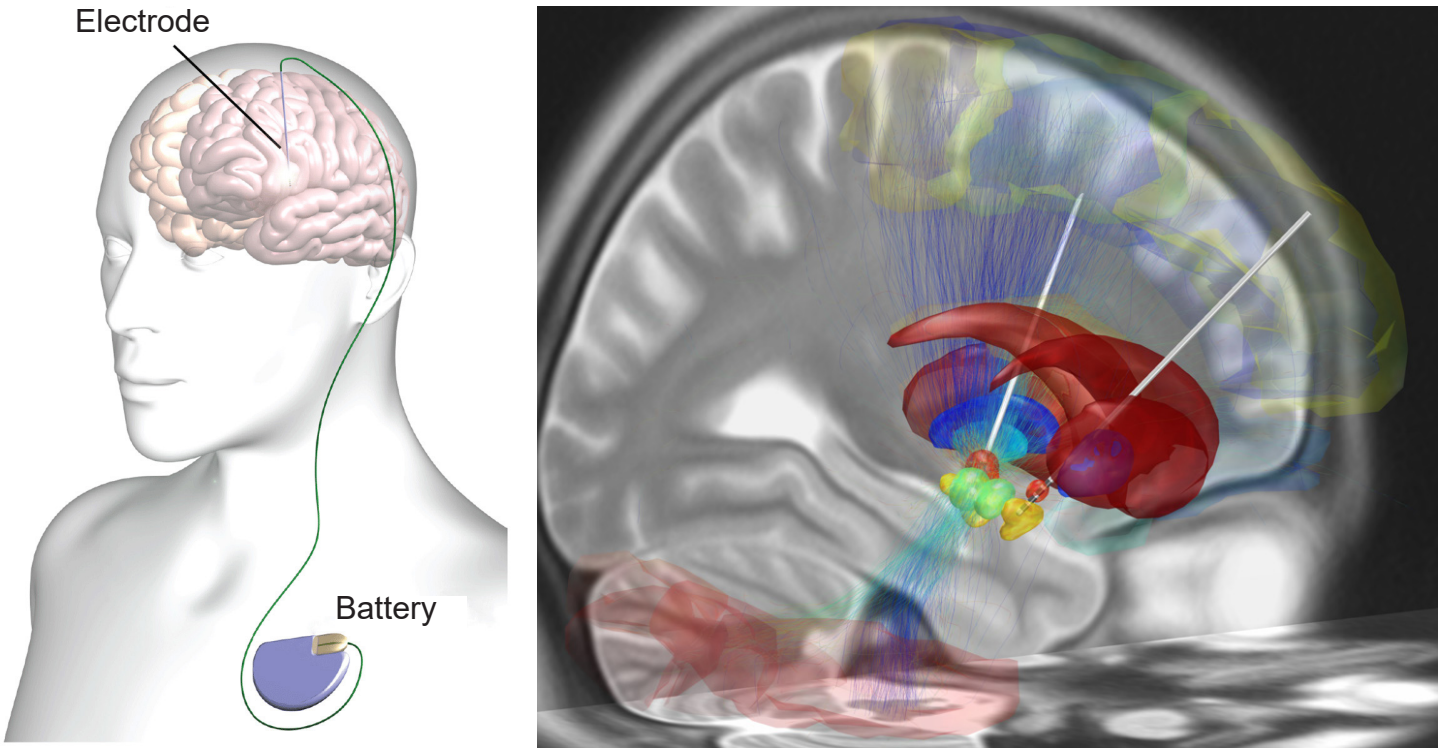


Figure 10.19 A deep brain stimulator (left) is controlled by an external battery. During surgery, the STN (right) can be targeted as a treatment for Parkinson's disease.

Substantia Nigra

The SN is a midbrain structure that appears darker in dissection due to heavy expression of neuromelanin across cells in these areas. It is the largest midbrain structure.

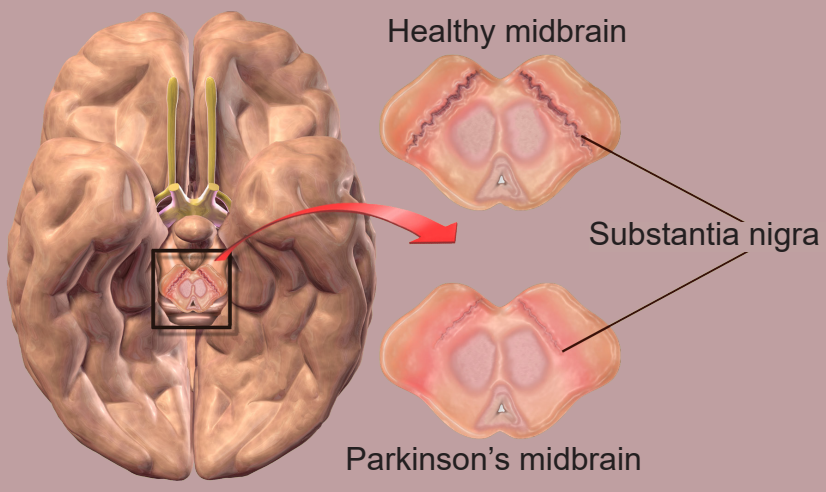
SN is divided into two areas. The **substantia nigra pars compacta (SNpc)** contains several dopamine expressing neurons that project into the dorsal striatum. Of clinical significance, these neurons experience selective neurodegeneration in Parkinson's disease.

The **substantia nigra pars reticulata (SNpr)** is an anatomically distinct area of the SN. It receives GABA-ergic inputs from the dorsal striatum and excitatory inputs from STN. The major output of SNpr is GABAergic; these axons terminate in the thalamus

Clinical connection: Parkinson's Disease

Parkinson's disease is a neurodegenerative movement disorder that causes **bradykinesia** (slowness of movement), a resting tremor, muscle rigidity, and changes to posture and locomotion. Although most symptoms are motor, there are mild cognitive and psychiatric changes, such as apathy, anhedonia, mood disturbances, or depression. Advanced age is the primary risk factor, as an estimated 1% of people over the age of 60 develop PD. Other environmental factors contribute to risk, such as repeated traumatic brain injury (suspected in Muhammad Ali) or occupational exposure to heavy metals, insecticides, or other neurotoxins. A small percentage of cases are early onset (21 - 50 years old; Michael J. Fox was diagnosed at 30), and have a strong genetic component. The disease causes significant decreases in life expectancy and quality of life.

A loss of dopaminergic neurons of the SNpc contributes to basal ganglia circuitry disruption. When less dopamine is released into dorsal striatum, the dSPNs are less active while the iSPNs are more active. This imbalance in the output leads to the motor symptoms of the disease. However, the cause of PD is still unknown; the formation of intracellular protein aggregates called **Lewy bodies** and oxidative stress leading to neurodegeneration are among some of the proposed mechanisms.



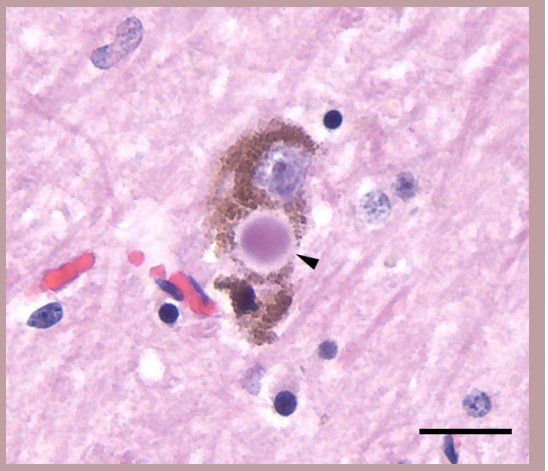


Figure 10.20 The substantia nigra selectively degrades in PD (left), possibly due to accumulation of Lewy bodies inside the neurons (right, arrow head).

10.3 The Spinal Cord

Spinal cord anatomy

The spinal cord extends posteriorly from the brain stem, surrounded by the vertebrae. In a typical human, the spinal cord is about 1 cm in diameter. Generally, the diameter of the spinal cord decreases from anterior to posterior, with two exceptions: an enlargement in the mid-cervical region to accommodate the extra nerves of the arms and hands, and a smaller enlargement in the lumbar region to accommodate extra nerves of the legs and feet (chapter 2.1).

The output neurons of the motor cortex carry voluntary motor information and send their axons down through two major descending communication routes. At the level of the brain stem, both pass through the **medullary pyramids**, paired anatomical structures found on the ventral surface of the medulla. Because of this, both descending axonal tracts are also called the **pyramidal tracts**.

The **corticobulbar tract** is made up of axons that terminate in various brain stem motor nuclei. These nuclei communicate directly with several cranial nerves, particularly those with motor function: The trigeminal nucleus (CN V) modulates muscles of chewing, the facial nucleus (CN VII) regulates facial expressions, the glossopharyngeal nucleus (CN IX) controls the muscles involved in speech and swallowing, and the hypoglossal nucleus (CN XII) controls the muscles of the tongue.

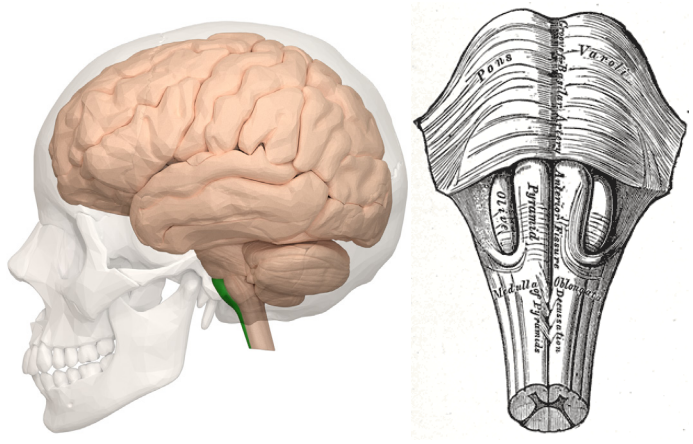


Figure 10.21 Position of the pyramids in the ventral spinal cord (left, green) and its location along the midline (right).

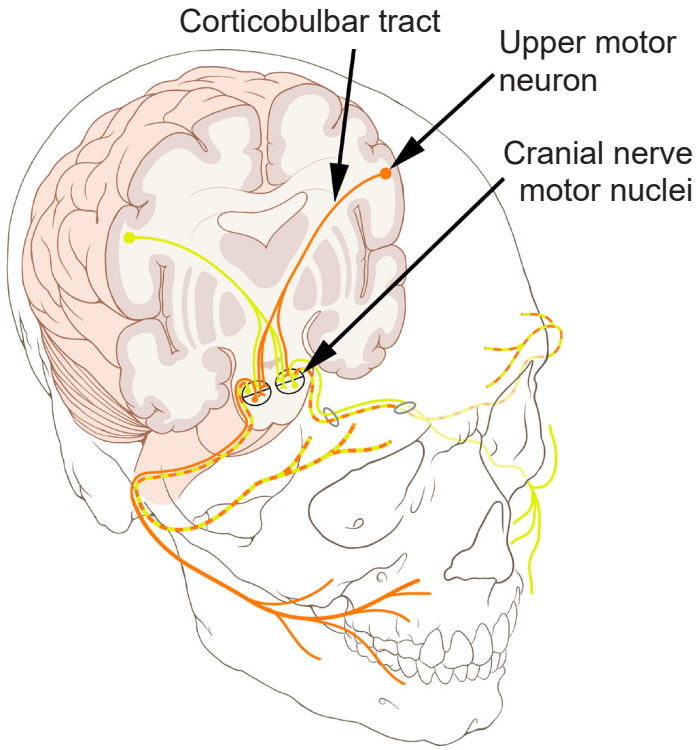


Figure 10.22 The corticobulbar tract originates in M1 and projects into the brain stem motor cranial nerve nuclei that control facial muscles.

The other major output of M1 are the upper motor neurons whose axons are the **corticospinal tract**, which runs down the spinal cord through white matter. This pathway has two different communication components.

Upwards of 90% of the corticospinal tract are axons of the **lateral corticospinal tract**, which sends voluntary motor control to the contralateral

distal limbs, such as in the arms, hands, legs, and feet. Some neurons of M1 are **Betz cells**, the largest known neurons in the brain. Most of this tract crosses hemispheres at the **pyramidal decussation**, runs down the contralateral spinal cord, then forms synapses onto the lower motor neurons in the ventral horn. These lower motor neurons then send nerve rootlets out of the ventral surface of the spinal cord, then form a neuromuscular junction with the muscles.

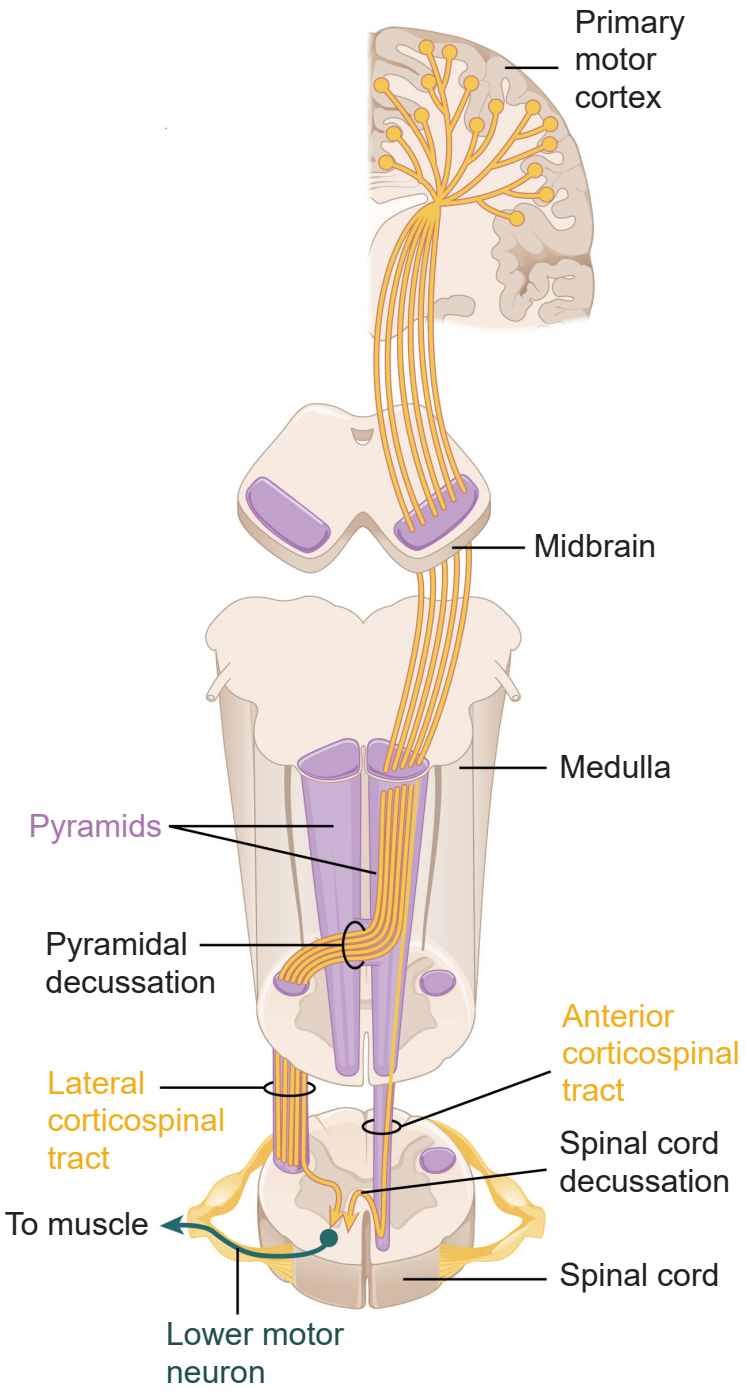


Figure 10.23 Connectivity of descending motor control pathways.

The remaining output is the **anterior corticospinal tract**, which carries information to the muscles of the trunk, such as our shoulders or pectorals. This signaling from the brain stem descends ipsilaterally, but eventually decussates at the spinal cord, resulting in contralateral control. Here, as above, the neuron forms synapses with the lower motor neurons found in the ventral horn.

Reflexes

Reflexes are involuntary motor responses that are performed automatically and independent of brain signals (although some can be suppressed voluntarily, with extra effort). Reflexes involve very simple circuits, sometimes consisting of only two populations of neurons: Sensory information comes in from the periphery, synapsing onto motor neurons in the brainstem or spinal cord. Reflexes can take place as quickly as 1/100th of a second!

A simple reflex is the **patellar reflex**, or the "**knee-jerk**" or **myotatic reflex** (described in more depth in chapter 8; figure 8.27). The stretching of the tendon causes the muscle to elongate, sensory receptors detect this change in muscle length, and the corresponding signal is sent into the spinal cord. This sensory neuron synapses directly onto a motoneuron, which then tells the thigh muscle to contract, making the kicking "knee-jerk" movement. Since only one synapse is involved in producing this response, this is an example of a **monosynaptic reflex arc**.

The reflexive kick is controlled at the level of the spinal cord and cannot be intentionally suppressed by descending motor pathways no matter how hard you concentrate.

Central pattern generators (CPG)

Central pattern generators (CPGs) are networks of cells that are capable of producing intrinsic motor responses even in the absence of sensory or brain inputs. These motor responses are usually well-rehearsed, repetitive, and happen at the unconscious level, meaning that certain behavioral outputs can be performed independently of signals upstream in the motor cortex. Some example motor responses driven by CPGs include diaphragm movement (respiration), alternating leg swinging and foot flexing (walking), and the progressive contraction of up to 25 pairs of muscles in the tongue and mouth (swallowing). Some CPGs are located in the brain stem (respiration) and others throughout the spinal cord (locomotion).

CPGs have been observed in several organisms, ranging from insects, crayfish, birds, and mammals including humans, hinting that their evolution was highly adaptive for survival. Imagine how difficult it would be to survive if every breath required a conscious thought!

The CPG is not a standalone driver of motor activity, however. The circuits receive signals from higher brain areas which can modify their characteristics. Imagine that you want to consciously hold your breath, or intentionally walk in some unusual or goofy way. In both cases, the descending signals from the brain are able to override the output motor command pattern that the CPG normally produces. Once you stop intentionally changing your motor behavior, the muscles return to their normal activity in response to CPG output.

CPGs of respiration

Getting oxygen is one of the most essential functions that an animal needs to perform for survival. As such, involuntary control of respiratory muscles through CPG activation is likely an evolutionary older mechanism. Respiratory control circuits are found at the level of the brain stem, specifically, the medulla and pons. One suspected CPG is a population of medullary cells called the **pre-Bötzinger complex**, which contains cells that change their activity in a cyclic pattern. Neurons in these areas are also sensitive to changes in blood chemistry (Chapter 9.3).

The major nerves involved with respiration are the **phrenic nerve**, which innervates the diaphragm, and the **intercostal nerves**, which predictably innervate the intercostal muscles. When we inhale (inspiration), the diaphragm contracts and the intercostals move the rib cage to increase the volume of the chest cavity, which allows the lungs to expand. During exhalation, the reverse process is true. These nerves receive innervation from neurons of the CPGs in the respiratory centers of the hindbrain (medulla and pons, specifically.)

The activity of the respiratory CPGs is so potent, that they can sometimes function while unconscious. Some deficits in respiratory CPGs are believed to be one cause of central sleep apnea, a disorder where a person stops breathing while asleep (chapter 12).

CPGs of locomotion

It is much more complex to describe the action of the muscle groups needed for a simple behavior of just walking down the sidewalk: the ankle dorsiflexors, the hamstrings, hip flexors, gluteus maximus and many more function at very precise moments, allowing us to propel ourselves forward. Importantly, this pattern of activity is rhythmic, and repeats at the frequency of stepping.

Many earlier studies of spinal cord CPGs were conducted in different animal groups with

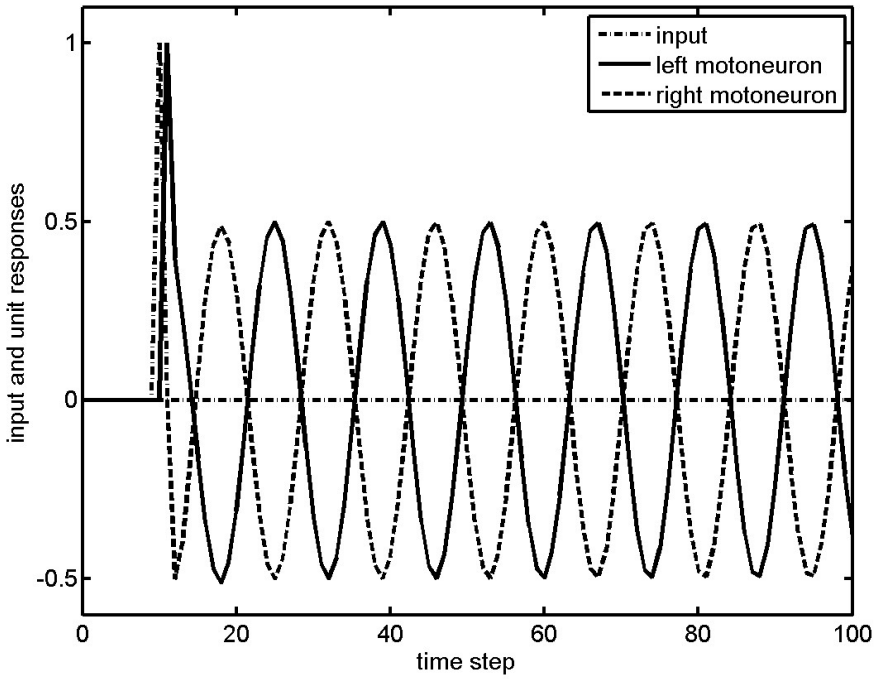


Figure 10.24 Locomotor CPGs produce alternating cyclic activity of motor neurons, which simplify walking.

unique locomotor patterns. Mollusks have crawling activity that is rhythmic, fish swim by sending alternative patterns of muscle activity through their body, and birds fly by rhythmic flapping of their wings.

In mammals, spinal CPGs were demonstrated in cats with a thoracic-level spinal cord transection. Following this surgery, the signals from the brain were unable to be communicated down to the hind legs, eliminating voluntary movement. They exhibited weakened muscular power, but were still able to stand for short times. If their front paws were put onto a stable platform while their hindlegs were put on a moving treadmill, the cat would involuntarily walk to keep up.

Humans also have CPGs for movement. One remarkable study looked at 37-year-old man who had injured his spinal cord in a football accident as a young man. His injury was at the level of C5, resulting in complete paralysis and a loss of sensation from the neck down. Over the following years, he gradually recovered a small amount of function. Fine manipulation skills and bladder bowel control never recovered, and he had not walked unassisted for ten years, experiencing tremendous weakness after taking a few steps.

Just days after beginning a physical therapy regimen consisting of assisted upright walking, upon lying down, the man reported alternating muscle flexion and relaxation of the hips, knees, and ankles similar to the pattern of activity seen in locomotion. He was unable to voluntarily stop these motions, and could only get some rest by turning to the side. This muscle activity was smooth and rhythmic, as would be seen in a healthy person walking.

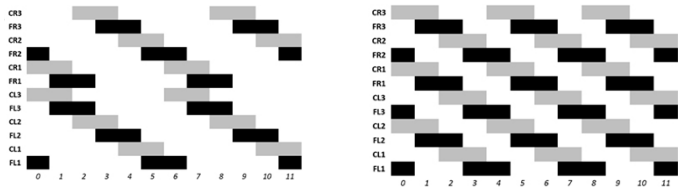


Figure 10.25 Six-legged animals like insects have central pattern generators that simplify locomotion through coordinated neural activity that regulates muscle movement.

10.4 The Muscles

Neuromuscular junction (NMJ)

The signal from lower motor neurons communicate with the muscle at the **neuromuscular junction (NMJ)**. The NMJ is similar to other chemical synapses, however the postsynaptic cell is a muscle cell separated by about 30 nm. The postsynaptic site is the **sarcolemma**, the cell membrane of muscle cells, which has several folds to increase surface area.

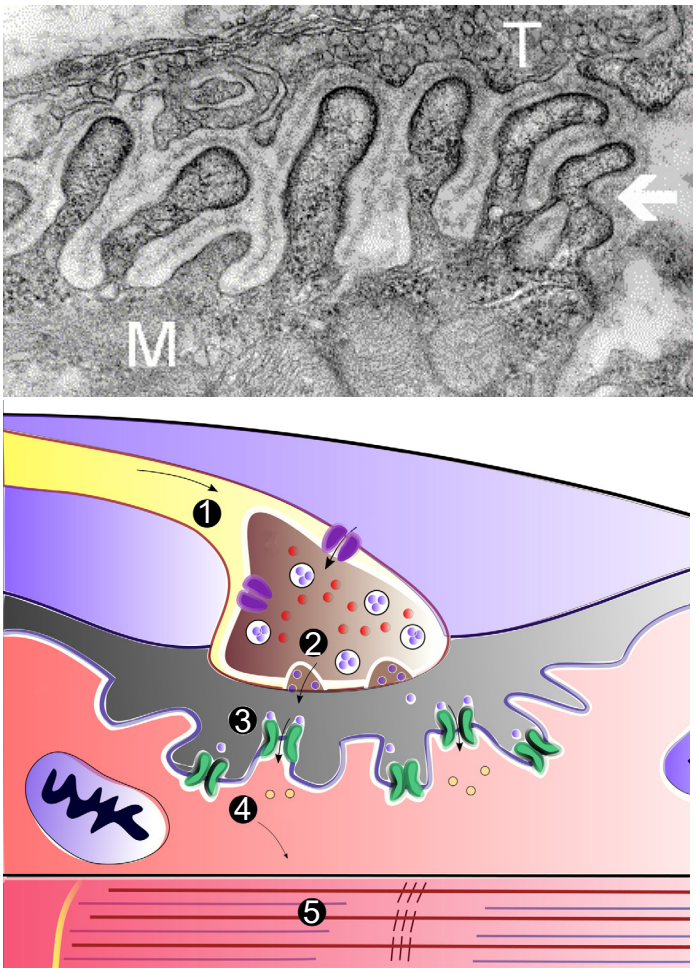


Figure 10.26 Electron micrograph image of an NMJ (top; T = axon terminal, M = muscle cell). Cartoon depicting the phases of signaling at an NMJ (bottom). The action potential (1) allows release of aceylcholine (2), activating nAChRs (3), allowing depolarization of muscle cells (4) and the contraction (5).

The axons of the lower motor neurons synthesize and release acetylcholine. Densely expressed in the sarcolemma are nicotinic acetylcholine receptors (nAChRs), ionotropic receptors that allow sodium influx and subsequent muscle cell depolarization upon acetylcholine binding. This depolarization results in muscle cell contraction.

Muscle anatomy

As mentioned above, motor functions can either be voluntary (moving your arm above your head) or involuntary (muscle contraction that leads to bowel motility, or heart beating). Although the nervous system influences both types of muscle activity, most of our discussion revolves around deliberate, voluntary skeletal muscle movement. The main action of muscles is to contract, a physical change of their shape where they become wider and shorter. For example, as you flex your arm, your bicep changes from long and thin to short and thick.

Different muscles have different characteristics driven by their shape and composition. For example, some muscles can be active for a long time without getting tired (maintaining your posture in your chair as you read this) while others can exert a lot of force but get tired quickly (lifting weights in the gym).

In studying muscles, some of the key measurements are where they connect to the bone, how long they are, how thick they are, and what type of muscle fibers they are made of. These measurements can be used to calculate how much force a muscle can generate and how quickly the joint can move.

Clinical connection: Myasthenia gravis (MG)

Myasthenia gravis (MG) is an autoimmune disorder characterized by muscle weakness, resulting in difficulty with speech, trouble with movement and swallowing, drooping eyelids, and double vision. Each year, an estimated 20 out of a million people get diagnosed with MG.

The muscle weakness seen in MG results from immune system-mediated destruction of the nAChRs expressed at the NMJ. Thus, when the lower motor neuron releases ACh, the muscle cells are unable to detect this release, so they fail to contract appropriately.

One therapeutic strategy involves inhibition of acetylcholinesterase, the enzyme that degrades ACh. This causes the synaptic ACh to remain in the synapse longer, increasing the chance that receptors get activated. Alternatively, autoimmune diseases like MG can be improved with immunosuppressant therapy. With successful treatment, MG usually does not result in changes in lifespan.

Figure 10.27 A patient with MG may show weakness of facial muscles.

Individual muscles, such as your biceps brachii, are made up of several muscle fascicles, which in turn are made up of muscle fibers. Muscle fibers are the individual cells in which contraction occurs. The functional units for contraction are called sarcomeres. It is the aggregated activity of hundreds of thousands of **sarcomeres** within each muscle fiber that generates the force with which you walk your dog or chew your food.

The contraction within sarcomeres happens between two proteins actin and myosin. **Actin** are the thin filaments that form the scaffolding, and **myosin** are thick filaments that pull the actin together, shortening the sarcomere and contracting the muscle. Although each sarcomere is small (1.5 to 3.5 μm), many are stacked back to back along the length of each fiber, pulling against each other to contract through the whole range of your joints. Many sarcomeres bundled in parallel provide the combined force that give muscles strength. The strength in all of your muscles comes from these tiny threads pulling against each other.

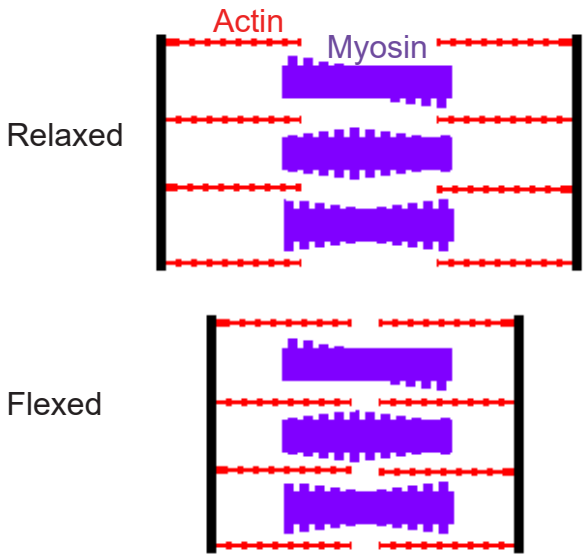


Figure 10.28 In a sarcomere, myosin (purple) and actin (red) overlap when the muscle flexes.

Muscle types

There are two key characteristics of muscles: their structure (striated due to the presence of sarcomeres or smooth), and whether they can be voluntarily controlled. Based on these characteristics, there are three types of muscles in humans and other vertebrates.

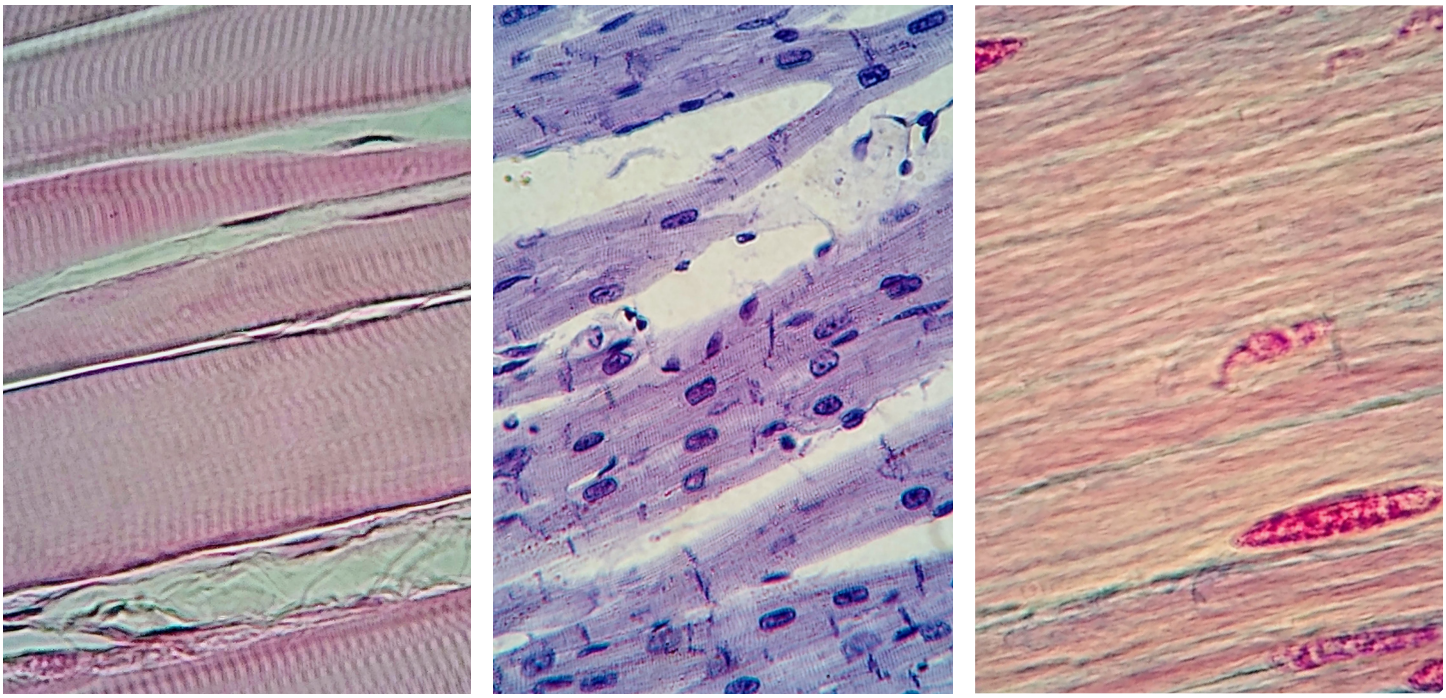
	Striated	Smooth
Voluntary	**Skeletal**	
Involuntary	**Cardiac**	**Smooth**

Skeletal muscle is what most people think of when they think "muscle." They attach to bones with tendons, and exert force on your skeleton to create movement and exert force on objects.

Skeletal muscles are also called "voluntary muscles" because they are the muscles that move when you choose to make a movement. They are voluntarily contracted (and in reflexes) to move your body by moving your skeleton. There are two types of skeletal muscles.

Fast twitch muscles generate a lot of force quickly, but also tire quickly. They are used mostly in high intensity exercise like lifting weights and sprinting. **Slow twitch muscles** generate less force, but can work for a long time. They are used in endurance exercise like jogging. Many exercises use a combination of both.

Muscles come in several different shapes, defined by the arrangement of their fibers and how they connect to tendons. The different shapes allow for different properties: some can change length quickly, others change shape less but are stronger. Another consideration for shape is the geometry of the joint they cover: for example, the pectoralis muscle reaches from your chest across your shoulder, and so the muscle is wide and flat



Skeletal Cardiac Smooth

Figure 10.29 Different types of muscle as seen at 400x magnification. Muscle fibers are seen running from left to right. In skeletal and cardiac muscles, striations can be seen perpendicular to the muscle fibers.

where it connects to your sternum, and narrows to a point where it connects to your upper arm.

Skeletal muscles work to move the body through a combination of cooperation and opposition. The **agonist muscle** is the main mover, like the **biceps brachii** when you flex at the elbow. It's movement is supported by **synergistic muscles**, including the brachialis and brachioradialis. These can also be **fixators**, providing stability and preventing or allowing rotation of the wrist while flexing at the elbow. **Antagonist muscles** are those that move in the opposite direction to the agonist. For the elbow, the antagonist is the triceps bracii, which lengthens during the bicep flex. Another example of an agonist-antagonist pair are the hamstrings to flex the leg, and the quadriceps to extend it.

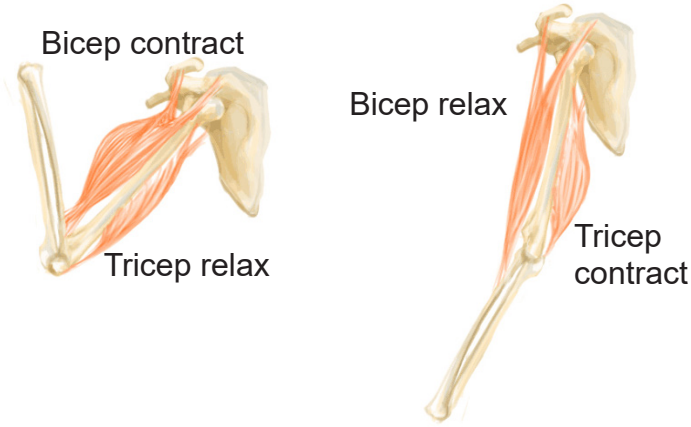


Figure 10.30 Skeletal muscles may be paired and produce opposing effects on a particular motion.

Smooth muscles are the muscles embedded within organs like your stomach and intestines, blood vessels, and bladder. They are also known as "involuntary" muscles because they are not under direct conscious control.

Cardiac muscle can be thought of as a hybrid between skeletal muscles and smooth muscles, in that they are striated like skeletal muscles but not under conscious control.

The Open Neuroscience Initiative is funded by a grant from the Vincentian Endowment Fund of DePaul University.

Cover: https://pixabay.com/photos/agility-ballet-dancing-athlete-1850711/
10.1 https://commons.wikimedia.org/wiki/File:Prefrontal_cortex_(left)_-_anterior_view.png https://commons.wikimedia.org/wiki/File:Prefrontal_cortex_(left)_-_lateral_view.png
10.2 https://commons.wikimedia.org/wiki/File:WisconsinCardSort.png modified by Austin Lim
10.3 https://commons.wikimedia.org/wiki/File:Posterior_Parietal_Lobe.jpg
10.4 Suzuki W, Banno T, Miyakawa N, Abe H, Goda N and Ichinohe N (2015) Mirror Neurons in a New World Monkey, Common Marmoset. Front. Neurosci. 9:459. doi: 10.3389/fnins.2015.00459
10.5 https://commons.wikimedia.org/wiki/File:UMN_vs_LMN.png
10.6 https://commons.wikimedia.org/wiki/File:Sensory_and_motor_homunculi.jpg
https://commons.wikimedia.org/wiki/File:Figure_35_03_04.jpg
10.7 https://commons.wikimedia.org/wiki/File:Intracranial_electrode_grid_for_electrocorticography.png
10.8 https://commons.wikimedia.org/wiki/File:Posterior_lobe_of_cerebellum_--_04.png
10.9 https://commons.wikimedia.org/wiki/File:CajalCerebellum.jpg
10.10 https://commons.wikimedia.org/wiki/File:Gray705.png
10.11 https://commons.wikimedia.org/wiki/File:CerebellumDiv.png modified by Austin Lim https://commons.wikimedia.org/wiki/File:1613_Major_Regions_of_the_Cerebellum-02.jpg modified by Austin Lim
10.12 https://commons.wikimedia.org/wiki/File:CerebellumDiv.png modified by Austin Lim
10.13 https://commons.wikimedia.org/wiki/File:Brain_chrischan.jpg https://commons.wikimedia.org/wiki/File:Ataxia.jpg
10.14 Consalez GG, Goldowitz D, Casoni F and Hawkes R (2021) Origins, Development, and Compartmentation of the Granule Cells of the Cerebellum. Front. Neural Circuits 14:611841. doi: 10.3389/fncir.2020.611841
10.15 https://commons.wikimedia.org/wiki/File:Basal-ganglia-2.jpg modified by Austin Lim
10.16 https://commons.wikimedia.org/wiki/File:Confocal_image_of_spiny_neuron_-_1.jpg
10.17 https://commons.wikimedia.org/wiki/File:Skinner%27s_box.jpg
10.18 https://commons.wikimedia.org/wiki/File:Huntington%27s_disease_(5880985560).jpg modified by Austin Lim
10.19 https://commons.wikimedia.org/wiki/File:Typical_deep_brain_stimulation_setup.jpg modified by Austin Lim
https://commons.wikimedia.org/wiki/File:Deep_brain_stimulation_electrode_placement_reconstruction.png
10.20 https://en.wikipedia.org/wiki/Pathophysiology_of_Parkinson%27s_disease#/media/File:Blausen_0704_ParkinsonsDisease.png Blausen.com staff (2014). “Medical gallery of Blausen Medical 2014”. WikiJournal of Medicine 1 (2). DOI:10.15347/wjm/2014.010. ISSN 2002-4436. - Own work
https://commons.wikimedia.org/wiki/File:Lewy_body_in_the_substantia_nigra_from_a_person_with_Parkinson%27s_disease.jpg
10.21 https://commons.wikimedia.org/wiki/File:Pyramid_(medulla_oblongata)_-_lateral_view.png https://commons.wikimedia.org/wiki/File:Gray679.png
10.22 https://commons.wikimedia.org/wiki/File:Motor_Facial_Route.png Patrick J. Lynch; illustrator; C. Carl Jaffe; MD; cardiologist Yale University Center for Advanced Instructional Media Medical Illustrations by Patrick Lynch, generated for multimedia teaching projects by the Yale University School of Medicine, Center for Advanced Instructional Media, 1987-2000. Patrick J. Lynch, http://patricklynch.net. Modified by Austin Lim
10.23 https://commons.wikimedia.org/wiki/File:1426_Corticospinal_Pathway.jpg modified by Austin Lim
10.24 https://commons.wikimedia.org/wiki/File:Cpg_output.jpg
10.25 Espinal A, Rostro-Gonzalez H, Carpio M, Guerra-Hernandez EI, Ornelas-Rodriguez M and Sotelo-Figueroa M (2016) Design of Spiking Central Pattern Generators for Multiple Locomotion Gaits in Hexapod Robots by Christiansen Grammar Evolution. Front. Neurorobot. 10:6. doi: 10.3389/fnbot.2016.00006
10.26 https://commons.wikimedia.org/wiki/File:Electron_micrograph_of_neuromuscular_junction_(cross-section).jpg https://commons.wikimedia.org/wiki/File:The_Muscle_Contraction_Process.png Modified by Austin Lim
10.27 https://commons.wikimedia.org/wiki/File:Myasthenia_Gravis.jpg
10.28 https://commons.wikimedia.org/wiki/File:Sarcomere_relaxed_contracted.PNG
10.29 https://commons.wikimedia.org/wiki/File:Skeletal_muscle_400x.jpg https://commons.wikimedia.org/wiki/File:Cardiac_muscle_histology_400x.jpg
https://commons.wikimedia.org/wiki/File:Smooth_muscle_400x.jpg
10.30 https://commons.wikimedia.org/wiki/File:Animation_triceps_biceps.gif Modified by Austin Lim

Chapter 11:

Neuropharmacology and Substance Use

Austin Lim, PhD (DePaul University)
Editor: Harriet de Wit, PhD (University of Chicago)

More than 275 million people worldwide have misused drugs such as heroin or cocaine at least once in 2016. Use of some of these drugs carry significant health risks, and their use has led to hundreds of thousands of deaths, making drug use one of the highest preventable causes of mortality.

For some people, drug use turns into a **substance use disorder**, such as addiction. These disorders are prevalent worldwide, affecting some 30 million people globally. On top of the direct health risks associated with compulsive drug use, drug addiction has a tremendous financial burden: Alcohol, tobacco, and illicit drugs cost the US more than $740 billion dollars in lost productivity, health care costs, and legal related fees and fines.

This chapter introduces you to some of the basics of **neuropharmacology**, the study of drugs that affect the nervous system. Drugs are defined as some chemical substance that comes from outside the body; drugs are exogenous substances that can influence a person's physical or mental state. The field of study of the effect of drugs on the body is called **pharmacodynamics**, while the reverse study, the effects of the body on the drugs, is called **pharmacokinetics**.

Chapter 11 outline

11.1 Common Routes of Administration

For a drug to influence the nervous system, it has to find a way to get into the body. Drugs can be used by several **routes of administration**, which can be roughly divided into two main classifications: **enteral**, which means they have to be absorbed through the gastrointestinal tract, or **parenteral**, which avoids the digestive tract. Routes of administration are particularly important in substance use, because drugs that reach the brain more quickly are more likely to produce pleasurable feelings, and thus be misused.

The most common route of administration is by **oral** administration, by mouth. Drugs consumed orally can come in a variety of preparations, with a pill being the most common, but they can also be foods or drinks. Swallowing substances is one of the most natural processes, making oral administration one of the most convenient ways

Figure 11.1 Pills, one of the most common preparations for oral administration of drugs.

to take a drug, generally very simple requiring no additional supplies. Orally administered drugs can be designed to release over a prolonged period of time, even up to a day, which makes it a preferred route of administration for substances that need to be present for extended periods of time. The substance can also be expelled from the body through vomiting or stomach pumping in case it is suspected as causing toxicity or in overdose. Absorption of the drug into the bloodstream can also be modified by other substances, such as foods in the stomach.

One of the major considerations of oral administration is the slow onset of effect of the drug. It requires digestive processes which work on the time scale of tens of minutes to hours, making it less effective in the case of major emergencies where speed of treatment effect is the difference between life and death. It also requires the conscious action of swallowing, so oral administration will not be useful when a person is unconscious. And if a person has nausea, they may be unable to keep the drug in their system long enough for it to be digested and absorbed.

Substances that are ingested orally are

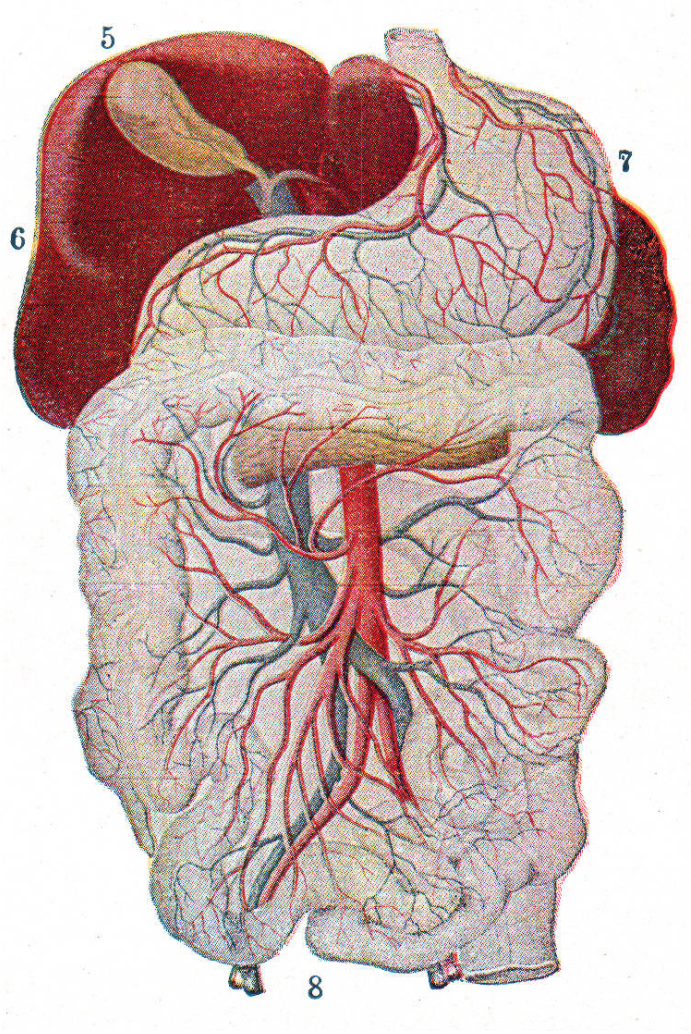


Figure 11.2 The hepatic portal system moves substances taken orally through the liver for enzymatic processing before entering total circulation.

absorbed through the gut wall and into the **hepatic portal system**, a branch of the circulatory system that surrounds the gastrointestinal tract. Before entering systemic circulation, this blood is passed through the liver, where a variety of degradation enzymes destroy some amount of the drug. This process is called **first-pass metabolism**, and causes a decrease in difference between the consumed amount of drug and the amount of drug that gets to its target of action.

On the contrary, some drugs are inactive until they get converted into another substance

Figure 11.3 Hypodermic needle for drug delivery via intravenous injection, a very rapid route of administration that can cause a drug effect within seconds.

after enzymatic processing. These substances are called **prodrugs**. One well known prodrug is **psilocybin**, a chemical found in "magic mushrooms." Psilocybin itself is inactive, but once it is exposed to highly acidic conditions, like the inside of the stomach, it becomes the psychoactive substance psilocin.

One other enteric route of administration avoids some of the issues with oral preparations. **Rectal** administration avoids a majority of the first-pass metabolism, and can be delivered even if a patient has severe nausea. There are two other related enteric methods worth mentioning. **Sublingual** refers to placement of the substance under the tongue, and may be seen with some preparations of LSD. **Transbuccal** administration refers to drugs that are absorbed by the vasculature in the gums as in oral nicotine preparations such as dip or Swedish snus.

Intravenous (IV) injection is the fastest route of administration, with drug effects being detected within seconds. When a needle is used to deliver the drug directly into the bloodstream, the drug is not subject to any first-pass metabolism, meaning that the dosage delivered is similar to the dosage experienced. Many substances, in particular heroin, are delivered via IV injection since it produces the most rapid onset high. In general, drugs that are taken via rapid routes of administration have a higher misuse potential. Clinically, the speed of an IV injection drug effect makes it preferable when time is essential, notably, when injecting substances like Narcan (naloxone) that can reverse the respiratory depressive symptoms of opioid overdose. Alternatively, it is also possible to deliver a pain-killer drug slowly and as needed using an IV drip, which minimizes the risks of overdose due to a sudden influx of drug.

IV injections are sort of a one-way street: the concentration of drug that enters the blood cannot be decreased. The drug rapidly mixes with blood, and can only be decreased with biological modifying agents or through natural enzymatic degradation. IV injections require syringes, which are an added cost to drug use, and bring with it a variety of health risks, including the potential transmission of bloodborne illnesses and infections at the injection site. Repeated injections can be painful and can lead to deflated veins, which is why some IV drug addicts resort to shooting drugs into other smaller veins such as the veins on their hands or between the toes.

Alternative injection routes are less common for misused drugs, and each has slightly different advantages and disadvantages. **Intramuscular (IM) injections**, like a flu shot

into the deltoid muscle of the arm, are somewhat slower than IV injections (tens of minutes rather than tens of seconds) but are much easier to perform than an IV injection. In a **subcutaneous (SC) injection**, solution is injected into the fat layer underneath the skin, making it ideal for lipid-soluble drugs.

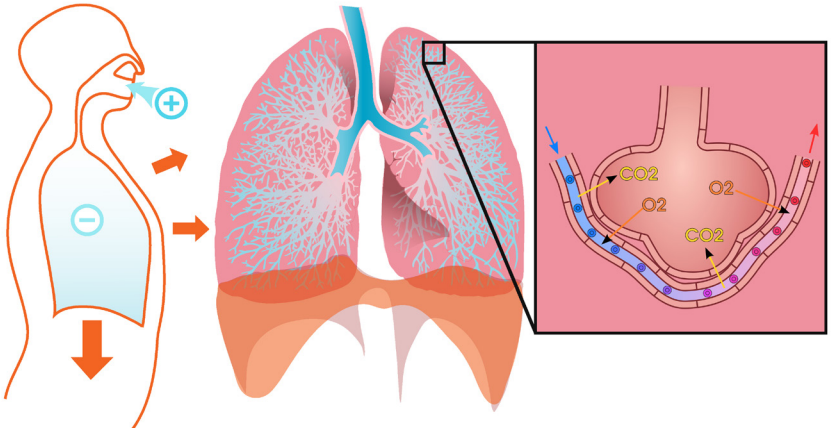


Figure 11.4 Inhalation allows for exchange of gases at the alveoli, so drug enters the bloodstream rapidly.

Inhalation is another rapid route of administration. Drug effects are experienced on the scale of tens of seconds. For a substance to have an effect on the brain through inhalation, it must first be volatilized then taken into the lungs through the nose or mouth. To inhale solids (such as with tobacco or marijuana), they are often burned and the smoke is inhaled, while liquids are turned into a vapor either through a heating element like vape or naturally because of the chemical properties of the liquid.

In some cases, drugs interact with foods in a way that changes their effectiveness. A common example of this is the compound *bergamottin*, commonly found in grapefruits. Bergamottin inhibits the activity enzymes called **cytochrome P450s (CYP),** a class of enzymes in the liver and the small intestines that metabolize a wide variety of substances that enter the body. When we drink grapefruit juice, inhibition of CYP can decrease the dosage of drug that gets into the brain. For example, the analgesic drug codeine is a prodrug, a substance that exerts influence on the body after it is modified during metabolism. CYP normally degrades codeine into the more potent analgesic morphine, but in the presence of grapefruit juice in the body less morphine is produced, resulting in less effect than expected.

On the other hand, other drugs are degraded by the CYP-3A4 enzyme before they are even able to affect the brain, like the anti anxiety drug buspirone. Usually, only 4% of the dosage of buspirone actually has an effect. But, when a CYP-3A4 inhibitor like grapefruit juice is present, a regular dosage of the drug results in a greatly elevated concentration in the brain, which may start to produce symptoms of overdose. More than 85 pharmaceuticals interact with grapefruit juice; some of the interactions can lead to life-threatening cardiovascular / respiratory side effects (like a buildup of the opioid fentanyl) or permanent kidney damage.

Physiologically, the lungs are extremely effective at transporting inhaled gases into the bloodstream, the same process that allows us to capture oxygen from the air and expel carbon dioxide as we exhale. Gas exchange occurs at the pulmonary capillaries, the dense network of blood vessels that surrounds the alveoli of the lungs. Inhaled air fills these alveoli and exchanges gases in the time span of each breath. This remarkable efficiency of exchanging inspired gases with the dissolved gases in the bloodstream provides one of the major clinical advantages of inhalation as a route of administration, especially during surgery. Inhalable anesthetics are very quickly cleared out of the body when a person goes from breathing the anesthesia to breathing normal air. Breathing is also one of the most natural, unconscious functions that our body possesses, which makes inhalation a very convenient route of administration.

However, inhalation of hot burnt solids causes significant damage to the airways. Inhalation of the chemicals in the propellants in compressed gases has been on the decline since manufacturers began adding bittering agents to the gases, thus making exposure to the substance highly undesirable. The dangers of this route of administration is that a person may experience **hypoxia**, low blood levels of oxygen, which can cause permanent brain damage. Many of the "rewarding" sensations associated with inhalant misuse are due to hypoxia.

Figure 11.5 Inhalation is often used for inducing anesthesia before surgery because of it's ease of administration and rapid onset.

Insufflation represents a different route of administration from inhalation because there are a few major differences of how the drug gets into the bloodstream. The substance is first pulverized into a fine powder then inhaled through the nostrils. Most of the solid sticks to the inside of the nasal cavity, where the chemical can then be absorbed by the thin blood vessels that line the inside of the nose. Insufflation, or "snorting," is a very common route of administration for the recreational use of cocaine, producing a very rapid drug effect within minutes. Because drugs that are insufflated are protected from first-pass metabolism, they yield a higher dose that affects the brain.

Transdermal is another parenteral route of administration that avoids processing by the digestive tract. In transdermal administration, a substance is placed on the skin surface, and over time the chemical diffuses through the skin into the blood vessels. The most common and well-known example of transdermal drug administration is the nicotine patch, a nicotine-replacement therapy. Transdermal patches have the advantage that the drug diffuses slowly over a prolonged period of time, up to 24 hours, minimizing the abuse potential of drugs taken in this route. However, because the skin is such an effective protective barrier against the outside world, there are major limitations to the types of drugs that can be delivered via transdermal

application: only small molecules and fatty lipids that are effective at milligram concentrations can be used.

Topical administration is closely related to transdermal administration, with the major difference being that topical administration does not significantly increase the blood concentration of the drug. Topical administration acts only locally, and some common substances that are applied topically include Tiger balm or Icy-Hot. Because chemicals introduced topically do not significantly get alter brain activity, they are not psychoactive.

The pharmacokinetic profiles of drugs are important for understanding their abuse potential, and potential strategies to develop effective treatments and medications to treat overdose.

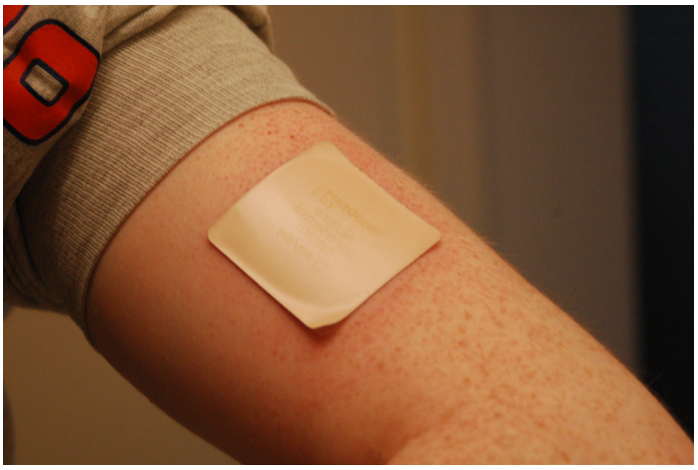


Figure 11.6 A transdermal patch that steadily delivers nicotine into the bloodstream.

11.2 Neural Circuitry Involved in Reward

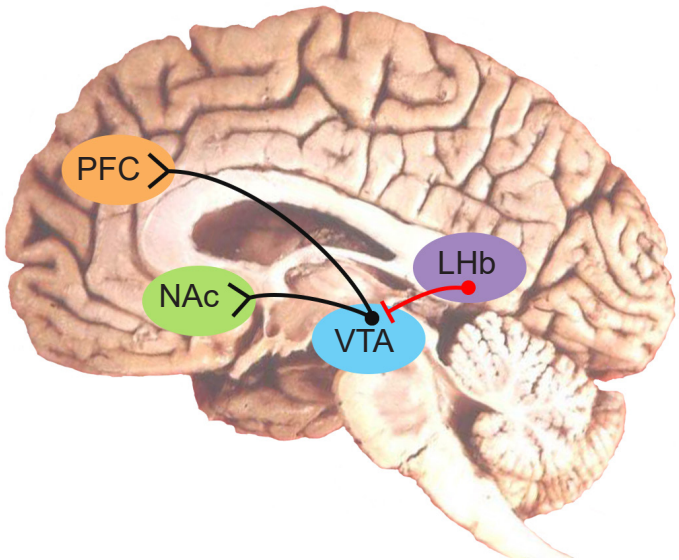
Having a sense of reward is a valuable adaptive trait. Consider two similar organisms living in the distant geologic past, both involved in Darwin's struggle to pass their genetic material to the following generation. One of these creatures gets an internal rewarding sensation whenever it drinks water while thirsty, eats high caloric-content food, or successfully reproduces. As can be expected, this organism is highly motivated to seek out those rewarding stimuli in the world. It is are more likely to crave sugary foods to maximize daily caloric intake necessary for activity and growth. The other creature does not experience pleasure from its actions, resulting in little motivation to seek out calories or to reproduce, for example. It stands to reason that the first of these two creatures has an evolutionary advantage to survive.

As can be expected, the regions of the brain that are concerned with reward are highly conserved through evolution. For us humans, our reward signaling neural circuitry starts with the **ventral tegmental area (VTA)**. Located in the midbrain, the VTA contains neurons that synthesize the neurotransmitter dopamine (DA), and is the largest group of DA cells (citation) in the brain. They are also called the A10 neurons. This population of DA cells similar to the cells in the substantia nigra, the midbrain neurons that are lost in Parkinson's disease (see chapter 10).

These VTA DA neurons send their axonal projections predominantly into two main areas. The **mesolimbic pathway** consists of dopamine producing neurons that release dopamine onto the cells in the **nucleus accumbens** (**NAc**, sometimes also called the **ventral striatum**). This seems to be the major pathway by which reward is mediated by the brain. Using microdialysis, it is possible to measure the dopamine concentration in the NAc during performance of some behavioral tasks. In studies with rats, engaging in rewarding activities such as eating or sex results in increased release of dopamine in the NAc. Similarly we see that release of dopamine into the human NAc is enhanced when we engage in activities we enjoy: binge eating sugary stuff, having sex, even playing video games. Notably, many drugs of abuse such as cocaine or amphetamine can induce a similar overflow of dopamine release when an animal is exposed to the drug.

The second area that receives VTA innervation is the prefrontal cortex (PFC), and this projection is called the **mesocortical pathway**. We think of the PFC as being involved in the conscious, decision making and inhibition of actions. It is likely that maladaptive changes in the dopamine projections into the PFC may explain the poor decision making and loss of

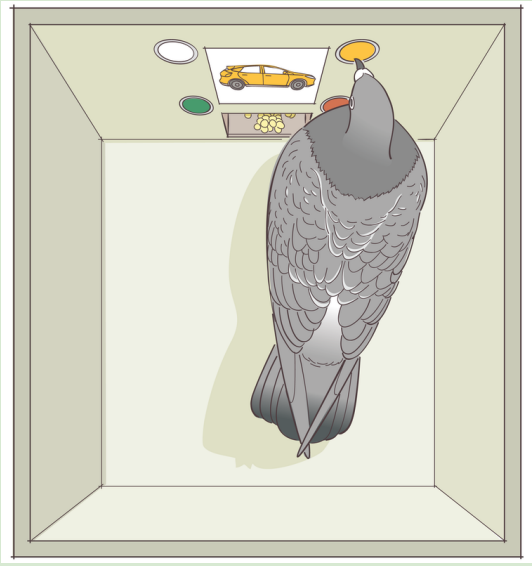
Figure 11.7 Neural structures involved in reward and aversion.



One of the earliest experiments demonstrating functional evidence of a "reward circuit" was conducted at McGill University in the early 1950's. Drs. James Olds and Peter Milner carried out surgery on anesthetized rats to implant metal electrodes into different areas of the brain. After recovery, the rats were placed into an **operant conditioning chamber**, also called a **Skinner box**. These special cages are equipped with a device which the subject can physically manipulate. For example, the cage may have a lever which the animal can push, or a hole into which they can poke their nose. In the case of Olds and Milner's experiments, a lever press led to activation of the implanted electrodes, resulting in neuronal activation. They had found that when the electrode was implanted into brain areas such as the corpus callosum or the hippocampus, the rats did not spend significant time pressing on the lever. However, the rats which had the electrodes implanted in the septal nucleus responded frequently on the lever. One such rat pressed the lever almost 2000 times in less than an hour, averaging nearly one response every two seconds! Electrical activation of reward centers is so desirable, rats will choose electrical stimulation over food even to the point of starvation.

This experimental paradigm is also called i**ntracranial-self stimulation (ICSS**). And while the original studies were conducted on rats, clinicians soon after tested their model in humans, with similar indwelling electrodes in different brain regions. Unsurprisingly, humans also have areas of the brain that are responsible for encoding reward, and when given the chance to electrically activate those areas, they do so extensively. When the experimenters brought in a tray of delicious food to the test chamber, the hungry patients - who hadn't eaten for upwards of seven hours - simply looked at the meal, but couldn't pull themselves away from the stimulation button long enough to eat. Artificial stimulation of these reward centers of the brain was more valuable to the patients than nutrition.

Figure 11.8 An operant conditioning chamber, where pushing a button or lever results in a response such as electrical activation of the brain.



control that addicts experience - their prioritization of immediate reward despite bad long term consequences (losing their job, alienating family, immediate health risks, for example). This connection has been heavily studied in the context of schizophrenia and other psychological pathologies.

Opposite of the reward pathway is a related set of neural circuits that signal an opposite reaction: **aversion**. If eating a sugary, chocolatey treat activates the reward circuitry, then eating a bitter-tasting bug activates the aversion circuitry. Just as with the VTA and related structures of the reward pathway, the ability to detect aversive stimuli is highly conserved throughout evolution.

For example, we know that noxious plant toxins, the kinds of foods that would have made our proto-human ancestors sick, are often bitter (see Chapter 9). Being able to detect these flavors as disgusting would be a pro-survival adaptation.

If the VTA to NAc projection is the reward pathway, the projections between the neurons in the **lateral habenula (LHb)** to the VTA is the "anti-reward" pathway. The LHb sends inhibitory GABA projections onto the VTA DA neurons, so when an aversive stimulus is presented, the LHb neurons increase in their activity, which decreases DA neuron excitability through GABA release, thus resulting in less DA in the NAc. These results were conducted with electrophysiological techniques and first demonstrated in monkeys in 2009. The researchers found that when a monkey was given a squirt of delicious fruit juice, LHb neurons spiked less, which relieved the inhibition on the VTA neurons, allowing them to increase in spiking - as can be expected when a rewarding stimulus is presented. On the other hand, when an aversive stimulus was presented, in this case an air puff to the face, the LHb increased in activity while the VTA neurons decreased in activity. Relatedly, if an animal is trained to expect a reward after some innocuous stimulus (Pavlovian conditioning), and that reward is withheld after presentation of the stimulus, the LHb increases in activity.

11.3 Molecular Pharmacodynamics

To talk more accurately about the action of drugs on a molecular level, it's important to establish some of the vocabulary that is used in pharmacology. As we begin discussing the action and movement of molecules, it's important to remember that the physical movement of molecules is more random than guided. When neurotransmitter molecules are released into the synapse, they move around in that space randomly. At many points in time, they float around the synapse, causing no effect. But sometimes, molecules will bump into a receptor, activating it as long as the molecule remains attached. Other times, those molecules will bump into reuptake proteins, large transmembrane proteins that decrease the concentration of neurotransmitters in the synapse by pulling the neurotransmitter back into the presynaptic axon terminal. And at other times, the molecule can even float outside of the synaptic space. Pharmacologists use the word **stochastic** to describe this randomness. The location of a molecule has no influence on where it will go next.

It is also helpful to become familiar with the **dose-response curve**, which is a graph that plots the activation of a receptor on the y-axis, and increasing dosage of a drug on the x-axis. In most cases, as neurotransmitter concentration increases, so does the activation of the corresponding receptor. The shape of the dose-response curve is described as sigmoidal, the S-shaped curve that is common to many fields of science, including ecology, microbiology, physics, and chemistry.

In molecular pharmacodynamics, we will describe three main categories of **ligands**, substances that are able to bind to receptors to form a ligand-receptor complex.

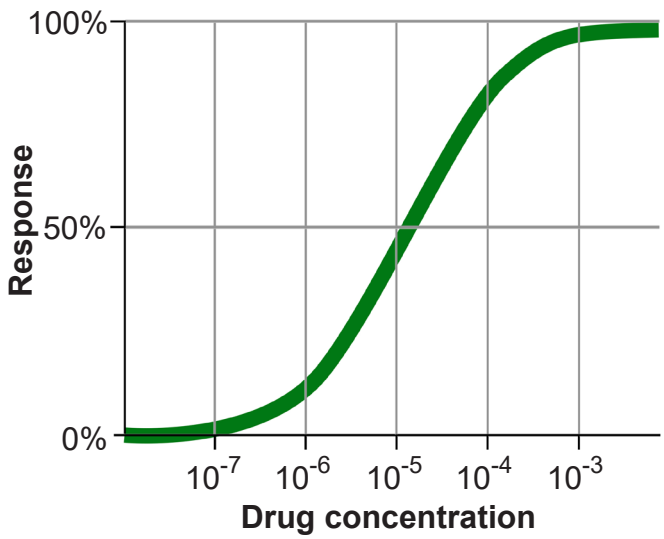


Figure 11.9 Generic dose response curve showing that drug effect (y-axis) increases as the dosage of drug (x-axis) increases.

1. **Agonists** are chemical substances that can activate receptors. Receptors themselves are large, transmembrane protein structures with a surface that is exposed to the extracellular side. On this exposed side, there is some three-dimensional arrangement of amino acids that the agonist is able to bind to. When the agonist-receptor complex is formed, the receptor physically changes conformation, which can then trigger a downstream reaction. In the case of an ionotropic receptor, the protein changes to either allow or hinder ion movement across the cell membrane. Nicotine is an example of an agonist that binds to ionotropic acetylcholine receptors. For metabotropic receptors, the activated agonist-receptor complex results in downstream activation of G-proteins, which then signal the cell to behave in a variety of ways. Morphine acts as an agonist at opioid receptors, which are metabotropic. The specific site on the receptor protein where agonists bind is called the

orthosteric site, or **active site**.

The interaction between a receptor and an agonist can be described by the analogy of a lock and key. The receptor is represented by a lock. Receptors are large protein complexes that span the cell membrane, just like a lock sits at the interface between two rooms. Receptors have an outward facing side that responds to the presence of an agonist by physically changing shape, similar to the way a lock will physically change when the correct key is inserted into the keyhole. Finally, receptors are very specific and will only respond to chemical structures that "match" the active site - the lock won't just open in the presence of any random key.

Agonists themselves can be divided into three different classes. A **full agonist** is a substance that can activate the receptor to the maximal degree at high concentrations. When we think of an exogenous neurotransmitter that is released by a neuron, such as glutamate, it activates the nearby glutamate receptors maximally. In a dose-response curve, full agonists substances that set the 100% value. The phrase **partial agonist** is used to describe substances that can also activate the receptor by binding to the orthosteric site, but are unable to fully activate the receptor, even at increasingly higher doses. On a dose response curve, the partial agonist activates the receptor to a lesser degree than the full agonist. Some partial agonists can be used clinically for treating a variety of disorders including anxiety, psychosis, and chronic pain. Another class of agonists are the **inverse agonists**, which causes an opposite response as an agonist.

2. If agonists are the chemical substances that activate receptors, **antagonists** are the substances that prevent agonists from acting.

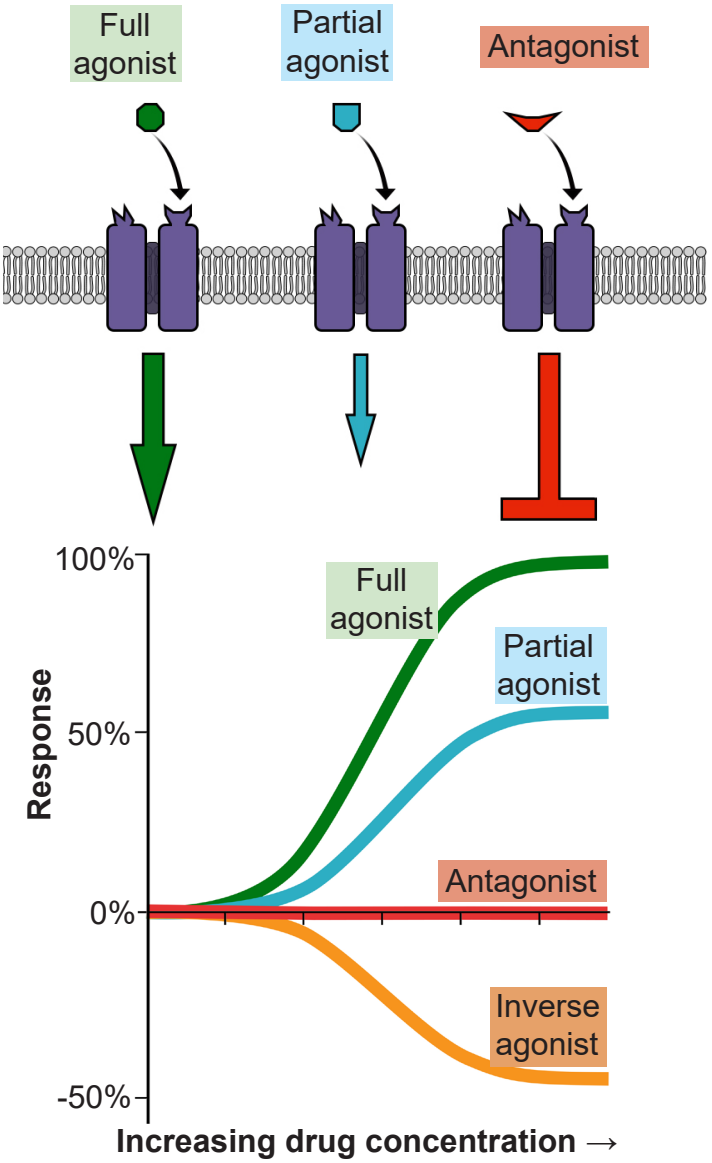


Figure 11.10 Different classes of ligands and their action on the receptors.

More specifically, **competitive antagonists** are substances that bind to the orthosteric site, the same site on the receptor where an agonist would bind. Because competitive antagonists physically block the active site, and because molecules move around randomly, the presence of the antagonist can prevent the agonist from activating the receptor in a concentration dependent manner: The more antagonist present, the higher the concentration of agonist must be in the synapse for activation. Therefore, adding a competitive antagonist to the system will shift the dose response curve to the right.

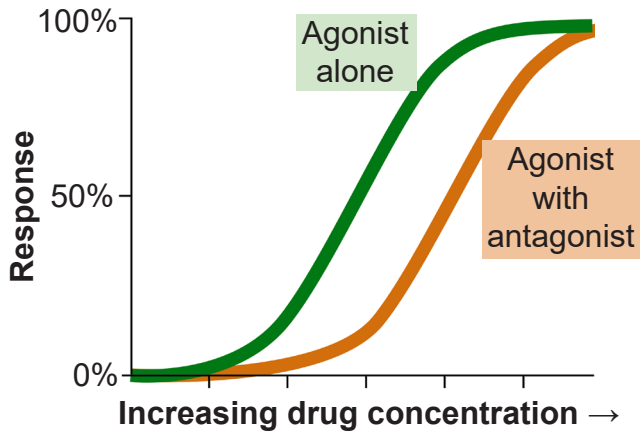


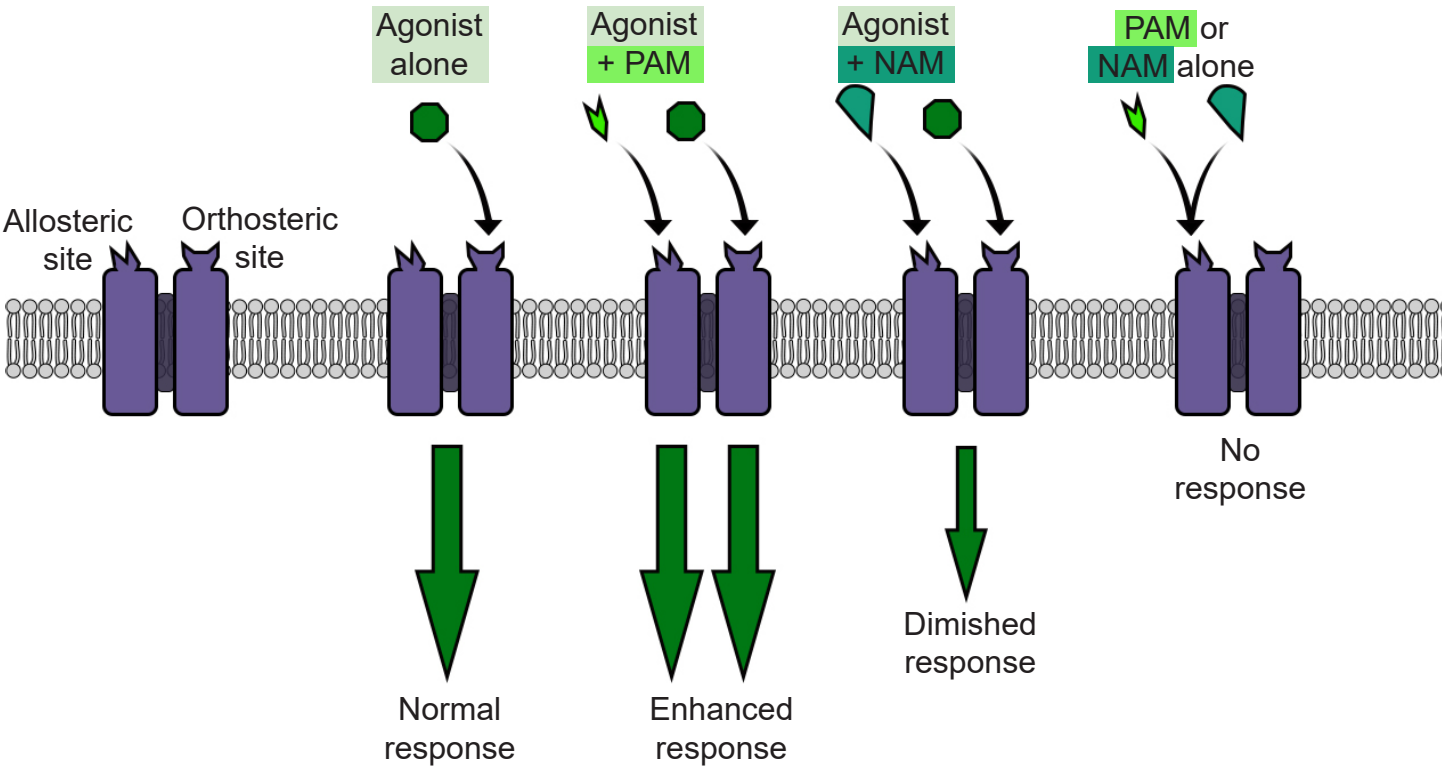
Figure 11.11 Dose response curves for agonist and for agonist plus competitive antagonist.

One of the most well known competitive antagonists is the drug Narcan (naloxone) which inhibits certain opioid receptors. One reason opioid overdose can be fatal is the drugs are able to shut down CNS drive of the respiratory system through acting as agonists at opioid receptors. In order to treat the person in this situation, you will have to find a way to block the action of the opioid molecules that are still acting on the receptors. When Narcan is given via an IV injection or nasal spray, it works within minutes to block the action of opioid drugs that are affecting respiratory control, thus restoring normal breathing.

It is important to note that in the absence of an agonist, an antagonist alone will have no effect on cell excitation. Therefore, if an antagonist alone induces a change in some cellular property such as excitability, then it is reasonable to conclude that the cell is being acted upon by some agonist at rest.

Separate from agonists and antagonists are a different class of chemical substances called **allosteric modulators** that also interact with neurotransmitter signaling. These chemicals belong to a unique classification of substances distinct from agonists and antagonists. Instead of binding to the orthosteric active site, they target a different region of the receptor, a site called the **allosteric site**. The allosteric site generally can

Figure 11.12 A receptor interacting with an allosteric modulator changes it's response following agonist binding to the orthosteric site.



be found on the extracellular side of the receptor, and like the active site, consists of a special three-dimensional arrangement of amino acids. When allosteric modulators bind to the receptor, they change the potency of agonists that activate the receptor. They can either increase the action of the agonist (**positive allosteric modulator**, or **PAM**) or decrease the action (**negative allosteric modulator**, or **NAM**.) In either case, some agonist, either endogenous neurotransmitter or exogenously-introduced drug, must be present for the allosteric modulator to have any effect.

Allosteric modulators act at the level of the receptor to modulate excitability of the neurons. For example, consider an ionotropic receptor that allows Na+ to move across the cell membrane when an agonist is present. When a PAM is present simultaneously with the agonist, more Na+ ions will move across the receptor than if the agonist was acting alone. But, when a NAM is present with that same agonist, fewer Na+ ions will move across the membrane, resulting in decreased excitability.

Many of these allosteric modulators are used experimentally but some are used clinically. The most well known allosteric modulators are the barbiturates and benzodiazepines. Both are PAMs which act at the GABA receptor, meaning that when a benzodiazepine (like Valium) is present, the same amount of GABA will have a stronger effect. Since GABA is an inhibitory neurotransmitter, enhancing this inhibition of neural circuitry can be an effective way to decrease brain disorders that result from overexcitation of the brain: anxiety, epilepsy, and insomnia.

A single pharmaceutical agent can have different actions at different classes of receptors. For example, the common antipsychotic clozapine is an antagonist at some dopamine and serotonin receptors, while also acting as a partial agonist at a different population of serotonin receptors. In pharmacology, drugs with many sites of action are commonly called **dirty drugs**.

11.4 Commonly Misused Substances

In the early 1900s, US laws regulating the manufacture, distribution, possession, and use of drugs were complicated. Certain substances were banned, but regulation of the substances was left largely to individual states. It wasn't until 1970 when a nationwide proposed regulatory measure would combine all of the current federal mandates regarding misused drugs into a single bill, and Richard Nixon signed the Controlled Substances Act (CSA) into law.

Under the terms of the CSA, the Drug Enforcement Agency and the Food and Drug Administration would be responsible for developing regulations about a wide variety of chemical substances. For each drug, they would be tasked with evaluating the health related properties of the drug:

1. Impact on health, including toxicity harms and addiction potential

2. Applications in the medical field

With these characteristics in mind, the substances were classified into one of five categories, rated Schedule I through Schedule V. Drugs put into Schedule I are substances that were perceived to have the highest addiction potential, most severe health risks, and no medical applications. On the other end of the spectrum are Schedule V substances, drugs with low addiction potential, few health risks, and legitimate uses within the clinical setting. Substances classified under Schedule I include heroin, for example, while a Schedule V substance might be an antiepileptic drug like Lyrica. Today, the list of controlled substances contains more than 400 chemicals. Interesting, neither tobacco or alcohol

Figure 11.13 The Controlled Substances Act was signed by President Richard Nixon in 1970.

is regulated under the terms of the CSA, despite these drugs being among the most misused in the US with tremendous health risks, high addiction potential, and little medical application.

There are probably thousands and thousands of **psychoactive substances**, chemicals that can act on the nervous system which can induce a change in behavior or mindstate. Here, we will focus largely on six classes of drugs.

Alcohol

According to the 2015 National Survey on Drug Use a nd Health, alcohol is the most widely used substance among these drugs, with more than 85% of adults reporting having used alcohol

at some point in their lives. Chemically speaking, many substances are classified as “alcohols,” but the most common misused is **ethyl alcohol** or **ethanol**. Ethanol is the easily-obtained substance with a highly intoxicating effect. The overwhelming majority of ethanol use is recreational. The World Health Organization reports that alcohol use is the leading risk factor for premature death among males aged 15-59 globally, making it one of the overall top causes of preventable death.

Acute exposure to ethanol has effects that differ depending on dosage. At low concentrations, ethanol can induce an elevation in mood, decreased anxiety, increased risk taking, slowing of reflexes, and impaired judgment. At high concentrations, ethanol can cause memory deficits, loss of consciousness, analgesia and areflexia, and possible death through respiratory depression. Chronic exposure to low levels of alcohol has some pro-health benefits such as decreased risk of death by cardiovascular events, but in general, these benefits are outweighed by negative health outcomes, such as the elevated risks of liver disease or cancer and alcohol-related violence or injuries.

Figure 11.14 Ethanol is the most commonly misused drugs, easily accessible in public settings, at grocery stores, sporting events, and others.

Ethanol is a metabolic byproduct synthesized by yeast as they break down sugars. It has the chemical structure of C2H6O, making ethanol a tiny molecule with properties similar to water, meaning that ethanol can diffuse easily through cell membranes. Whereas other chemical substances often have highly specific molecular or protein targets, ethanol has several: GABA receptors, glutamate receptors, potassium channels, serotonin receptors, and more. In general, ethanol has a depressant effect on neurons, but it has been demonstrated to activate DA neurons in the VTA.

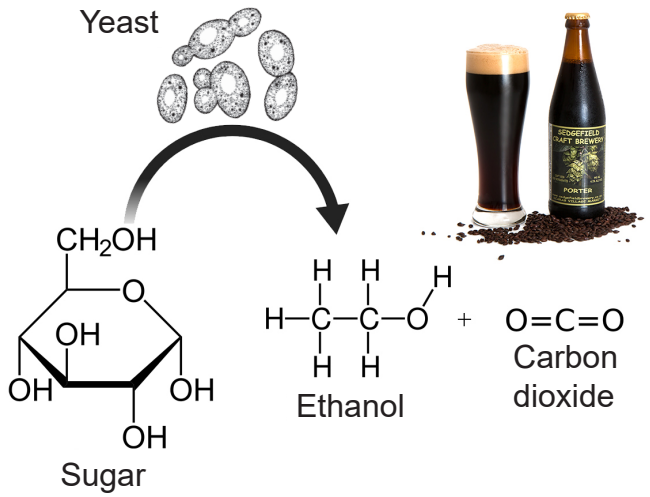


Figure 11.15 Ethanol is a byproduct of yeast metabolism. As carbon dioxide is dissolved into the solution, fermented products like beer become fizzy.

Nicotine

Nicotine is the main psychoactive ingredient of tobacco products. Nicotine is an alkaloid compound that is synthesized in nature by a variety of plants, with the tobacco plant having the highest nicotine content. Other plants of the Solanaceae family (Nightshade family),

such as eggplants and tomatoes, also contains nicotine, as it was probably an anti-herbivory evolutionary adaptation. The most common routes of administration for nicotine include inhalation (cigarettes, cigars, vape), transbuccal (chew, dip, snuff), or transdermal (in the case of the nicotine replacement therapeutic strategy, the nicotine patch). Nicotine is a highly addictive substance, often rated as addictive as cocaine or heroin. It is currently the leading cause of preventable death in the US, causing nearly one out of every five deaths.

Pharmacologically, nicotine has specificity as an agonist at **nicotinic acetylcholine receptors (nAChRs)**. These receptors are excitatory ionotropic receptors, and are expressed widely across the body. Nicotine is a stimulant, and induces release of norepinephrine by activating the sympathetic nervous system. When a person smokes a cigarette, they experience a rapid rush or a high within tens of seconds. Chronic smoking greatly increases the risk of developing lung, throat, and mouth cancers, coronary heart disease, or stroke.

Cigarette smoking over the past several decades has seen a sharp decline, in part due to the success of anti-smoking campaigns in changing public opinion toward smoking, and government-led changes in legislation resulting in stricter marketing laws, especially in regards to targeted advertising aimed at youth. In the mid 1970's, almost 30% of 12th graders reported smoking cigarettes daily, and that number has decreased to only 4% in 2018. On the other hand, the rising popularity of vape among teenagers has encouraged nicotine consumption through an alternate route of administration: Teen e-cigarette users are more than 3 times more likely to start smoking cigarettes compared to non e-cigarette users.

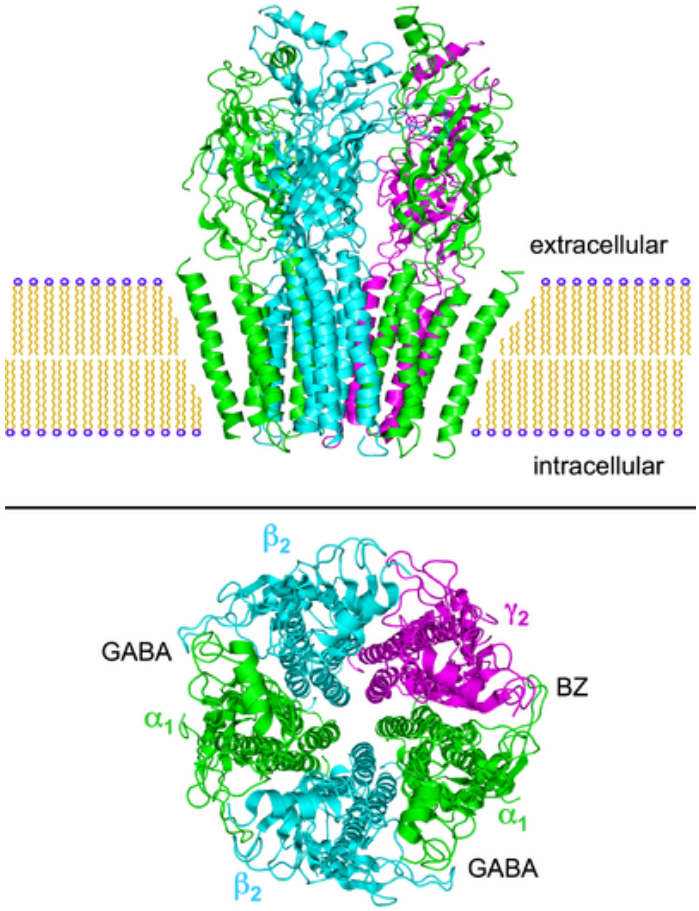


Figure 11.16 NAChRs are transmembrane receptors made up of five subunits.

Figure 11.17 E-cigarettes are a popular smoking cessation strategy, but many younger people have started using e-cigarettes recreationally.

Cannabis

Cannabis is a drug that is derived from the flowering buds or other parts of the Cannabis sativa plant. The most common route of administration is inhalation, as when people smoke a joint or a blunt. As cannabis becomes decriminalized or outright legalized across the US, oral preparations are becoming more commonplace. The effects of cannabis are often felt within minutes, and can include a sense of euphoria, relaxation, distortion in the sense of perception, and potentially a lightness (high) or a heaviness of body (stoned).

Although the cannabis plant contains several chemical substances, the main psychoactive ingredient is **delta-9-tetrohydrocannabinol**, or **THC** for short. THC, once it enters the bloodstream, is capable of activating receptors that our body uses endogenously, called the **cannabinoid receptors (CB receptors)**. There are two types of CB receptors, with **CB1** exerting the majority of the psychoactive effect and **CB2** expressed mostly throughout the immune system. Our bodies naturally produce substances that activate these receptors; these substances are called the **endocannabinoids**, or **eCBs**. Activation of the CB1 receptors, either through eCBs or THC, produces the elevation of mood.

Cannabis is currently classified as a Schedule I drug federally, but as of 2019, 10 states have legalized recreational use of cannabis, and 20 other states permit cannabis for medicinal use. Recently, Illinois has permitted the 55 medical dispensaries to sell recreational cannabis starting in January, 2020. The stigma surrounding the "evils" of cannabis use are largely a remnant of a morality propaganda campaign driven by the publishing of such "educational" films as Reefer Madness in 1936. Much of our current medical understanding of cannabis is that it is far less harmful than once believed, although it does have a negative impact on the developing brain, as significant adolescent exposure to cannabis increases the likelihood of developing psychiatric conditions such as schizophrenia later in life.

Figure 11.18 The flowering buds of the cannabis plant contain high concentrations of the psychoactive substance THC.

Opioids

Opioids are a class of drugs, either natural or synthetic, that can bind to and activate the body's endogenous complement of **opioid receptors**. The most well known natural opioid is **opium**, but the other popular drugs of this class include the street drug heroin or the medicinal substances morphine or fentanyl. Derived from the poppy plant, evidence of human use of opioids for both medical and recreational use dates back more than 5000 years to ancient Sumer.

As with the endocannabinoids described above, the body naturally produces opioid substances, **endorphin** being the most well-known. Activation of opioid receptors, either by

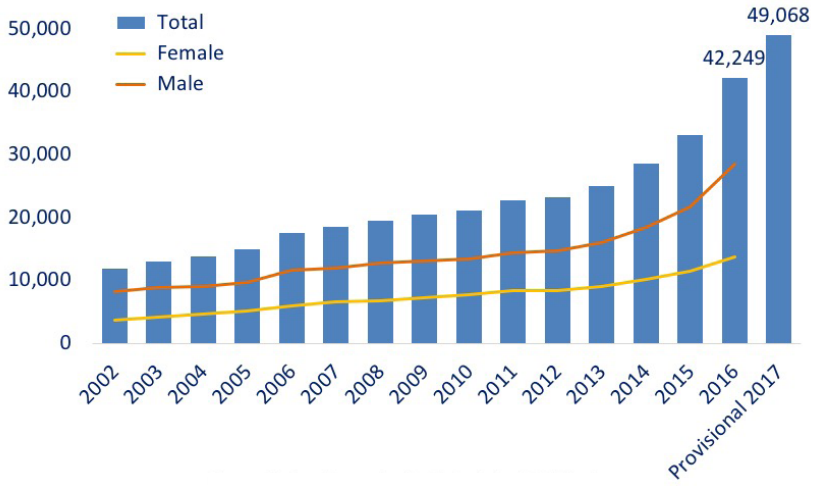


Figure 11.19 Opioid-overdose related deaths in the US have risen steadily over the past few decades, with a steep rise over the past five years.

endogenously produced opioids or exogenous opioids, induce a potent analgesia, dampening the sensation of pain. Opioid drugs the current gold standard for pain relief. They also cause sedation and feelings of euphoria. Heroin and morphine are frequently taken via IV injection, but many prescription opioids, particularly for long-term pain conditions, are consumed orally.

Over the past 20 years, there has been a dramatic increase in the number of opioid-overdose related deaths. The most recent spike in deaths is due to overdose on synthetic opioids such as the extremely potent painkiller fentanyl, which saw a 10-fold increase between 2013 and 2017. The so-called **opioid epidemic** kills an estimated 130 Americans each day.

Cocaine

Cocaine is a psychostimulant that is derived from the coca plant. The most common routes of administration are insufflation (powder cocaine), inhalation (crack cocaine), but it can also be solubilized and injected or taken via transbuccal and oral administration by chewing the leaves of the plant.

Cocaine is considered to be a potent **sympathomimetic**, meaning the substance activates the sympathetic nervous system. Therefore, the physiology of a person on cocaine will be similar to an activated "fight-or-flight" response: elevated heart rate and blood pressure, dilated pupils, and increased respiration.

Cocaine acts on a molecular level by functioning as an inhibitor of reuptake proteins. At the synapse, there are often presynaptically expressed proteins called **transporters**. These large membrane-spanning proteins function to move molecules from the extracellular space back into the presynaptic terminal, where they are then repackaged into vesicles for release at a later time. Cocaine is a **reuptake inhibitor**, physically preventing the transporter protein from clearing the neurotransmitters dopamine, norepinephrine, and serotonin out of the synapse. This leads to an elevation of the concentration of those neurotransmitters, which results in a greater likelihood that the synaptic receptors get activated.

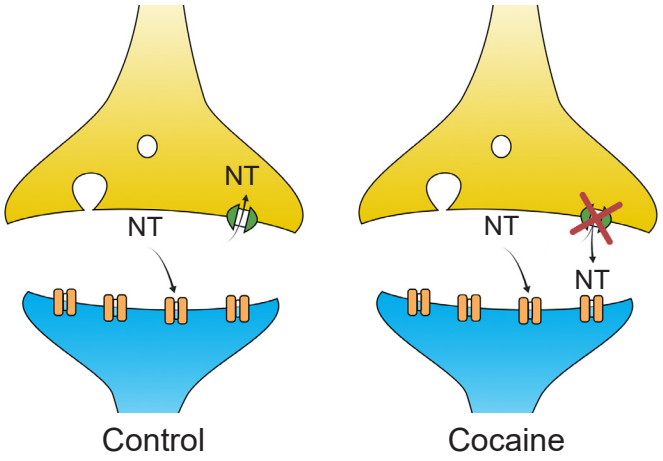


Figure 11.20 Cocaine blocks reuptake of neurotransmitters, increasing signaling.

Cocaine is a Schedule II substance. Although it has a very high addiction and harm potential, it has a handful of medicinal applications as well. For example, it is a potent local anesthetic and vasoconstrictor, so it is often used in surgery, especially in facial or nose surgeries.

Psychedelics

Psychedelics are the class of drugs formerly called hallucinogens. They can be either natural, such as **psilocybin** which is derived from mushrooms, or man-made, in the case of **lysergic acid diethylamide (LSD)**. Psychedelics can cause a person to experience visual distortions, synesthesia, an altered sense of self (ego dissolution), or sometimes a sudden connection with nature or a higher power. Chemically, many psychedelics are similar in structure to the endogenous neurotransmitter serotonin and therefore activate serotonin receptors.

Although psychedelics have been used for centuries in religious ceremonies, they have had a troubled history in the US. The adoption of LSD (acid) by the counterculture movement during the 1970's, secret CIA funded research, and some loosely-regulated research ethics in the clinical setting all led to notoriously negative press regarding the substance, tainted the general public opinion about this class of drugs.

Since 2010, there has been a growing body of evidence supporting the notion that psychedelics can serve a therapeutic purpose, most notably in treating psychiatric conditions such as PTSD, terminal illness-related depression, and drug addiction.

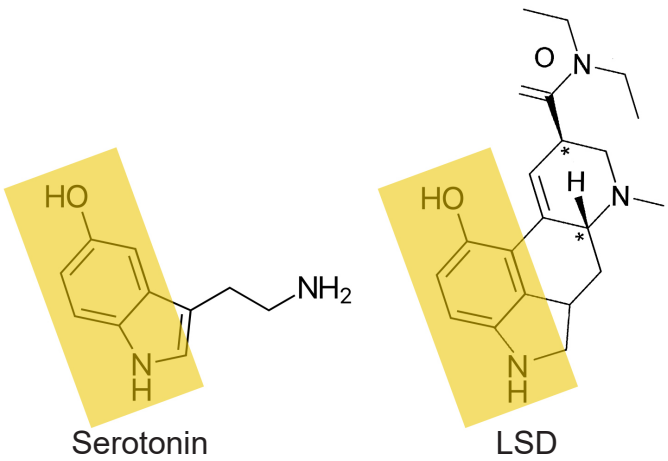


Figure 11.21 LSD is chemically very similar to the endogenous neurotransmitter serotonin.

11.5 Tolerance, Withdrawal, and Dependence

Pharmacologically speaking, **tolerance** can be defined as a decrease in the action of a drug due to repeat exposure. This phenomenon happens as a result of homeostatic adaptations that the body undergoes when it has been exposed to a substance for a prolonged period of time. In a person who is tolerant, they must take progressively higher doses of the drug to experience a desired effect.

Tolerance can be represented graphically by a rightward shift in the dose response curve. For a drug naive person, someone with no experience with the drug, a low dose of drug may have a strong effect. However, for a person who has become tolerant to the effects of the drug, that same low dose will produce a small drug effect. (The other way to think about the dose response curve is the following: For two people to experience the same magnitude of effect, a drug tolerant person would require a higher dose than someone who has never taken the drug before.)

There are different forms of tolerance that a person can experience.

In **metabolic tolerance** (sometimes also called **dispositional tolerance**, or **pharmacokinetic tolerance**), there is a decrease in the amount of the drug that gets to the site of action. The body becomes more efficient at eliminating the substance. Since exogenous substances usually get degraded by enzymatic reactions, an increase in the activity or number of enzymes will decrease the amount of the substance that is capable of acting.

The most common example of metabolic tolerance is seen in frequent alcohol use. The main psychoactive substance found in alcoholic beverages like beer, wine, or liquor is ethanol. Our liver, which filters circulating blood, contains enzymes which break down chemicals in the blood. Specifically, the standard degradation process of ethanol uses two enzymes. The first of which, called alcohol dehydrogenase, oxidizes molecules of ethanol into the toxic byproduct acetaldehyde. The second step uses the enzyme acetaldehyde dehydrogenase to convert the acetaldehyde into acetic acid, a generally harmless compound (the same chemical structure as vinegar) that is used by the body in a biochemical reaction that drives the citric acid cycle. Because most alcohol is taken via oral consumption, ethanol is subject to first-pass

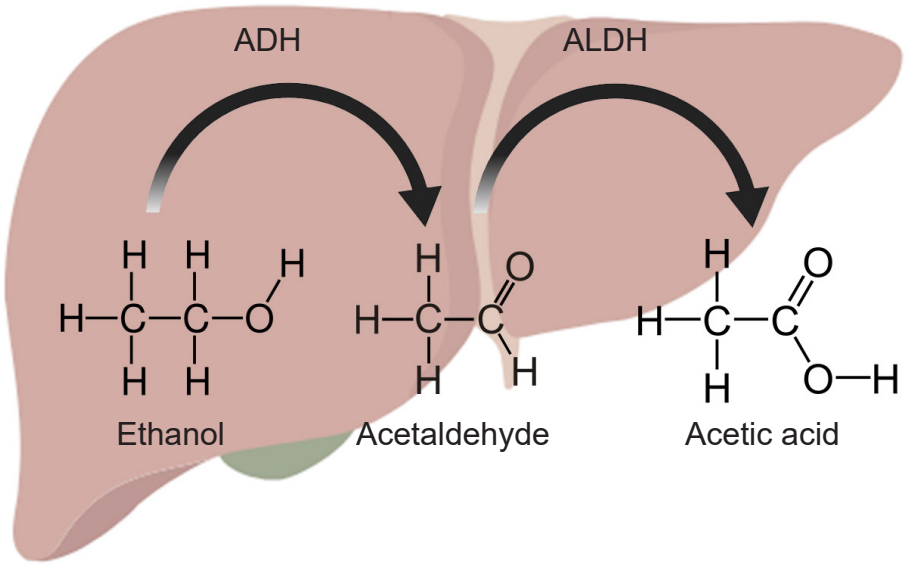


Figure 11.22 The liver breaks down alcohol in a two-step enzymatic process. Frequent alcohol use upregulates the speed by which degradation happens, causing tolerance.

metabolism before reaching systemic circulation.

However, after prolonged exposure to alcohol, the body undergoes homeostatic compensatory changes. The two liver enzymes responsible for degradation of ethanol increase in either function or amount. When these metabolic enzymes become upregulated, the circulating ethanol in the bloodstream becomes broken down more rapidly and more efficiently, resulting in a decrease in the amount of ethanol that reaches general circulation, and consequently, the brain. This explains why people who have never drank alcohol (low amounts / function of enzymes) may experience a stronger drug effect compared to someone who drinks regularly. It also accounts for why there is some genetic variation between a person's sensitivity to alcohol's intoxicating effect, as enzyme efficiency is partially encoded by genetics.

Another way by which tolerance can manifest itself is through changes at the level of the molecular components of cells. This type of tolerance is called a **functional tolerance**, or **pharmacodynamic tolerance**. This tolerance leads to a decreased sensitivity to the substance.

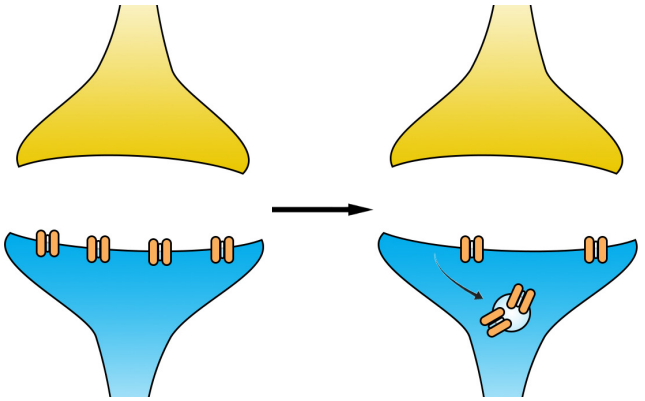


Figure 11.23 Pharmacodynamic or functional tolerance causes a synapse to be less sensitive to drug.

This functional tolerance can be characterized by a decrease in receptor expression after chronic exposure to an agonist. When the opioid etorphine is present in the body, it activates the opioid receptors. Frequent exposure to etorphine causes a molecular change in the receptors, which leads to internalization, the process by which cell surface-expressed receptors become taken into the cell. Since etorphine is only able to activate receptors on the surface, these internalized receptors diminish the strength of the etorphine signal, which manifests as tolerance. By downregulating the number of receptors that are expressed on the surface, the cell reacts to the environment by becoming less responsive in the face of drug.

Sometimes, the body physiologically prepares for the drug effect before the drug is even present. The way it does this is to initiate an opposite somatic effect from what the drug is expected to do. This "reverse" effect is likely a protective mechanism to minimize the drug effect, and these anticipatory bodily changes can be precipitated by drug-associated cues. For example, heroin causes analgesia. Cues that predict administration of the drug, such as preparation for an intravenous injection or being in a physical environment where drug is frequently administered, can cause the reverse physiological effect, hyperalgesia. Because of this, a person may need to take more drug to overcome the anticipatory changes. This unconscious form of learning leads to a form of tolerance that is dependent on Pavolvian cues; this is called **conditional tolerance**.

If not accounted for, conditioned tolerance can have potentially lethal side effects. A clinical description (Siegel, 2001) of a patient with pancreatic cancer describes one such "failure"

of tolerance. Bedridden and suffering from tremendous pain, a man was regularly given morphine in his bedroom, dark and filled with humming apparatus for his medical care. As expected, he developed a tolerance towards the morphine, requiring higher and higher doses in order to inhibit pain. One day, he went into the brightly lit living room to receive his normal dose of painkiller. Soon after, he died from opioid overdose.

Sensitization is a phenomenon related to tolerance, but the polar opposite: instead of a drug effect being lessened after chronic exposure, a sensitized person experiences an increase in drug effect. It is sometimes called “reverse tolerance.” Nicotine as well as many psychostimulants such as cocaine or amphetamine can produce psychomotor sensitization after repeated dosings.

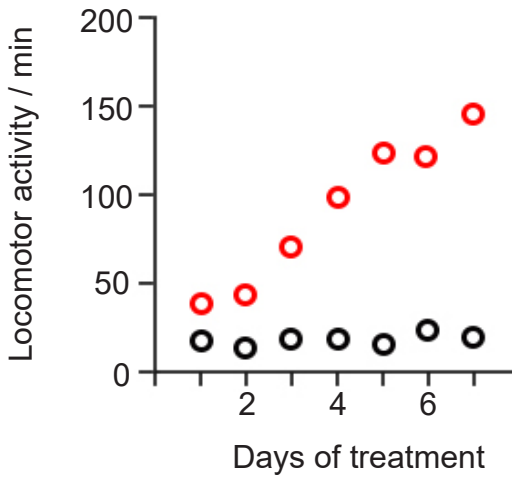


Figure 11.24 A rat exposed to amphetamine over multiple days shows locomotor sensitization, an increase in activity in response to drug.

Since mechanisms of tolerance result in long lasting homeostatic changes, they may lead to **withdrawal**, a set of symptoms that a person experiences when they are abstinent from the substance. This produces a highly aversive state that can be relieved by administration of more drug. The symptoms of withdrawal are frequently the opposite of the side effects of the drug. One example is the physiological response to heroin. When heroin is present, a person may experience euphoria, analgesia, relaxation, and constipation. But after the homeostatic changes have taken place resulting in downregulation of opioid receptors, a person may experience withdrawal symptoms such as dysphoria (depression), pain hypersensitivity, restlessness, and diarrhea.

It is good to note that tolerance does not only happen in the presence of agonists like ethanol or morphine. In fact, antagonists are well known to produce tolerance and withdrawal. If you have ever had difficulty staying alert a headache or without your morning coffee, you are experiencing the effects of withdrawal. Caffeine is an antagonist for the adenosine receptors. When the body is used to a certain amount of caffeine that blocks adenosine signaling, there are mechanisms that lead to homeostatic upregulation of adenosine receptors. And so when caffeine is no longer present, the endogenous levels of adenosine overactivate adenosine signaling, leading to the withdrawal symptoms.

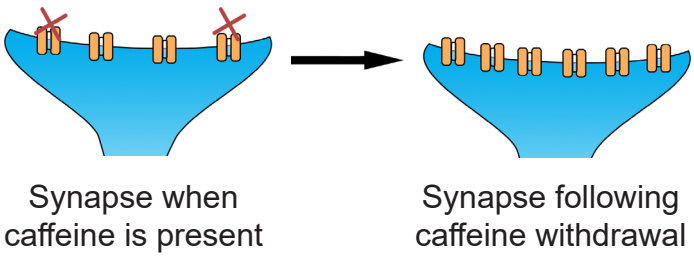


Figure 11.25 The regular presence of an antagonist like caffeine may cause increases of adenosine receptor levels postsynaptically.

When someone is withdrawing from a drug, they often feel the urge to take the drug. This is a sign of **drug dependence**. A physical dependence can be a physical dependence,

where the person seeks drug in order to either relieve the physical symptoms of withdrawal, or to experience the positive sensations associated with the drug. Alternatively, a psychological dependence can also develop, where the person has intense cravings for the drug such as a fixation on drug acquisition, or mood or behavioral changes in the absence of drug.

Long term drug use may cause people to become heavily dependent. These people often engage in recurring patterns of drug seeking and drug taking. Their patterns of use can be highly destructive, as they make drug acquisition as their main objective, fixating on drug taking even at the cost of their jobs, relationships, and even their own health. Because this form of pathological drug use is extremely harmful, much attention is focused on understanding the mechanisms that underlie this condition.

11.6 Theories of Addiction

"Why do some people transition from user to addict?"

Many people are able to use the most addictive substances such as cocaine or heroin casually without ever becoming an addict. For these people, their drug use pattern can remain recreational, even though they may use the substance frequently and report a great sense of satisfaction when they use the drug. On the other hand, there are some people who have a very high propensity to quickly becoming a pathological addict. One of the big challenges in drug abuse research is to understand the differences between these two populations - what factors, either environmental or genetic, are protective and prevent the transition from casual drug user to compulsive drug addict? And what are some ways we can use this knowledge to help addicts find ways to break their pattern of harmful compulsive drug seeking and drug use?

One possible way to get at the answer is to create animal models of drug dependence and try to determine the nature of the underlying biological changes that contribute to the addiction-like pattern: Do some cellular circuits change permanently? Are certain genes or proteins upregulated or downregulated? Can pharmacological or behavioral interventions change drug-taking or drug-seeking behavior?

Non-human models of addiction

One of the most instructive non-human animal models of drug use disorder is called **self-administration**. In a self-administration experiment, the subject (usually a monkey, rat, or mouse) gets a surgical implantation of an indwelling intravenous tube that is connected to a pump. Then, the subject is put into an operant conditioning chamber. When they press the lever or poke their nose into a hole, the pump triggers, resulting in an infusion of drug directly into the bloodstream.

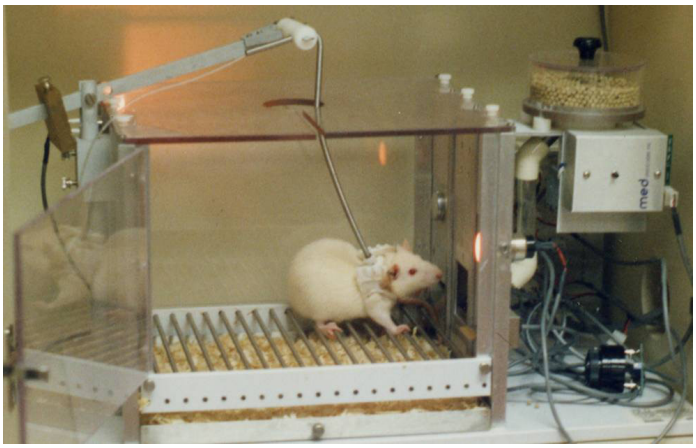


Figure 11.26 In a Skinner box, animals can be trained to give themselves a variety of drugs.

This experimental paradigm closely mimics human drug seeking behavior. For example, rats and mice will very quickly learn to press on the levers for intravenous infusions of cocaine and morphine. Many of them will give themselves drugs even despite severe adverse consequences, such as enduring foot shock - arguably similar to human behavioral patterns where addicts may do all kinds of behaviors for access to drug. Monkeys, if given the option to self administer for cocaine, must be restricted in the amount of drug they receive: if given unrestricted access, they give themselves so much cocaine that they have seizures and die.

A second experimental test created to study drug abuse is called **conditioned place**

preference (CPP). A CPP test works on the principle that animals learn to associate certain environments with positive feelings, and will prefer to spend significantly more time in those environments. To conduct CPP, a three-room testing chamber is used. This experimental chamber has two large rooms, each with characteristics that are unique to that room. For example, one room may have mesh flooring with dark-colored walls, while the other room may have woodchip flooring with light-colored walls. The two rooms are connected by a smaller, neutral room, and the rooms can be isolated from another by means of a dividing door. To conduct CPP, an animal will be given the drug of interest, then placed into one of the two rooms, where they are free to explore the room, but cannot leave. As a control, at a different time, they will be injected with an innocuous solution and placed into the other room. On test day, the animal will be put into the neutral middle room, and given the opportunity to move freely between the two chambers. If the animal developed positive associations with the room where they experienced the drug effect, then they will choose to spend significantly more time in that room. On the other hand, if they experienced negative feelings associated with that room, they will avoid being in the room. Whereas self-administration assesses "drug seeking," CPP measures "drug liking."

Below are three major biology-driven hypotheses that have been proposed to explain drug addiction. It's important to note that none of them tell the complete story - otherwise drug addiction will have already been cured! Also, these theories are not mutually exclusive from one another. A likely answer to curing addiction probably incorporates some combination of these approaches in addition to societal policy changes.

The "hedonia hypothesis"

What drives addicts to behave in the way they do? One of the earlier theories put forward to explain addiction is the **hedonia hypothesis** proposed in the late 1970s by Roy Wise. The hedonia hypothesis hinges on the assumption that dopamine is the "pleasure neurotransmitter," and that any substance or behavior that increases dopamine will be desirable. The theory is supported by some behavioral data. For example,

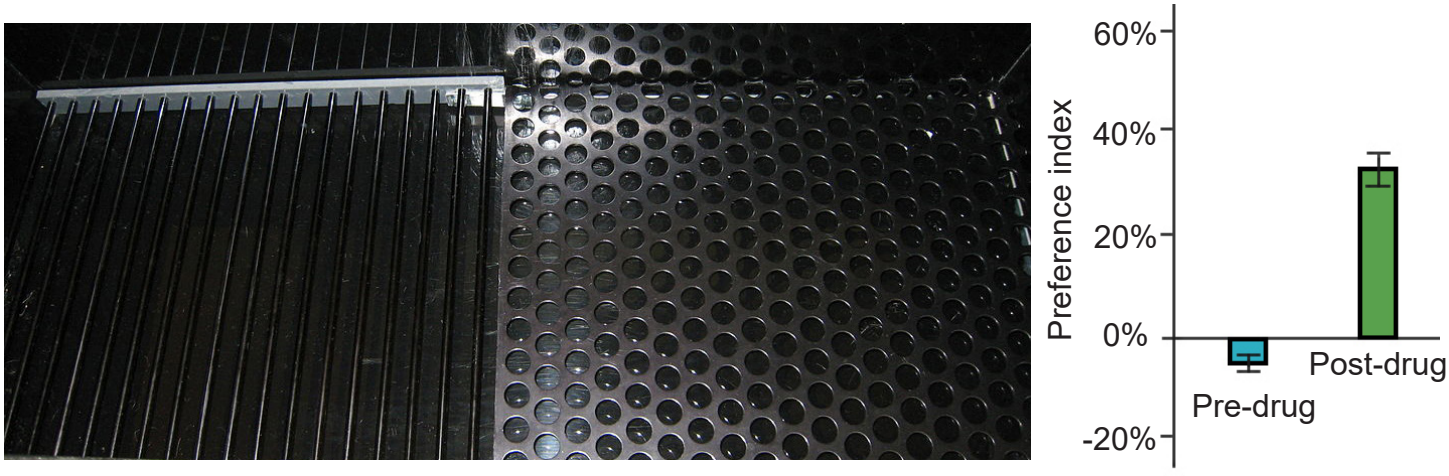


Figure 11.27 In a conditioned place preference test, increased time spent (preference index) in a chamber previously paired with drug is interpreted as an increase in drug liking.

using microdialysis, it was found that dopamine release by the mesolimbic pathway increases in the presence of many drugs of abuse. Additionally, administration of dopamine antagonists are able to block some drug taking behavior. This dopamine-centric approach to drug use disorder was so influential that pharmaceutical companies, while developing novel antidepressants, quickly dismissed drugs that increased dopamine out of fear of the risk of addiction potential.

Despite the attractiveness of the straightforward "pleasure neurotransmitter" model, the dopamine hypothesis does not give us the whole picture. For one, dopamine release at the mesolimbic pathway is only notable for psychostimulant drugs, such as cocaine and methamphetamine. But there are plenty of stimulants that still produce the significant overflow of dopamine without causing any rewarding sensation. Furthermore, if dopamine is the signal that mediates the positive hedonic feelings, one might predict that by blocking dopamine receptors chronically, you can also prevent long-term cocaine addiction, but this therapeutic strategy isn't very useful. Many of the most popular misconceptions about dopamine that you frequently see in the media take this early hypothesis as the truth. It even was featured as the cover of a 1997 *Time* magazine. That same year in an interview with the original proponent of the hedonia hypothesis published in *Science*, Roy Wise agreed with recent results that his model was incomplete.

The incentive sensitization model

A more recent theory to explain drug addiction is a model developed by researchers from the University of Michigan, Terry Robinson and Kent Berridge. Under their model of drug use, initially a person takes a drug because they like the positive, euphoria-producing side effects of the substance. But after chronic exposure to the substance, the person becomes fixated on obtaining the drug rather than how the drug makes them feel. At this point, they continue to use the drug even though they no longer enjoy the drug to the same degree that they once did. Instead of "liking" the drug, the addict fixates on "wanting" the drug.

The incentive sensitization model suggests that the drug-exposed brain develops an increased sensitivity to drugs, but also more importantly, drug related cues. Pavlovian learning takes place when an addict takes a drug. Each drug has an associated set of cues, and re-exposure to those cues can trigger relapse.

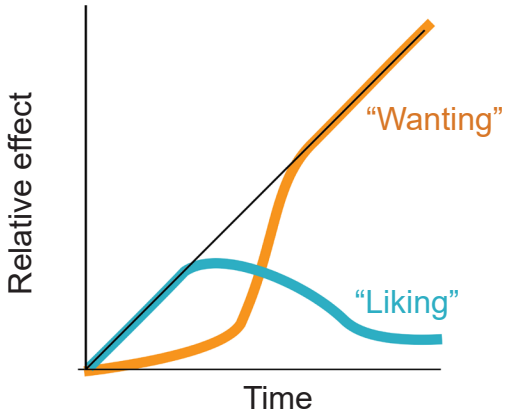


Figure 11.28 The incentive sensitization model suggests a difference between "wanting" and "liking" the effects of a drug.

Over repeated pairings, those cues trigger a progressively stronger response.

Their model accounts for why it is difficult to remain abstinent from an addictive drug for long durations of time. Despite a decrease in drug seeking behaviors, many addicts eventually **relapse**, or return to the previous destructive pattern of drug use. One phenomenon that makes staying drug free difficult is the observation that

prolonged abstinence can increase the intensity of drug cravings. This is called **incubation of craving**, and is seen with several different classes of drugs, notably, nicotine.

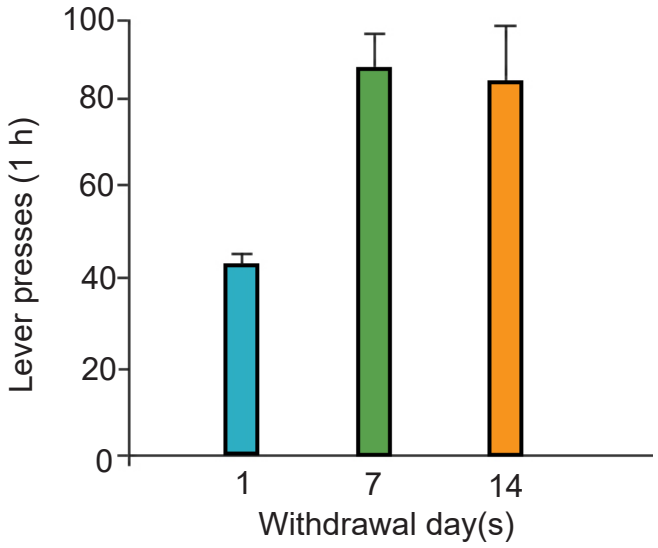


Figure 11.29 In a self-administration paradigm, lever pressing increases the longer an animal is abstinent from drug.

The Brain Disease Model of Addiction (BDMA)

Much of our current understanding of addiction is based on the principles that repeated drug use leads to biological changes in the brain that resemble aberrant learning. This model for drug addiction is called the **Brain Disease Model of Addiction**. Genetics has a strong influence on the likelihood that someone develops a disease, and addiction is no different. Inherent risk-taking behaviors, magnitude of a drug effect, and risk of relapse are all modified by genetics.

One of the major implications of the BDMA is that compulsive drug use and seeking is not a choice that a person makes. Addicts are not the way they are because of some poor moral guidance or a weakness of willpower. Addicts compulsively seek drug because of some underlying circuitry of their brain that makes cessation difficult. One substantial accomplishment of the BDMA is that public policy has adjusted to increase access to mental health resources for addicts to help cover the financial challenges of recovery programs (Mental Health Parity and Addiction Equity Act of 2008), and to improve the outcomes of non-violent drug offenders as they reintegrate with society outside of prisons.

The BDMA is not without fault, however. The major criticism of this framework to explain compulsive drug use is that most people stop their addictive use patterns by themselves spontaneously without any treatment. Another harm of the BDMA is that it puts heavy emphasis on biochemical therapies rather than on a public policy approach to drug addiction treatment. Lastly, the BDMA has not produced a successful therapeutic strategy to help with addiction.

The Open Neuroscience Initiative is funded by a grant from the Vincentian Endowment Fund of DePaul University.

Image credits

Cover https://pixabay.com/photos/poppy-poppy-butt-plant-flower-1540284/

11.1 https://pixabay.com/photos/hand-tablets-encapsulate-health-3632914/

11.2 The New Student's Reference Work, 1914

11.3 https://pixabay.com/photos/syringe-needle-medical-medicine-3902915/

11.4 https://pixabay.com/vectors/lungs-diagram-anatomy-body-human-41562/, https://upload.wikimedia.org/wikipedia/commons/7/7d/Gas_exchange_in_the_aveolus.svg modified by Austin Lim (Grapefruit) https://pixabay.com/photos/grapefruit-grapefruit-red-citrus-3434196/

11.5 https://upload.wikimedia.org/wikipedia/commons/3/3d/Preoxygenation_before_anesthetic_induction.jpg

11.6 https://upload.wikimedia.org/wikipedia/commons/9/94/Nicoderm.JPG

11.7 https://commons.wikimedia.org/wiki/File:Human_brain_midsagittal_cut.JPG modified by Austin Lim

11.8 https://commons.wikimedia.org/wiki/File:Operant_Conditioning_Involves_Choice.png modified by Austin Lim

11.10 https://commons.wikimedia.org/wiki/File:Scheme_simple_diffusion_in_cell_membrane-ca.svg modified by Austin Lim

11.12 https://commons.wikimedia.org/wiki/File:Scheme_simple_diffusion_in_cell_membrane-ca.svg modified by Austin Lim

11.13 https://commons.wikimedia.org/wiki/File:Richard_M._Nixon_signing_-_NARA_-_194618.tif

11.14 https://pixabay.com/photos/alcoholic-beverages-bar-beer-1845295/

11.15 https://upload.wikimedia.org/wikipedia/commons/0/04/Yeast_%28PSF%29.png https://pixabay.com/photos/craft-beer-dark-beer-porter-brewery-2468985/

11.16 https://upload.wikimedia.org/wikipedia/commons/4/46/NAchR_2BG9.png

11.17 https://pixabay.com/photos/ecigarette-juul-electronic-cigarette-3576177/

11.18 https://pixabay.com/photos/northern-lights-medical-marijuana-1826032/

11.19 NIH

11.24 Reprinted with consent of Dr. Eugenia Gurevich. https://www.sciencedirect.com/science/article/pii/S0028390817301715?via%3Dihub#undfig1

11.26 https://upload.wikimedia.org/wikipedia/commons/c/c4/SelfAdmin-2.jpg

11.27 https://upload.wikimedia.org/wikipedia/commons/0/0a/Conditioned_Place_Preference_-_Floor_Texture.JPG

Reprinted with consent of Dr. Ewa Niedzielska-Andres

https://www.sciencedirect.com/science/article/pii/S0166432819302803?via%3Dihub

11.29 Reprinted with consent of Dr. Xia Li https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5446935/

Chapter 12:

Sleep and the Circadian Rhythm

Austin Lim, PhD (DePaul University)
Edited by: Dana Simmons, PhD (University of Chicago)
Ben Marcus, PhD (University of Chicago)

Sleep is such an important part of our lives that a lack of it strongly correlates with negative outcomes on nearly every measure of health. People who sleep less than approximately 7 hours a night are at a greater risk for heart disease, stroke, asthma, arthritis, depression, and diabetes. Nearly 20% of all car crashes, both fatal and nonfatal, are attributed to drowsy driving. The cancer research branch of the World Health Organization has determined that disruption of regular sleep is "probably carcinogenic to humans," putting it in the same risk category as the infectious agents malaria and human papillomavirus (HPV), as well as the biochemical weapon mustard gas. Sexual health is affected by sleep deprivation as well, as men with the worst sleeping habits have significantly lower sperm counts, decreased circulating testosterone, and even testicular shrinkage.

Despite all that we know about the benefits of sleep, sleep is often the first time commitment to get cut, often getting squeezed as people stay awake later while waking up sooner. Consider that the CDC estimates that more than a third of American adults fail to get enough sleep each night. Almost 70% of college students fail to get the recommended amount of nightly sleep, and half of all college students report experiencing daytime sleepiness as a result.

The current medical recommendation is 7-9 hours of sleep each night. But why is sleep so important?

It is possible to study sleep using a combination of techniques; the output of sleep studies are visualized on a **polysomnogram** (pahl-e-SOM-nuh-gram; somn- is the prefix referring to sleep). Several physiological measures are taken in a polysomnogram, including heart rate, blood pressure and oxygenation level, respiratory depth and pattern, muscle activity, eye movement, and one of major interest to neuroscientists, brain wave activity.

While everyone knows what sleep is, it is useful to try to more precisely define sleep as a biological function. Sleep is characterized by the following:

A decrease in physical activity

Compared to waking behavior, a person's physical activity is greatly decreased when they sleep. While asleep, people are relatively inactive, and as a result the body uses about 10% less energy.

This is not to say that people do not completely cease all movement during sleep. It is very common to readjust posture many times in the middle of the night. Some people may grind their teeth together or talk in their sleep, sometimes carrying on full conversations by themselves! About 15% of people have experienced **somnambulism**, or sleepwalking: full on wake-like behaviors such as navigating down a flight of stairs or preparing a sandwich, performed entirely in the absence of intent of

Figure 12.1 Sleep is usually characterized by a decrease in physical activity, but some people experience somnambulism. Delacroix, Eugene. *Lady Macbeth Sleepwalking.* 1849-1850.

memory recall. Despite these rare occurrences of physical activity, the average movement of the person over a night's rest is still less than their average activity when awake.

A decoupling from external inputs

When we sleep, our conscious brains are "distanced" from the outside world. Sleep causes a heightened threshold for detection of stimuli, so we do not receive the same magnitude of inputs from our sensory systems as when we're awake. This is why someone else might have to talk loudly or even shake you physically to wake you up.

Changes in brain wave activity

Sleep was once thought to be a period of time characterized by low brain activity. After all, the person looks like they are not moving. Shouldn't brain activity be reflective of that decreased state of activity? With the advancement of EEG technology in 1924, and the rise of sleep laboratories in the 1970s, scientists who studied brain activity noticed that, at different times throughout the night, the brain of a sleeping person was very similar in activity to the brain of an awake person!

Chapter 12 outline

12.1 Phases of sleep

Each night when we go to sleep, our brains undergo a very stereotyped pattern of activity changes. At times, neurons in the cortex exhibit synchronized patterns of firing. And at other times, cortical activity looks very similar to an awake brain.

We can divide sleep roughly into two different phases depending on one of the first physiological measures that sleep scientists studied: eye movement. A study published in 1953 used a device to detect eye movement while a person was asleep. Interestingly, they noticed that at some points in the night, usually first occurring around three hours after falling asleep, the patient's eyes would dart rapidly and jerkily back and forth, a pattern of activity that the University of Chicago researchers called **rapid eye movement (REM)**. This first period of REM activity lasted for about 20 minutes, after which the eyes would stop moving again. This activity pattern repeated every hour or two for the rest of the night. They used eye movement to separate sleep into two phases: **REM sleep**, and **non-REM sleep (NREM sleep.)**

In addition to eye movement, they observed and measured other physiological behaviors. Respiration rate and heart rate both increased during the REM phase of sleep and dipped during NREM sleep. They also (rudely) woke up patients throughout the two phases of sleep, and found that patients were more likely to recall dreams with visual imagery if their REM sleep was interrupted. Those woken during NREM sleep were less likely to recall dreams, hinting that dreaming is more likely to happen during REM sleep.

While eye movement could differentiate between two phases of sleep, another common diagnostic technique, **electroencephalography (EEG)**, could further subdivide NREM sleep. EEG measures electrical activity at the scalp, detecting the firing of large numbers of cortical neurons. Using EEG, scientists discovered three distinct NREM phases based on neuron activity patterns: **NREM1**, **NREM2**, and **NREM3** sleep. Currently, readings collected via EEG are considered to be the gold standard for measuring the stages of sleep.

But before we describe EEG traces while asleep, we should describe the EEG of a person who is awake. Usually, the awake EEG is dominated by high-frequency waves, falling in

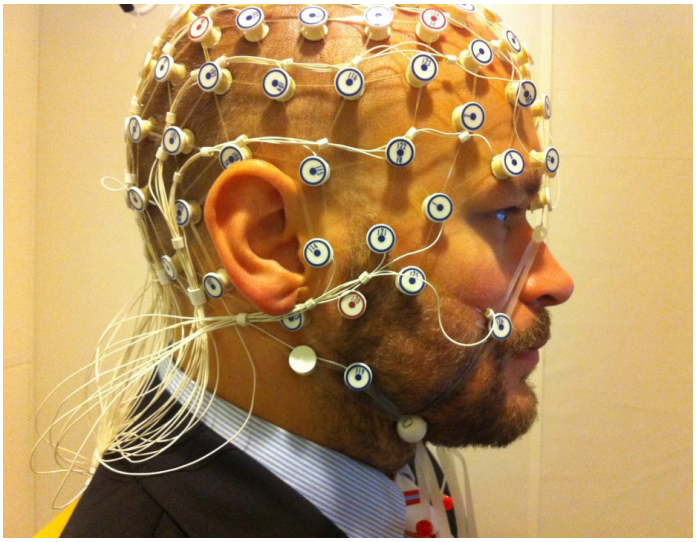


Figure 12.2 In an EEG, electrodes are placed on the head that can detect and record neuronal activity of the cortex.

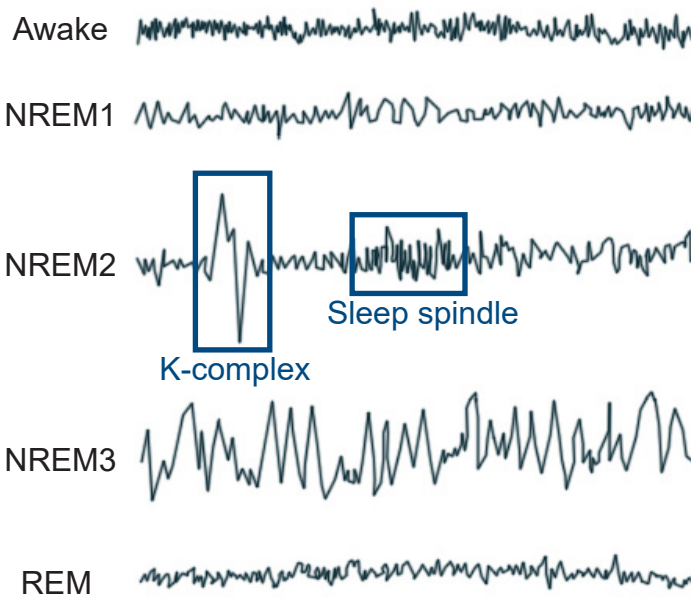


Figure 12.3 Throughout the night, the EEG shows that the brain cycles through different patterns of activity.

the **beta** band range of frequencies: between 13 and 30 Hz. The proportion of neurons firing at the beta frequency increases with attention and mental activity: When a person is concentrating on a task, such as reading a textbook, the beta wave frequency dominates.

NREM1 is the earliest stage of sleep. It's also described as relaxed wakefulness, drowsiness, or light sleep. During NREM1, a person's muscles are still somewhat active, their eyelids may open and close every so often, and they may still respond to questions. In NREM1 sleep, the beta frequency amplitude decrease as the slower frequency alpha waves (8-13 Hz) increase in amplitude. Late in NREM1, theta waves (4-8 Hz) become more prevalent. Basically, the deeper into NREM1 sleep a person becomes, more waves with lower and lower frequencies start to emerge.

During **NREM2** sleep, theta waves predominate. In a healthy adult, about 50% of a night's sleep is spent in NREM2. NREM2 is characterized by the appearance of two patterns of activity that interrupt theta activity. **K-complexes** are large amplitude events that are observed about every minute. These are the largest amplitude events in a healthy human EEG. Following a K-complex you may see a **sleep spindle**, a high-frequency burst of rapid neural activity in the low beta range that lasts for about a second. It is unknown exactly what the function of these sleep spindles are, but some research suggests they may be involved in memory processes or to minimize perception of outside noises, which can help a person stay asleep even in the face of disruptive stimuli.

NREM3 is also called **deep sleep**. At this phase of the night, a person's physiological activity drops to its lowest point of the night: heart rate, respiration, blood pressure, and metabolism all reach minimum during NREM3. In this stage of sleep, many of the cortical neurons fire in synchronicity with one another, and the subsequent change in potentials cause large

amplitude deflections in the EEG, at the low-frequency delta band (1 - 4Hz). Because the delta frequency is much slower than frequencies detected in lighter stages of sleep or wake, NREM 3 is also called slow wave sleep.

The EEG trace of a person in REM sleep is quite the opposite of what is seen in deep sleep. Instead of large amplitude events at a low frequency, the REM brain has a lot of low amplitude events at a high frequency. In fact, the brain in REM sleep has a pattern of activity that is more similar to a person who is awake than asleep! Because of this asynchronous firing activity, REM sleep is sometimes also called **paradoxical sleep**.

To illustrate the stages of sleep that a person experiences each night, we can use a **hypnogram**. These charts to plot time on the x-axis, and stage of sleep on the y-axis. Awake is represented at the top, and deep sleep is at the bottom.

For an average night's rest, neural activity will fluctuate through the four phases relatively predictably. When a person first falls asleep, they will move from NREM1 down through NREM2 then NREM3, before coming back out of deep sleep progressively back to NREM1. After NREM1, they may enter REM sleep before transitioning back through the stages to NREM3 again. This cycle of activity repeats roughly every one and a half hours.

People spend a larger percentage of each cycle in deep sleep and very little time in REM sleep early in the night. On the other hand, in the last few cycles before waking up from a full night's rest, people spend a larger percentage of of each cycle in REM sleep, and almost no time in deep sleep.

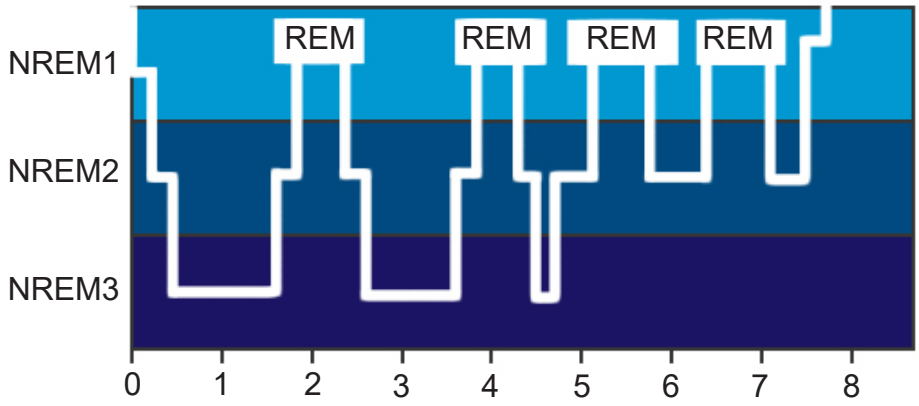


Figure 12.4 A hypnogram can be used to visualize the time spent in each phase of sleep over the night.

12.2 Why do we sleep?

All organisms that we know of experience some type of sleep. But we still haven't figured out exactly why animals sleep. Here, we will discuss three theories that have been proposed to explain sleep. None of these theories alone fully explains the complex phenomenon of sleep, and they are not mutually exclusive. The most likely reason we sleep is probably some combination of the following three theories.

Recuperation Theory

The recuperation theory of sleep is centered around the idea that being awake is stressful and exerts a physically demanding toll on the body. The body therefore needs a period of time when energy usage decreases and the body's natural repair systems can work without disruption. Sleep is how the body "wipes the slate clean" and resets.

Evidence for the recuperation theory comes from experiments tailored around the idea of looking at what happens when a person doesn't get sufficient sleep. As anyone who has ever pulled an all-nighter can attest, a single night of sleep deprivation often leads to significant psychological changes, including anxiety, irritability, and mood swings. Staying awake even longer than 24 hours can cause more severe changes in mind state, such as temporary psychosis, hallucinations, or delusions.

The recuperation theory is supported by several pieces of evidence, three of which we will discuss.

1. Enhanced metabolic cleaning during sleep. When you're awake, your cells produce several biological waste products during cellular respiration. These chemical byproducts can be potentially toxic to the body when they

Case study: Peter Tripp

In 1959, New York City radio DJ Peter Tripp ran a publicity stunt to raise money for the charity March of Dimes: he stayed awake for 201 hours. Sitting inside a glass booth in the middle of Times Square, Peter played music and broadcasted his experiences across the airwaves. The first night of sleep deprivation wasn't awful, but the next seven consecutive days and nights were a real challenge, for both Tripp and the doctors who kept an eye on him.

Just a few days in, Peter began experiencing severe psychological side effects. Tripp developed intense paranoia: He mistook his psychiatrist, who was wearing all black, for being the undertaker here to collect Peter's body for a funeral. Tripp even began hallucinating, seeing spiders and rodents in his clothes - prompting him to strip naked and run into the street screaming.

Some believe the long-term sleep deprivation severely affected his brain beyond this temporary psychosis. After the experiment, Tripp was involved in a huge commercial bribery scandal, lost his job, and got divorced (Of course, life has many variables, so we'll never be completely sure of the precise effect long-term sleep deprivation may have had on his relationships and decisions.) Today, the Guinness Book of World Records no longer allows people to compete for the prolonged-wakefulness record due to health concerns.

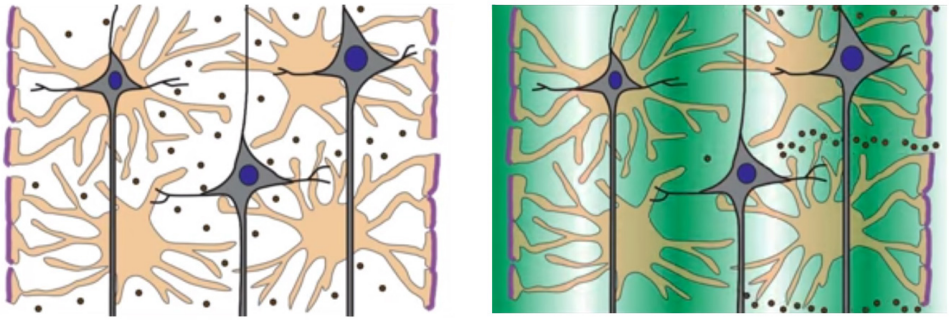


Figure 12.5 During sleep, the glymphatic system increases flow of CSF (green waves) that are able to wash out the amyloid-beta protein (black dots) from the interstitial space.

accumulate. The **glymphatic system**, a sort of cellular rinse that floods the extracellular space with CSF, clears these cellular waste byproducts when they accumulate in the brain. When we sleep, the extracellular space expands by about 60%, which increases the ability for CSF to penetrate deeper into the brain tissue.

One byproduct of interest is the molecule **beta-amyloid**, a protein that exists in the healthy brain. But, accumulations of beta-amyloid are found at high levels in the post-mortem analysis of brains from patients with Alzherimer's disease. During sleep, the glymphatic system's "rinsing" activity washes away beta-amyloid from the interstitial space and degrades the protein.

2. Immune system function improved with sleep.

Sleeping more enhances your ability to fight pathogens. Sleeping for fewer than 6 hours per night increases the likelihood of catching a variety of transmissible illnesses, such as cold, flu, or gastroenteritis. Sleep increases the effectiveness of vaccines, and each hour of sleep over 6 hours increases said effectiveness by about 50%.

3. Increased production of growth hormone during deep sleep.

Most cells of the body are turned over regularly as cells die and get replaced. One of the signaling molecules that encourages the replacement process is **growth hormone (GH)**, which enhances cellular repair, muscle and bone growth, and protein synthesis. Normally, throughout a 24-hour day, GH is produced and released throughout the body by the hypothalamus. The largest wave of GH production occurs early in sleep, during NREM 3 sleep. During this huge burst of GH release, a person's circulating plasma GH concentration may be 10 times higher than at baseline.

While the support for the recuperation theory is generally true for almost all people, there are about 1% of people who seemingly gain the restorative benefits of sleep, even with fewer than 6 hours of sleep each night. They may wake up at 4:30 in the morning feeling completely refreshed. And yet, despite getting so little sleep, these short sleepers have similar health outcomes with respect to body mass index and psychiatric measures such as depression and overall optimism.

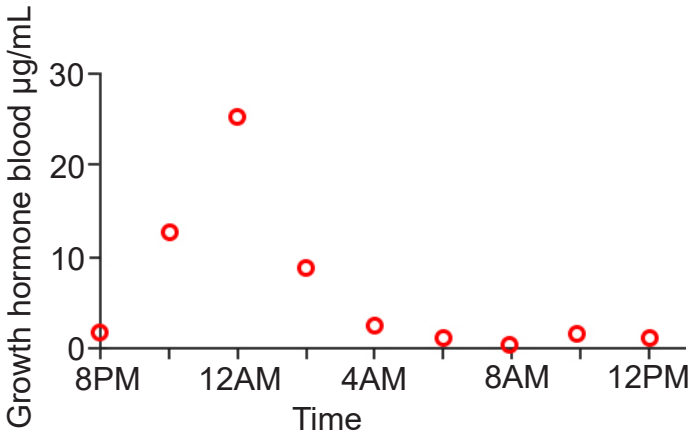


Figure 12.6 Growth hormone levels peak early in the night during deep sleep.

Something about the circadian rhythms of these short sleepers allows them to “maximize” their sleep efficiency. Many have very short sleep latencies, meaning they fall asleep within minutes after lying down - as quickly as someone with narcolepsy. They also spend a larger percentage of their night in deep sleep and REM sleep while minimizing NREM1 and NREM2.

Evolutionary Adaptation Theory

The evolutionary adaptation theory is the idea that animal sleep patterns are different across species for reasons that most efficiently benefit each animal. Over millions of years of evolution, individuals with the most ideal sleep patterns have an advantage, and their sleep habits will be selected for in the following generation.

For example, consider humans. As an animal highly dependent on light and the visual system for navigation and accurate performance of tasks, the dark is a very dangerous time to be active. The risks of wandering off a cliff, running head-first into a tree while escaping a predator, or eating a wrong-colored poisonous berry would all be elevated in the dark. We benefit from behaviors that minimize those risks, such as inactivity until the sun rises. During this inactive period, sleeping decreases our metabolism and our body’s need for energy.

Figure 12.7 In a traveling pod of dolphins, the animals on the edges sleep with half of their brain in order to remain a vigilant lookout.

Humans are just one animal that has an evolutionarily fine-tuned sleep pattern. If you look across the animal kingdom, you’ll find all varieties of sleep behaviors that are best fitted to the needs of the individual species. In the wild, for example, dolphins are generally prey. They evolved with the ability to put one half of their brain to “sleep” at a time, allowing the “awake” half to keep an eye out for potential predators. Small prey animals, like squirrels, are faced with the threat of being attacked at night. For them, remaining very still, quiet, and hidden improves their survival. Tigers, the top alpha predators in any ecological niche, have almost no predators to hide from, allowing them the luxury of sleeping up to 20 hours a day.

This evolutionary adaptation theory argument has a major weakness, however. In almost all animals, sleep represents a period of time when an organism’s ability to use their sensory organs to detect the hallmark signs of an approaching predator, like the flurry of feathers from a hungry owl or the soft padding of a wolf footsteps, decreases drastically. For an animal that can’t hide very effectively, sleep represents a period of vulnerability, as they would be unable to sense incoming threats.

Figure 12.8 Alpha predators have no fear of being attacked, and are given the luxury to sleep for prolonged periods of time.

Brain Plasticity Theory

The brain plasticity theory suggests that the brain needs some period of time for critical changes to occur. During sleep, circuits in the brain undergo consolidation processes that are important for memory formation. For example, academic performance and examination grades worsen as a person's nightly sleep decreases.

Both the REM and NREM3 phases of sleep are important for different types of memories, and studies suggest that **declarative memory**, pieces of information about facts, benefits more from slow-wave sleep while **procedural memory**, the learning of motor skills, is enhanced by REM sleep. Although the exact mechanisms about how sleep improves memory are unknown, we theorize that brain activity during sleep helps move memories held in "temporary" areas into areas of stable, long-term storage.

The evidence in support of this theory starts with looking at the brains of newborns. When you were first born, during those first few weeks of life, you slept close to nearly 70% of the day, almost 17 hours! At this point in your life, your brain starts to experience all sorts of new sensations: your eyes detect visible light for the first time, the skin feels the air blow past it, and the ears sense new frequencies and combinations of sound waves. As a result of these stimuli, scientists hypothesize that your brain undergoes as many learning events as rapidly as possible. This rapid learning helps you remember what you learned each day so you can respond to your environment as you grow up.

Figure 12.9 According to the brain plasticity theory of sleep, newborns spend most of their day sleeping in order for their brains to adapt to all the new senses they are experiencing.

12.3 The circadian rhythm

Almost every living organism that we know of on Earth exhibits some sort of cyclic pattern of activity that closely matches the rising and setting of the sun. The discovery of 24 hour patterns of behaviors began in 1729 with the French scientist Jean-Jacque d'Ortous de Mairan, who documented the movement of the *Mimosa pudica* plant. This unique organism was chosen since the plant exhibits **heliotropism**, light-seeking movements. In particular, this plant opened up the leaves in the daytime to capture sunlight, then closed at night, minimizing predation. When the plants were put into a dark room with no exposure to sunlight, to his surprise they still opened and closed their leaves in time with the clock. Mairan concluded that the plant did not change its behavior in response to light, but rather in response to some internal 24-hour clock. His work laid the foundations for future **chronobiologists**, scientists who study day-night dependent periodic phenomena in living beings.

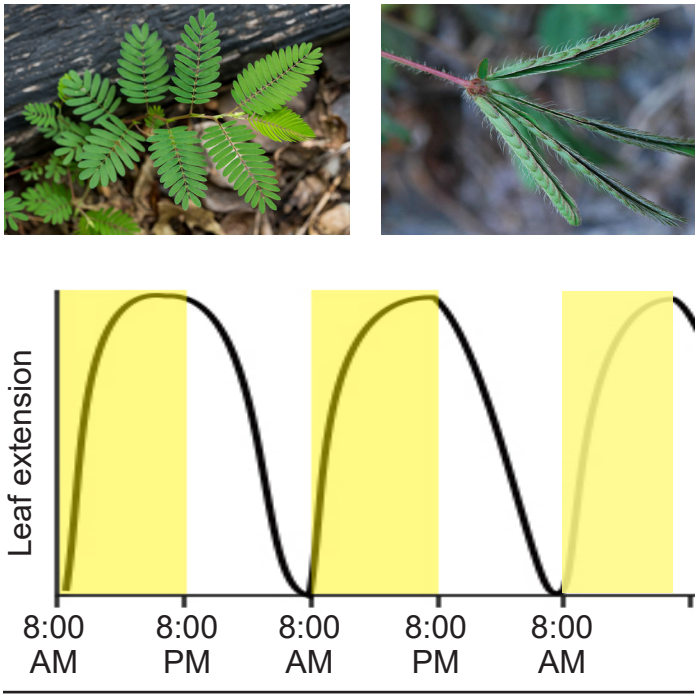


Figure 12.10 The leaves of the *Mimosa pudica* plant open and close their leaves in time with the 24-hour cycle of sun rising and setting, even in complete darkness.

Any behavior or physiological measure that intrinsically cycles on a 24-hour pattern is said to be a **circadian rhythm**. The word circadian comes from the Latin words circa- meaning "around", and diem meaning "day". Compare this with an ultradian rhythm, any cycle that is faster than 24 hours, such as the cycling between deep sleep and REM sleep every 90 minutes. Alternatively, compare with infradian rhythms, patterns that are longer than 24 hours, like the 4-week-long human menstrual cycle.

Although we mostly think of the circadian rhythm in the context of sleep and wake, many other physiological measures fluctuate reliably throughout the day. Blood pressure peaks at 11 AM, making the morning the time with the highest risk for cardiac events. Body temperature dips late in the evening, putting the body into a low-energy state that helps promote sleep. Withdrawal reflexes peak around midnight, hunger-driving hormone production rises before lunch and dinner, attention is usually highest in the morning - all manner of behaviors that rise and fall depending on the time of day can be said to be a part of a circadian rhythm.

Bizarrely, even organisms with a life span shorter than a full day still exhibit 24 hour cycling patterns of circadian rhythm-like behaviors. Cyanobacteria, or blue-green algae, are capable of a process called nitrogen fixation, where they convert atmospheric nitrogen into organic compounds like amino acids. These organisms are capable of asexual reproduction every 6

hours or so, but they still have nitrogen fixation patterns that align with daily patterns of light and dark.

Circadian rhythms on a behavioral level

Luckily, a person's circadian rhythm is not permanent. Anyone who has traveled overseas to a different time zone for more than a few days has experienced that uncomfortable sensation called **jet lag**, where a person experiences psychological symptoms such as difficulty concentrating and mood swings, and physical symptoms like daytime fatigue, insomnia, and gastrointestinal distress (nausea, constipation, or diarrhea). Jet lag happens when there's a mismatch between the internal environment and the signals that the brain receives from the outside world. If you flew east from Chicago to Cairo, for example, your circadian clock will be off by 7 hours. When your internal "Chicago" clock is telling you to start getting sleepy around 11 PM, the sun will be rising in Cairo, as the locals are starting to wake up. You may be eating when you're not hungry or laying down in bed when you're not sleepy, and this mismatch contributes to jet lag.

However, with a few days of adjustment, you will be able to overcome jet lag. You will start sleeping as the sun sets, you will get hungry at the same time as the Egyptians, and your physiological measures will start to align with your time zone. This adjustment is only possible because our circadian rhythms are **entrainable**, meaning they are able to change and fit the surroundings. Our circadian rhythms entrain in response to **zeitgebers**, the German word for "time givers": environmental cues, such as increased light exposure when the sun comes up, or social cues, such as increased sensory input from heightened activity of the people around

Figure 12.11 Traveling to a part of the world where day and night misalign with your internal circadian rhythm can cause jet lag.

you. A rise in the neurohormone melatonin is an important signal that contributes to helping the brain entrain, which is why taking a melatonin supplement late at night when in a new time zone may help someone get over jet lag more quickly.

If sunlight is a trigger that helps a person entrain their circadian rhythm to new environments, what would happen if a person is completely isolated from sunlight? In other words, what does a free-running circadian rhythm look like? One of the early documented case studies addressing this curiosity was conducted by a French cave explorer named Michel Siffre. In 1972, Siffre (voluntarily) spent six months deep in a Texan cave to evaluate what would happen to a person completely isolated from zeitgebers. At the end of the experiment, he found his circadian cycle was much longer than 24 hours, and very unpredictable - some of his so-called days would consist of being awake for 36 hours and asleep for 12.

A more rigorous scientific study, conducted in a group of people living in an underground bunker with unchanging lighting conditions for

severeal days estimates the typical free-running circadian cycle to be close to 26 hours. In other words, a person absent from external cues begin to fall asleep and awaken 2 hours later each day.

Free-running circadian disruption is a potential issue for scientists aboard the International Space Station, who experience about 16 sunsets and sunrises per day, since the orbiting space vessel completes one trip around the earth every 90 minutes. In order to minimize the negative effects of jet lag on the researchers aboard, NASA has the inside of the vessel set to an artificial 24-hour cycle. Bright blue LEDs illuminate the cockpit in the "daytime," while dim, red-shifted wavelengths are used in the evening to induce sleepiness.

Figure 12.12 The free-running circadian rhythm is slightly longer than 24 hours. In conditions with unusual light conditions, such as aboard the International Space Station's research laboratory, "night-time" is mimicked every 24 hours.

Circadian rhythms on a molecular level

The circadian cycle functions on the molecular scale at the level of cellular transcription. In the mid 1980s, the gene period was discovered in the genome of the fruit fly, Drosophila melanogaster. Normal fruit flies are no exception to 24-hour cycles of behavior, as they generally exhibit periods of wakefulness that are paralleled by the rising and setting of the sun. But, if this period gene was mutated, there was an unusual change in the sleeping habits of the flies. Some of them slept on a 29-hour cycle, some had a shorter rhythm at 19 hours, and some had no predictable sleep-wake pattern whatsoever.

Later, it was discovered that another gene was related to the cycle of sleep and wake, called timeless. This gene codes for a protein called TIM, which interacts closely with the protein coded by period, the protein PER. When PER and TIM interact with each other, they form a **dimer** (a pair of molecules) with the ability to enter into the nucleus, bind to a specific sequence on the genome, and prevent further transcription of both PER and TIM. Therefore, the paired proteins function as a negative feedback regulatory system.

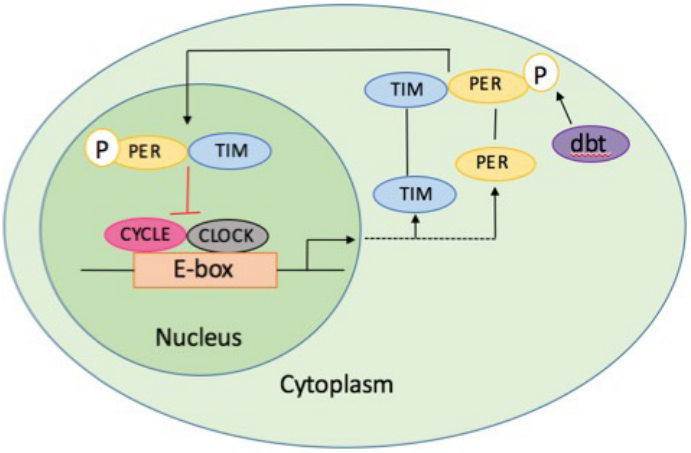


Figure 12.13 The molecular basis for the circadian cycle depends on transcriptional repressors.

The TIM protein, however, is degraded by light, so during the daytime, the concentration of TIM in the cell is very low. As a result, the PER protein is left by itself, where it can no longer repress transcription. The cells, without the

active repression of transcription, then proceed to create more protein. Once night falls, light no longer breaks down TIM, and TIM begins to accumulate again, forming the dimer with PER, which prevents further protein transcription. The cycle repeats itself the next morning.

Since behavior can be driven or modified based on protein levels, the capacity for an organism to change transcriptional activity on a 24-hour cycle suggests that genetic level changes can possibly influence the activity of the whole organism. These gene transcription-level changes were discovered in Drosophila in 1990, and three scientists, Hall, Rosbash, and Young, were recently awarded the 2017 Nobel Prize in Physiology or Medicine.

12.4 Neurochemical signals

Many of the neurotransmitters that the brain uses for signaling are capable of modifying some aspects of sleep. For example:

- Glutamatergic signaling is heightened during the awake state, and many glutamatergic neurons increase their activity during REM sleep.
- Drugs that increase the action of GABA by acting as positive allosteric modulators are used as sedatives and sleep aids.
- Norepinephrine, acting through increasing activity of the sympathetic nervous system, enhances alertness.

Be aware that many neurotransmitters can affect sleep behavior. However, we are only going to describe the function and actions of three sleep-related neurochemicals.

Adenosine

Adenosine is a molecule that has a variety of functions in the body. In addition to being one of the four main building blocks of DNA (the "A" of the A:T G:C base pairing combinations), it acts as a signaling molecule in the body that is involved in inflammation, the immune response, and modulation of heart rate.

It is also used as part of the molecule that stores cellular energy: ATP, or adenosine triphosphate. Each molecule of ATP has phosphate bonds that release tremendous energy when the bonds get hydrolyzed (broken apart). Vesicles, in addition to containing neurotransmitters, have many molecules of ATP. Throughout the day, as the body uses up cellular energy, there is an increase of adenosine as a result. Therefore, as our energy consumption increases, so do adenosine levels, thus signaling to the brain that we are sleepy.

If you are able to chemically block the action of adenosine, you can stave off sleepiness, increasing alertness and ability to focus. Odds are good that you have used a psychostimulant to block adenosine signaling this morning, or are drinking some right now. **Caffeine**, for example, is an adenosine receptor antagonist, and is the world's most popular unregulated psychostimulant drug. Other common adenosine receptor antagonists include theobromine and theophylline, both of which can be found in tea and chocolate. They are chemically very similar to caffeine and adenine, part of the chemical structure of adenosine.

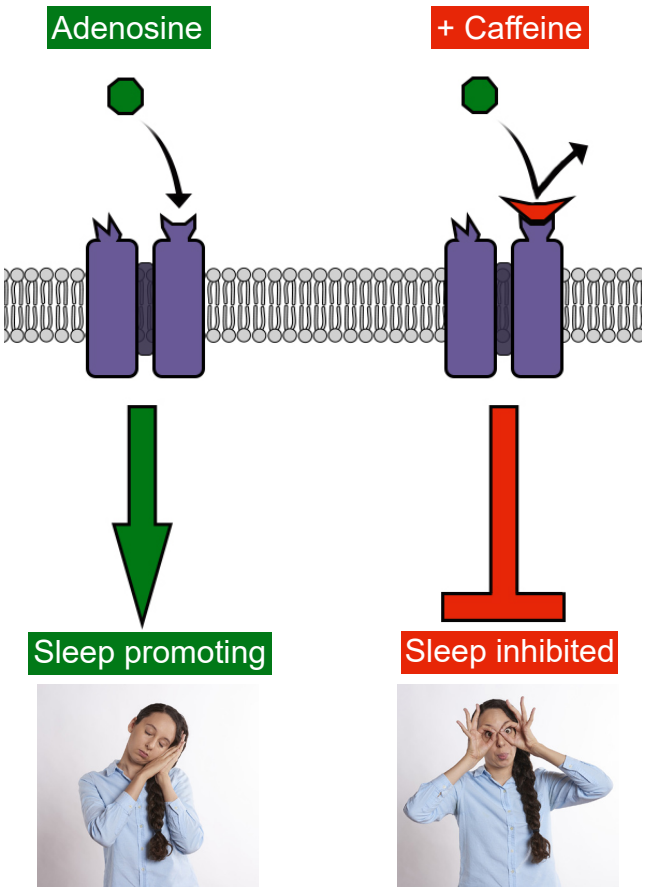


Figure 12.14 Caffeine acts as an antagonist for adenosine signaling.

Melatonin

Melatonin is an endogenous hormone that helps the brain regulate the sleep-wake cycle. Melatonin is produced by a single gland in the brain called the **pineal gland**, so named for its pinecone-like shape. Specialized cells in the pineal gland convert the amino acid tryptophan into melatonin, which is then secreted into the bloodstream. Increased melatonin levels helps to signal the body to prepare for sleep.

The production of melatonin is heavily dependent on exposure to sunlight. While some of the cells in the retina are responsible for passing specific visual information into the brain, other cells (such as the photosensitive retinal ganglion cells) communicate whether or not it is daytime. These cells send their axonal projections separately from the optic nerve that sends most visual information. Rather, they project through a pathway called the **retinohypothalamic tract (RHT)**, synapsing on a clump of cells in the hypothalamus called the **suprachiasmatic nucleus (SCN)**. In turn, these cells of the SCN send inhibitory projections to the pineal gland. In the day time, the SCN tonically inhibits activity of the pineal gland, resulting in low production of

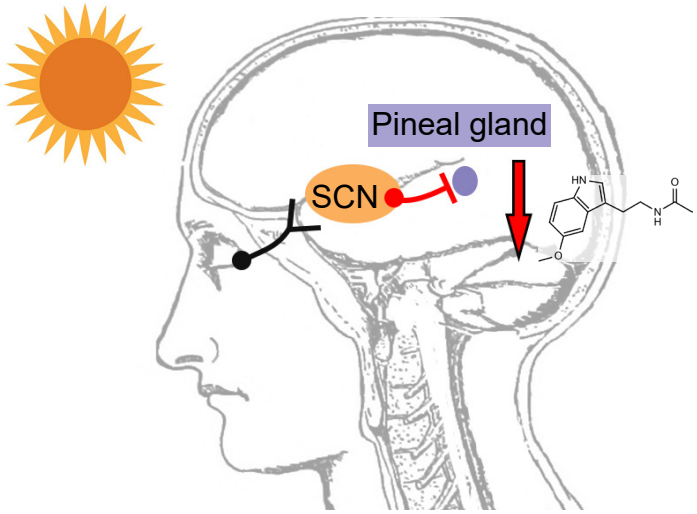


Figure 12.15 Exposure to light sends information via the retinohypothalamic tract and signals to the suprachiasmatic nucleus of the hypothalamus, which inhibits pineal gland production of melatonin.

Half-life

Exogenous substances that enter the body usually get degraded over time through natural enzymatic processes. The **half-life**, also written as $\mathbf{t_{1/2}}$, is the time that it takes for the concentration of the substance to be degraded to one half of what it originally was.

The half-life of caffeine is about 5 hours. This implies that if you drink a full cup of coffee at 2 PM and plan on going to bed at midnight, the caffeine in your bloodstream is similar as if you just chugged a fourth of a cup of coffee before trying to sleep.

The half-life was originally used to describe the decay of radioactive atoms, but has also been adopted by biologists and pharmacologists. It is safe to say that after 5 half-lives, the substance has effectively been eliminated from a clinically relevant perspective.

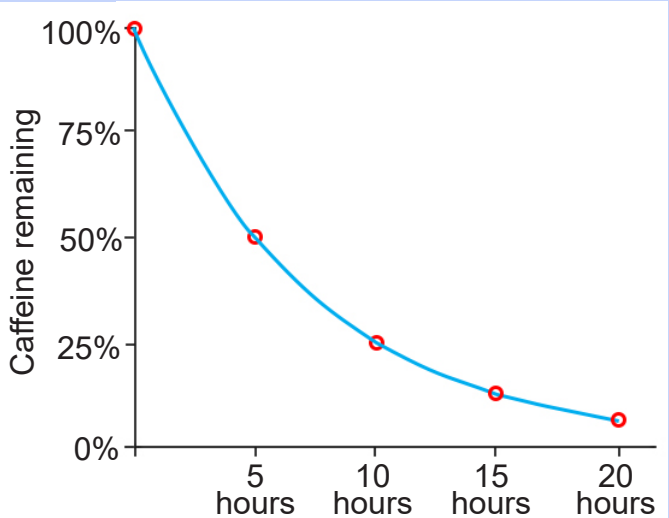


Figure 12.16 Exponential decay of caffeine given a half life of 5 hours.

melatonin. But when daylight starts to decrease, the RHT sends a weaker excitatory signal onto the SCN, which allows increased pineal gland activity.

Light exposure is the main environmental influence that decreases melatonin levels. But, not all wavelengths of light are equally potent at dampening melatonin production. The shorter wavelengths of light, down in the violet-blue range, are much more efficient at activating the RHT compared to longer, yellow-red wavelengths. Increased RHT activation leads to decreased melatonin production, which can delay the onset of sleep. Fluorescent or LED lighting and digital devices, such as computer screens and cell phone screens, use blue wavelengths of light. Therefore, the best advice to optimize your sleep habits is to eliminate exposure to all digital devices about one hour before your intended bed time.

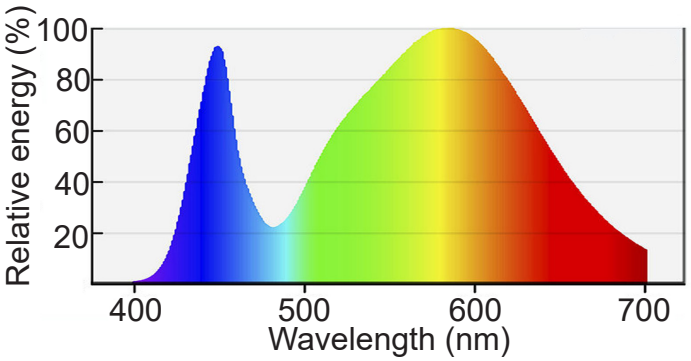


Figure 12.17 Production of melatonin is most strongly suppressed by blue wavelenghts of light, which are emitted by the LEDs that backlight digital devices like computer or cellphone screens.

Histamine

Histamine is a small signaling molecule that has a variety of functions. In the body, histamine mediates the sensation of itch, participates in the inflammatory response, and activates the immune system. In the brain, histamine is a neurotransmitter that acts as a pro-wakefulness signal (the opposite of adenosine and melatonin, which both increase drowsiness.)

Most people understand the role of histamine in the context of allergies. Seasonal allergy sufferers often take a histamine antagonist (**antihistamine**) to decrease the severity of allergen exposure. Many antihistamines warn against operating heavy machinery while taking these drugs, since drowsiness is one of the major side effects. Newer generations of antihistamines are more effective at minimizing drowsiness, so they often advertise “non-drowsy” on the packaging.

12.5 Brain structures involved in sleep

When scientists use an EEG to measure electrical activity, they look at the activity of neurons in the cortex, the outermost layer of brain cells. However, sleep and wake behaviors are driven by the action of cells that lie buried deep within the phylogenetically older areas of the brain. The signals that originate here communicate broadly throughout the rest of the brain, and these signals are the ones that cause us to sleep or wake.

It is not just a single part of the brain that controls sleep behavior, but most likely a network of communication activity between different areas. Here, we will only address a few of the brain areas that are heavily involved with sleep.

Hypothalamus

Early studies on the role of the hypothalamus in sleep began with Viennese neurologist Constantin von Economo. He described a series of patients with a disease called **encephalitis lethargica** in 1917. In his observations, patients presented with one of two sets of symptoms. Some had progressive lethargy, starting with drowsiness, moving to extended sleeping periods, worsening to coma. On the complete opposite end of the spectrum, some patients had severe insomnia - a clinically significant difficulty with falling asleep.

When von Economo performed the autopsy on the patients, he found very specific injuries in the hypothalamus that correlated with symptoms. Among those with persistent sleepiness, the posterior hypothalamus was severely damaged. Von Economo concluded therefore that the posterior hypothalamus contained structures that are needed for maintenance of wakefulness in the

Clinical connection: Encephalitis lethargica

In the midst of World War I, a strange disease of unknown origin called encephalitis lethargica ravaged the globe. A worldwide pandemic, an estimated five million people were affected. About half of the patients died from early stage symptoms, and those who survived often failed to recover fully.

Encephalitis lethargica was also called “sleeping sickness” because of the symptoms that patients experienced: a lack of energy, extreme muscle weakness, and loss of all desires. Neurologist Oliver Sacks described these patients as being “insubstantial as ghosts, and as passive as zombies.”

Today, cases of encephalitis lethargica are extremely rare. The cause of the disease was never figured out, but one theory suggests that malfunctions in the immune system are related to the onset of the disease, since a major outbreak appeared in the wake of the Spanish flu pandemic of 1918.

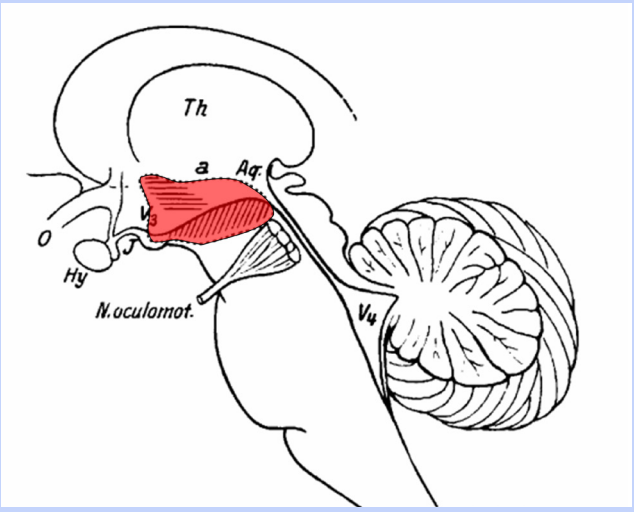


Figure 12.18 Hypothalamus damage was found in people who experienced encephalitis lethargica (sleeping sickness).

healthy individual. In the patients with insomnia as their main symptom, their anterior hypothalamus was injured, leading von Economo to conclude that this area was important for promoting sleep.

Von Economo's findings represented a shift in the way scientists thought of sleep. Most researchers believed that sleep was brought on simply by an overall decrease in brain activity. However, his discovery of hypothalamic localization of a "sleep center" demonstrated that for normal sleep to happen, certain areas in the anterior hypothalamus actually need to increase their activity.

The hypothalamus can be subdivided further into populations of neurons that have previously been addressed in some context. The suprachiasmatic nucleus (SCN), the neurons that receive light information via the retinohypothalamic tract and help regulate melatonin release, is part of the hypothalamus. Elsewhere in the hypothalamus is the **tuberomammillary nucleus**, the major site of neuronal production of the wakefulness signal histamine. The lateral hypothalamus has neurons that produce the pro-wakefulness signaling molecule **orexin** (sometimes also called **hypocretin**), and these neurons are lost in people with severe narcolepsy.

Reticular formation

The reticular formation is found in the brainstem. Like a large net of interconnected clumps of neurons, it is difficult to anatomically classify the structures of the reticular formation, since they do not have a clearly defined border or boundary. Activity of these neurons contribute to a variety of behavioral states, such as alertness and consciousness. The reticular formation is vulnerable to ischemia, lesions, or physical trauma. Often, severe injuries may result in a loss of consciousness or coma.

Information flow through the reticular formation passes in the upwards (towards the cortex) and downwards (towards the body) directions. The upward pathway, also called the **ascending reticular activating system (ARAS),** receives inputs from all the sensory systems before sending wide projections all across the cortex. (For your curiosity, the downward pathway is the reticulospinal tract that is involved with motor control of the skeletal muscles, and is not strongly involved with sleep behaviors.)

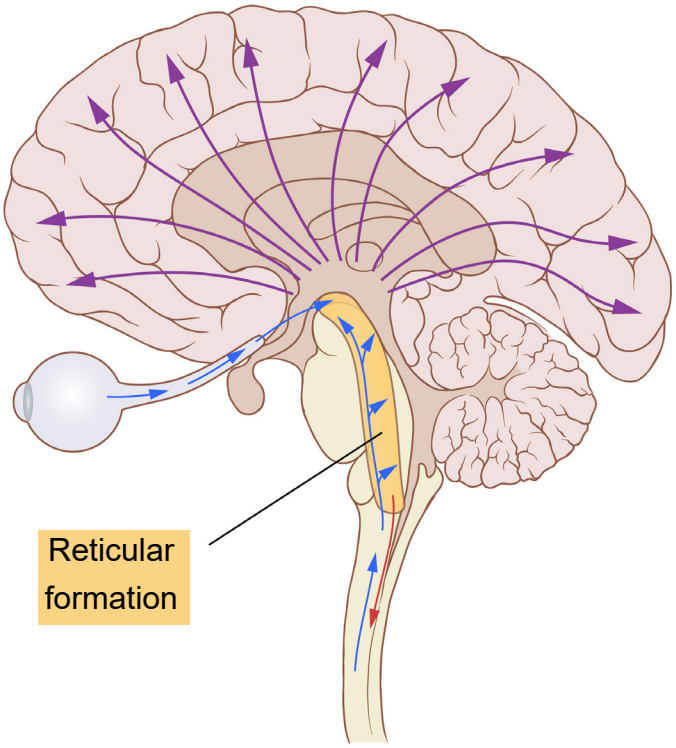


Figure 12.19 The reticular formation in the brain stem is a series of interconnected neurons that control some functions of consciousness.

12.6 Sleep disorders

Insomnia

Almost everyone has experienced difficulty sleeping at some point in their lives, often as a result of stress or anxiety. For example, it might be difficult to fall asleep the night before a big interview, or you may wake periodically in the hours before an important early morning flight.

There is no strict definition for insomnia. The major clinical symptoms are self-reported measures, such as a dissatisfaction with nightly sleep or a change in daytime behavior, such as sleepiness, difficulty concentrating, or altered mood states. The lack of clear diagnostic criteria makes estimating prevalence difficult, but some guesses put the number of people with insomnia close to one third of the US population.

Common triggers for insomnia include heightened anxiety, stress, or advanced age. It may also be downstream of other diseases, such as Parkinson's disease, diabetes, depression, or chronic pain conditions. Lifestyle can also be a major risk factor for insomnia, as jet lag and working late-night shifts can disrupt sleep patterns.

We can describe insomnia as acting at two stages. **Onset insomnia** is defined as a difficulty with initially falling asleep. People with onset insomnia will frequently lie in bed for a long time before finally drifting off. **Maintenance insomnia**, however, is a difficulty with remaining asleep. People with maintenance insomnia experience many waking events throughout the middle of the night, or they may wake up very early in the morning and be unable to get back to sleep. The two are not exclusive, and people may experience both forms of insomnia in a single night.

The most effective treatments for insomnia begin at the level of behavioral changes. Improving sleep habits, such as minimizing arousal states before bedtime, developing a reliable pattern of sleep-wake timing, eliminating caffeine intake in the afternoon and evening, and increasing daytime physical activity can decrease insomnia. Prescription medications are less preferred for insomnia treatment, since these drugs are more effective at inducing unconsciousness rather

Clinical connection: Fatal familial insomnia

While many cases of sleeplessness last a day or two, and some cases are clinically significant and treatable with behavioral changes, a very small fraction of cases of insomnia are incurable and deadly. In people with **fatal familial insomnia (FFI)**, they experience severe insomnia. Some patients stay awake for up to six months at a time. As a result of either the disease or the sleep deprivation, they experience altered mood states, hallucinations, dementia, and eventually death, usually within two years after a diagnosis is made.

The cause of FFI is unknown. There is a strong genetic component associated with it, as it appears frequently within certain family trees. But, there have also been a few cases of sporadic FFI, people with no apparent family members with the disease. One observation in common among people with FFI is significant damage to the thalamus, as a result of misshapen proteins called **prions**, a similar disease-causing agent that is responsible for mad cow disease.

than biological sleep. These drugs can also have adverse psychological side effects such as mood swings and depression, and those adverse effects may be more severe than insomnia itself. Prolonged use of prescription sleep medications can lead to a "rebound effect," causing a person to experience even worse insomnia when they are unable to get sleep drugs. This is called **iatrogenic insomnia**, and can lead to a cycle of dependence.

Sleep apnea

Sleep apnea is characterized by nightly sleep that is frequently interrupted by the inability to breathe (a- meaning none, and pneu referring to wind, air, or breath, as in pneumonia or pneumatic pump). In turn, this lowers blood oxygen levels, causing the brain to wake the person in panic. Each wake event may only last a few seconds, but these disruptions are significant enough to prevent a person from getting the appropriate amount of restorative deep sleep. A person with sleep apnea may wake 30 times an hour, despite having no recollection of waking. The immediate result of sleep apnea is excessive daytime sleepiness, but over long periods of time, sleep apnea contributes to the development of heart disease and an increased risk for stroke.

There are two forms of sleep apnea, both of which may be seen in a person with this sleep disorder. **Obstructive sleep apnea** is the more common of the two. This happens when soft tissue in the back of the throat temporarily collapses, which can decrease or completely block airflow into the lungs. On the other hand, **central sleep apnea** is mediated by some biological change in the brain that results in a decrease in involuntary breathing patterns at night, possibly due to some damage in the respiratory centers of the brain.

There are several risk factors that contribute to sleep apnea, which are often a combination of genetic and environmental influences. Obesity is a major risk factor, resulting in increased soft tissue mass around the neck and torso, which can increase the likelihood of airway blockage. Advanced age contributes to sleep apnea, as the muscles that keep the airway open weaken and lose tone over time. Exposure to chemical irritants, such as cigarette smoke, contribute to

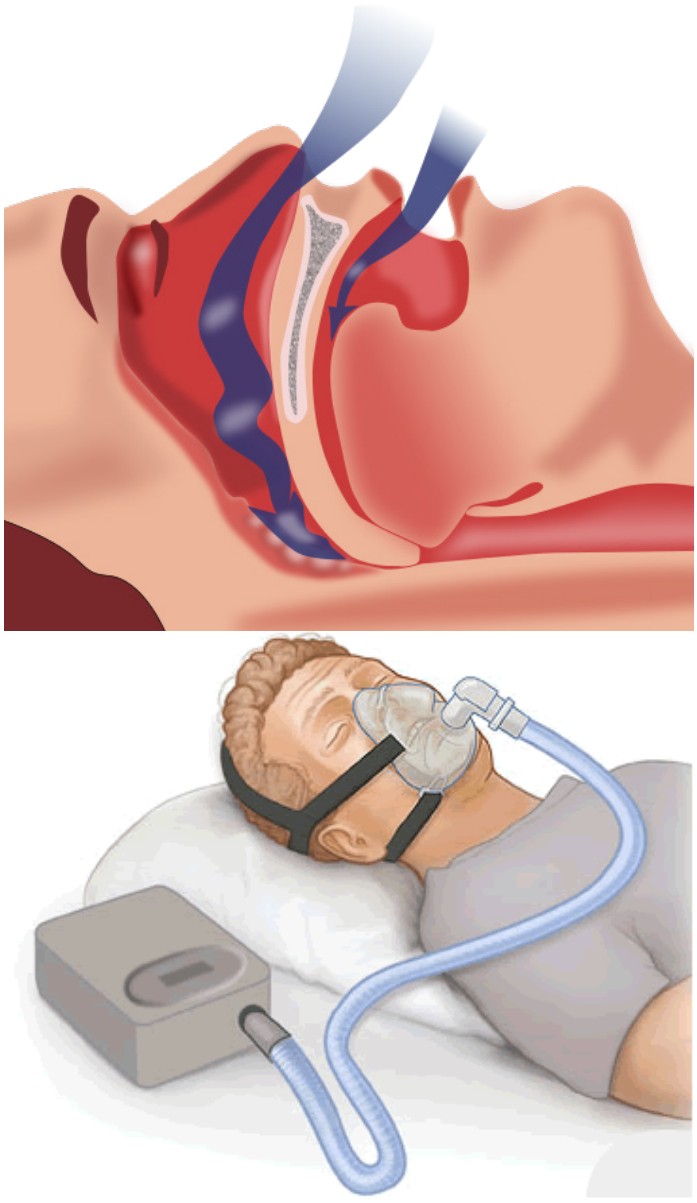


Figure 12.20 Sleep apnea is a blockage of the airway during sleep, causing patients to wake in the middle of the night (top). It can be treated by wearing a CPAP device (bottom).

inflammation and increased water retention in the soft tissue, both of which can decrease the size of the airway.

Sleep apnea is most often treated with a portable machine called a **continuous positive airway pressure** device, or a CPAP device. These machines are basically air pumps that connect to a mask that is worn over the nose and mouth. When used correctly, the CPAP forcibly pushes air into the person's respiratory system, acting as an external lung. However, the CPAP can be loud and bulky, and the mask must be airtight for the treatment to be effective, making the treatment very uncomfortable. Because of this, CPAPs can cause more difficulty with sleep compared to sleep apnea itself, so they often have a low compliance rate.

Narcolepsy

Unlike the previous two sleep disorders, which result in a deficit of sleep, narcolepsy can be thought of as an "excess" of sleep. More accurately, narcolepsy is inappropriate sleep, and it manifests as frequent sleep attacks throughout the day, each event lasting for seconds or minutes at a time. An estimated 1 in 2000 people experience narcolepsy.

One of the life-threatening symptoms that appears in narcolepsy is **cataplexy**, which is the sudden weakening of muscle tone that accompanies a sleep attack. A cataplectic attack may cause someone to physically fall over during a narcoleptic incident. Cataplexy often happens during high emotional states, such as excitement.

As with other sleep disorders, changes in lifestyle can improve the course of narcolepsy. Introducing short daily naps can be helpful, as can general good sleep habits (minimal digital device usage before sleep, regular sleep-wake timing, and physical activity). Drugs such as amphetamines (Modafinil) can be used in the daytime to stimulate activity in the CNS, and can be prescribed to treat severe cases of narcolepsy. Some antidepressant drugs can be used to treat cataplexy.

The exact cause of narcolepsy has not yet been identified. However, there are many clues that point to a dysregulation of the signaling molecule **orexin** produced by cells in the lateral hypothalamus. These neurons die off in people with narcolepsy, but the cause of why the neurons die is unknown. Also, having a genetic

Clinical connection: REM sleep behavior disorder

One very rare parasomnia (sleep disorder) called REM sleep behavior disorder (RBD) can cause people to carry out complex, highly coordinated motor actions while they are sleeping, sometimes acting out their dreams as if they were reality. People with RBD are at risk of injuring themselves or others.Their sleep actions may be in response to a violent nightmare, causing them to jump out of bed, kick or punch the air, run through the house, or throw things.

One of the most shocking instances of parasomnia-induced sleepwalking was the 1987 case of Kenneth James Parks, a Canadian man who, in his sleep, drove 23 km to the house of his in-laws and stabbed both of them with a kitchen knife. After his arrest, scientists discovered that his brain activity was highly abnormal during sleep. As a result of the medical examination, the Supreme Court of Canada acquitted him of murder in 1992.

predisposition to narcolepsy does not guarantee that a person will experience the symptoms, indicating that there is some combination of genetic and environmental factors that lead to narcolepsy.

Restless legs syndrome

A person with **restless legs syndrome** (RLS) experiences frequent unusual sensations in their limbs, such as a tingling or buzzing. Because these uncomfortable sensations disappear with movement, people with RLS often want to move their legs around. Technically, RLS is not exclusively a sleep disorder, since patients experience similar symptoms when they are simply at rest, while sitting and watching TV or studying. But, the sensation happens frequently as a person is lying down in bed, thus delaying the onset of sleep.

RLS is likely underdiagnosed, since it is a disease that exists on a spectrum. Those who are minimally affected probably do not experience significant changes in sleep. Estimates of prevalence of RLS range from 2.5% to 15%. Major risk factors for RLS are iron deficiency and dopamine dysregulation. The exact pathogenesis is unknown, but it does have some genetic component.

Periodic limb movement disorder

Periodic limb movement disorder (PLMD) is a motor disorder that causes abrupt limb movement such as kicking, flexing, or jerking. Usually, the limb movement is in the legs, but arm movement can also be seen. Since motor activity is actively suppressed during REM sleep, limb movement is most often seen in the first half of sleep when NREM sleep dominates the sleep cycle. Each series of limb activity repeats at about a 30 second interval.

PLMD is different from RLS, since PLMD results in involuntary movements that a person may not be aware of, while RLS related-movement is voluntary. As a movement disorder, the course of PLMD can be modified by dopaminergic drugs. Alternatively, sedatives can help minimize nighttime movement.

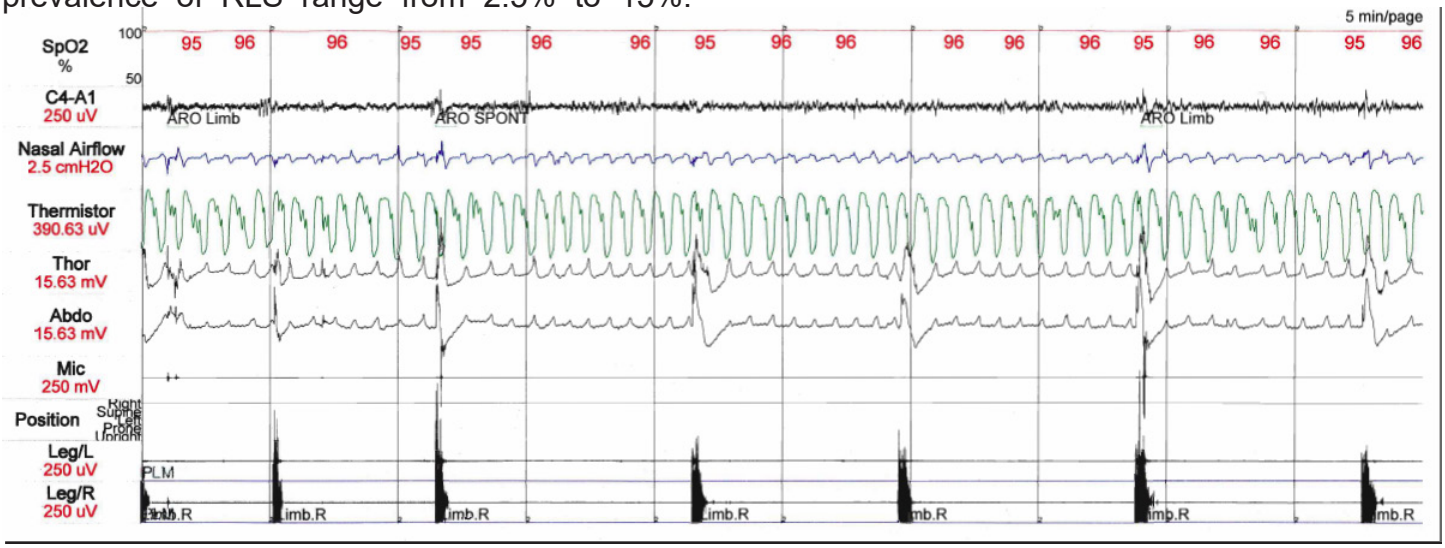


Figure 12.21 Polysomnogram showing repeated muscle activity in both left and right legs (bottom two traces) during sleep.

The Open Neuroscience Initiative is funded by a grant from the Vincentian Endowment Fund of DePaul University.

Cover: https://pixabay.com/photos/cat-sleep-wall-blue-grey-leisure-4358985/
12.1 Eugene Delacroix, lady Macbeth sleepwalking
12.2 https://commons.wikimedia.org/wiki/File:EEG_recording.jpg
12.5 modified from https://commons.wikimedia.org/wiki/File:Video_schematic_of_glymphatic_flow.ogv
12.7 https://pixabay.com/photos/dolphins-fish-mammals-delphinidae-380034/
12.8 https://pixabay.com/photos/tiger-big-cat-majestic-noble-zoo-3383616/
12.9 https://pixabay.com/photos/father-baby-portrait-infant-22194/
12.10 Brucewinter [CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0)]
Pancrat [CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0)]
12.11 Image by Arek Socha from Pixabay
12.12 https://commons.wikimedia.org/wiki/File:ISS-47_The_International_Space_Station%27s_Destiny_Laboratory_at_%27night%27.jpg
12.13 PranjalAgr [CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0)]
12.14 https://pixabay.com/photos/asleep-sleep-bored-tired-nap-3126444/ https://pixabay.com/photos/crazy-fun-person-happy-girl-funny-2840240/
12.15 Davide Gnocchi, Matteo Pedrelli, Eva Hurt-Camejo, and Paolo Parini [CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)]
12.17 By Lamiot - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=36046024
Kipala [CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0)]
12.18 Images are generated by Life Science Databases(LSDB). [CC BY-SA 2.1 jp (https://creativecommons.org/licenses/by-sa/2.1/jp/deed.en)]
12.19 https://commons.wikimedia.org/wiki/File:Reticular_formation.svg modified by Austin Lim
12.20 https://commons.wikimedia.org/wiki/File:Obstruction_ventilation_apn%C3%A9e_sommeil.svg https://commons.wikimedia.org/wiki/File:CPAP.png?wprov=srpw1_26
12.21 https://commons.wikimedia.org/wiki/File:PLMD_Polysomnogram.png

Chapter 13:

Learning and Memory

Austin Lim, PhD (DePaul University)
David Graykowski (DePaul University)
Editor: Alexandrina Guran, PhD (Universität Wien)

Think back to your favorite birthday party. Which of your friends were there? What did you do, where did you go, and did you have cake? Did you get gifts?

The ability to perform this task depends on our ability to create and recall memories. According to our current best understanding of the neuroscience of learning, the underlying biology of a memory mainly consists of subtle changes among synapses distributed across several brain areas. Our ability to learn new facts, recount the events of last week, or to perform new motor skills is the result of learning-induced neural plasticity. In this chapter, we will consider different aspects of learning and memory, starting from the behavioral level down to the molecular changes responsible for memory formation, as well as some disorders that disrupt healthy memory processes.

Chapter 13 outline

13.1 Patient HM

One of the most influential case studies in the neuroscience of memory is the story of **Patient HM**. HM was born in 1926 in a small Connecticut town. He had a mostly regular childhood: taking family road trips, riding bicycles, and learning about American presidents in school.

In his childhood, HM began having severe seizures, possibly the result of a head injury. In his teenage years, he started having **tonic-clonic seizures**, the most severe form of seizures that produces a loss of consciousness and convulsions (extreme muscle contraction or extension). In his early adulthood, he was having a tonic-clonic seizure monthly and several minor seizures daily, preventing him from working a normal job or living a normal life - despite taking a cocktail of anti-epileptic medications.

Neurosurgeon William Scoville proposed a “frankly experimental operation” to treat HM. It was known that most epilepsy originates in patches of neurons of the **medial temporal lobe (MTL),** and HM’s epilepsy was typical in this respect. Scoville suggested to surgically resect the MTL. In 1953, Scoville removed about

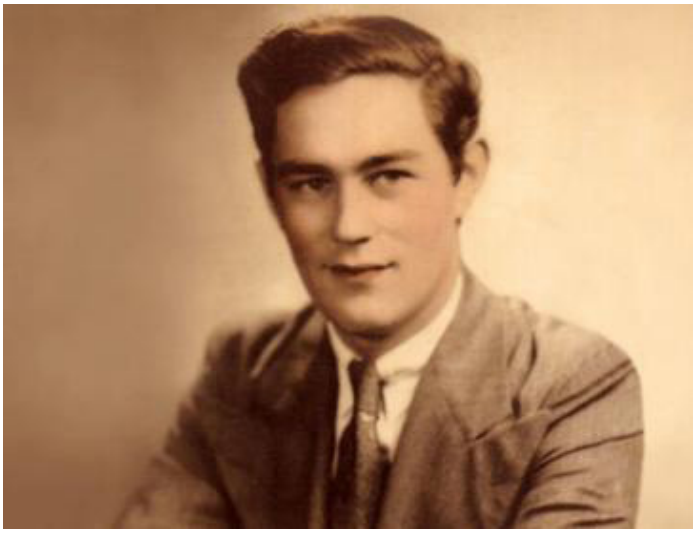


Figure 13.1 Patient HM at 27 years old.

8 cm of the MTL bilaterally, including part of the amygdala, and notably the **hippocampus**, the seahorse-shaped structure of the brain.

The surgery succeeded at its primary goal: HM's seizures were less frequent and less severe. However, HM was left with a highly unusual and life-altering side effect: He was unable to create new discrete memories, a memory deficit called **anterograde amnesia**. For example, he could not remember what he had eaten for lunch just minutes after finishing the last bite. Despite being an avid fan of watching the news, HM couldn't remember the names or the faces of different celebrities or public figures. It was as if he was permanently living in the present. (In contrast, **retrograde amnesia** affects the ability to successfully retrieve memory from one's past.)

However, despite his pervasive memory deficits, HM did not display any deficits in intelligence. His language and speech were unaffected, and word recall was excellent, as he loved completing crossword puzzles and often did so successfully late in life, with only the occasional spelling errors. He could learn to acquire new skills, such as keeping a pen still on a moving circular platform, or a tapping task (these skills

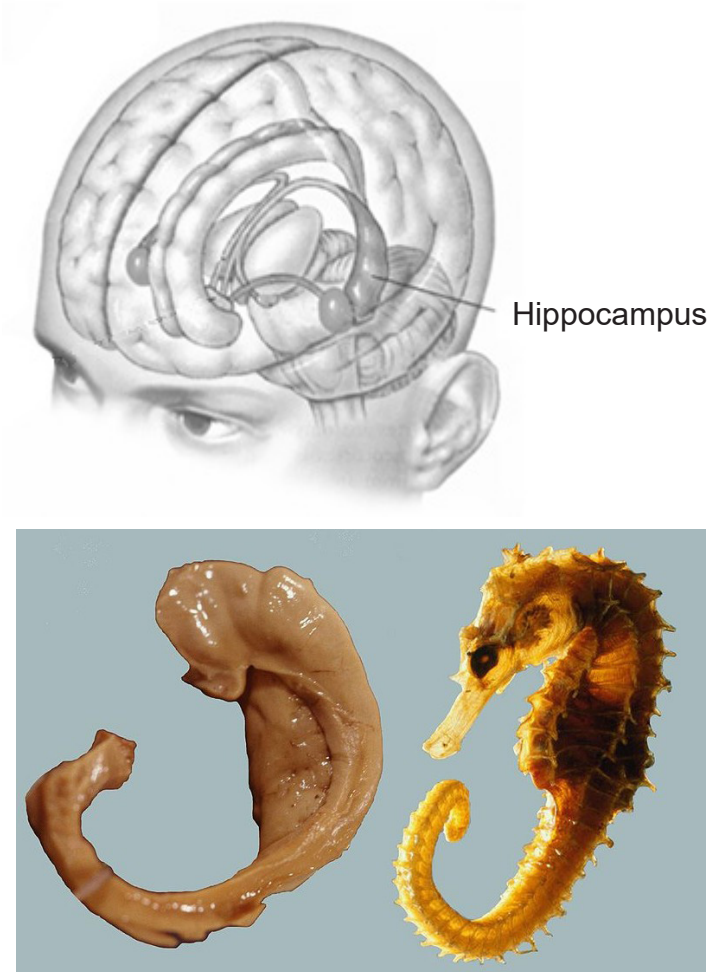


Figure 13.2 The location of the hippocampus in the medial temporal lobe (top). A dissected hippocampus and fornix (bottom left) looks like a seahorse (bottom right).

are different form of memory called **procedural memory**; see below). He was also capable of recalling things from his early childhood, such as geography facts he had learned in elementary school.

Types of memories

The fact that HM's MTL surgery disrupted some types of memories (e.g., memory for facts) while others were still intact (e.g., motor skills) inspired neuropsychiatrists to try to define the different forms of memory. Much of the research was led by Dr. Brenda Milner, who carried out several behavioral tests on HM to figure out what

types of memories are dependent on the intact MTL and which ones can function without MTL.

The most profound deficit was HM's inability to create new **declarative memories**. Declarative memories, also called **explicit memories**, are the pieces of information that can be consciously declared or stated explicitly. Declarative memories are thought of as a "knowing what". Declarative memories can be further subdivided into semantic memory and episodic memory.

Semantic memories are pieces of factual information. Some examples include:

1. "Jupiter is the largest planet of our solar system."
2. "Rosalind Franklin discovered the double-helix structure of a DNA molecule."
3. "The actor Keanu Reeves played the protagonist of the movie The Matrix."

An **episodic memory**, sometimes also called an **autobiographical memory**, is the recollection of a discrete moment in a person's life. It can be thought of as "mental-time travel" - what was it like when. The following memories are examples of episodic memories:

- "When I got home, I put my wallet and phone on the table."
- "I ordered pizza last night."
- "In 2019, I went to see my favorite musicians perform live."

Several tests concluded that HM had lost his ability to create new semantic memories. In one such study, HM was asked to determine if a word was made up or real. He was shown words with very old origins, such as "shepherd" or "butcher." On these words, he performed as well as the control group. When he was shown words that are made up, such as "phlage" or "thweise", he likewise performed as well as the controls. However, when shown words that were added into the dictionary after his 1953-surgery, such as "granola" or "jacuzzi," he scored about 50% correct - consistent with guessing at random, as if he never acquired the knowledge that these words have a meaning.

HM was also unable to create autobiographical memories. When asked to recall one of his birthday celebrations as an adult, he wouldn't be able to give any significant details about the event. Instead, his answers were often vague and generic.

One interesting observation was that HM's memory about details from his childhood were still intact. The inability to recall memories from the past, in this case, from before HM´s surgery, is called **retrograde amnesia**. Patient HM's retrograde amnesia was **temporally graded**, meaning that the farther back you examine, the more complete his memories were. Many of his memories for the two years before his

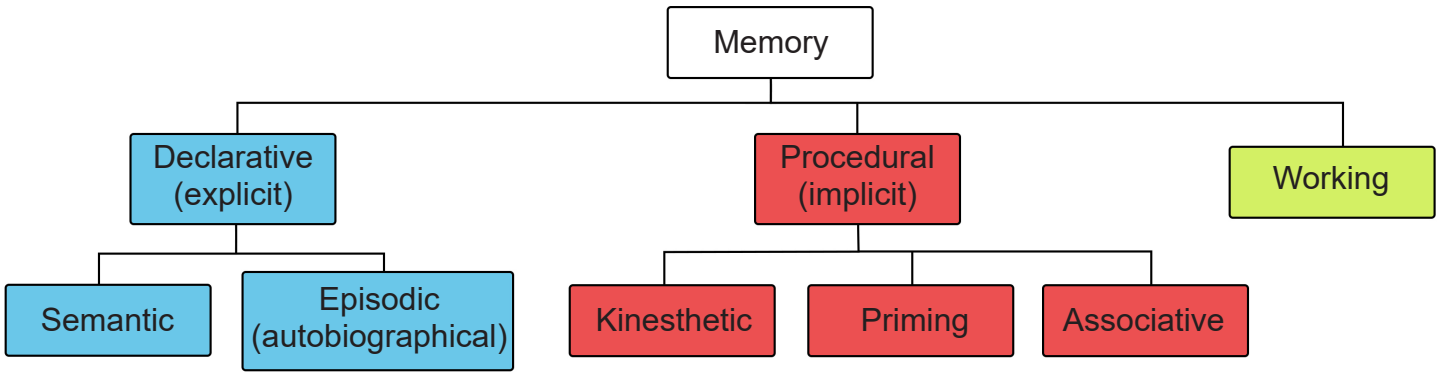


Figure 13.3 Summary diagram of some of the major subtypes of memory.

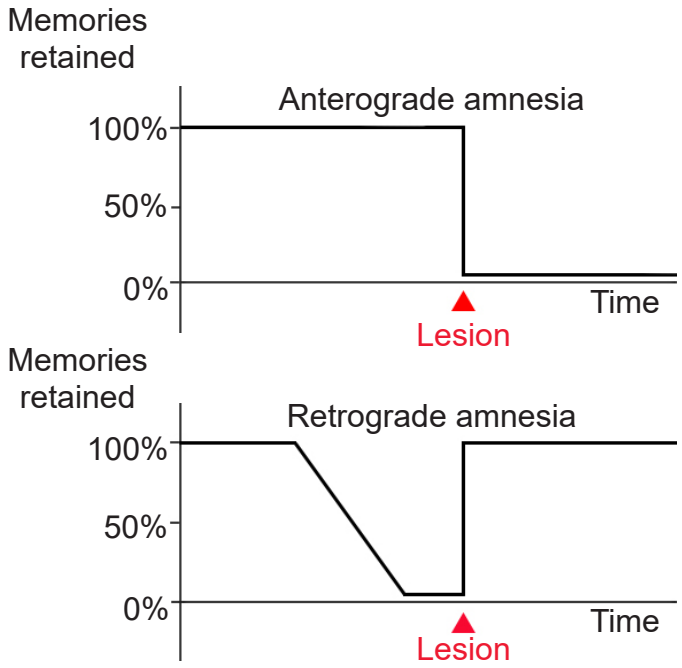


Figure 13.4 In anterograde amnesia, a person is unable to create new memories following a lesion (top). In temporally-graded retrograde amnesia, older memories are better retained while recent memories are more likely lost.

surgery were completely lost, but memories from his youth and teenage years were intact as much as healthy individuals (there is contention about this observation, because HM was taking several anti-epileptic drugs, which may have impacted memory formation.) From this observation, memory researchers concluded that the MTL functions as short-term storage site for memories, but after some years, those memories get relocated to other brain areas outside of the MTL. Currently, the scientific evidence suggests that memories are distributed across several networks of cortical and subcortical brain areas.

While HM lost the ability to create new declarative memories, he was still able to maintain a different class of memories, called **procedural memories** (or **implicit memories**). They are unconscious memories, and can't be explicitly stated. These can be thought of as "knowing how". Some examples of procedural memories include, for example, performance of a series of motor actions without conscious thought such as an experienced musician playing a simple scale (sometimes commonly called "muscle memory", even though the muscles do not store any actual memory!), or a priming effect (such as when a person sees pictures of bananas, they are more likely to answer the fill-in-the-blank prompt "b _ _ _ _ _" with "banana", whereas other people might guess "bubble" or "badger").

The original test of procedural memory conducted by Dr. Brenda Milner was called the **mirror tracing task**. In this test, the patient is told to draw a third star in between the two stars as quickly as possible without making any mistakes. The challenge is that the tracing is to be done while watching their hand and the star in their reflection in a mirror. Because of these unusual circumstances, completing this task is difficult. But over multiple days of practice, people become

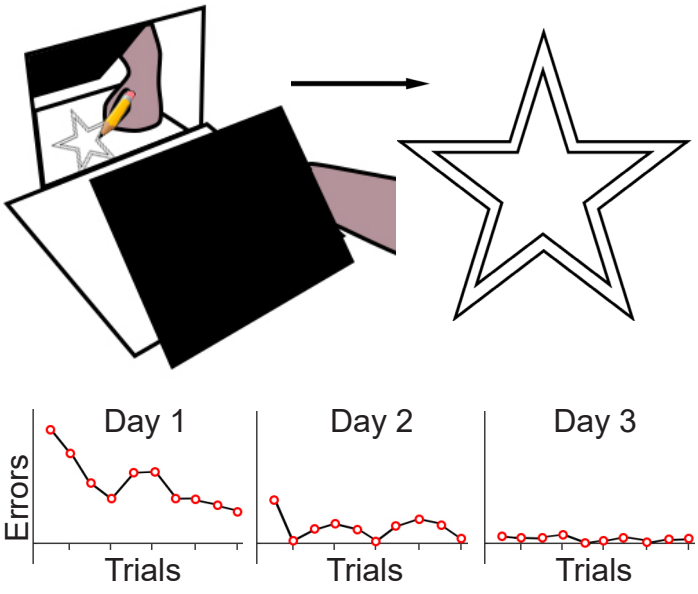


Figure 13.5 Patient HM performed poorly on the mirror tracing task (top), but improved at the task over time despite having no memory of performing the task (bottom).

better at this mirror tracing task, completing it faster with fewer errors. Improvement on this task indicates that a person is learning or gaining some memory about how to better perform the task.

After practicing this mirror tracing task, HM was able to finish drawing the star about ten times faster than when he first began. He improved his performance within each day's worth of training, and he also improved day-to-day. There is evidence that he maintained these skills up to one year later, despite not having regular training on this task. Surprisingly, each day Milner examined HM, she would need to reintroduce herself since he forgot who she was. She also had to re-explain what HM was supposed to do in the mirror tracing task. Hence, while HM was unable to form declarative memory about the experiment or the people involved, learning of the procedural memories and motor actions involved in this task remained intact.

Another type of procedural memory is an **associative memory**. Associative memories are the types of information that we learn through traditional Pavlovian conditioning. For example, recall the classic Nobel prize-winning experiment in physiology conducted by Ivan Pavlov in the late 1800s. Normally, the presentation of dog food, an **unconditioned stimulus (US)**, causes a dog to salivate, a naturally happening behavior, called the **unconditioned response (UR)**. Dogs are not particularly interested in the sound of a whistle: this neutral stimulus will produce a minor response, such as a head turn and attentional shift towards the origin of the sound, but not much more than that. However, when this stimulus is repeatedly paired with the presentation of food, dogs quickly learn to associate that the whistle signals food. After multiple pairings, upon hearing the whistle, a **conditioned stimulus (CS)**, the dogs begin to salivate, a **conditioned response (CR)**, independent of any food being presented.

Separate from declarative or procedural memories, a different form of memory called **working memory** was tested in HM. Working memory involves processes of storing information temporarily while simultaneously manipulating those pieces of information. This type of memory

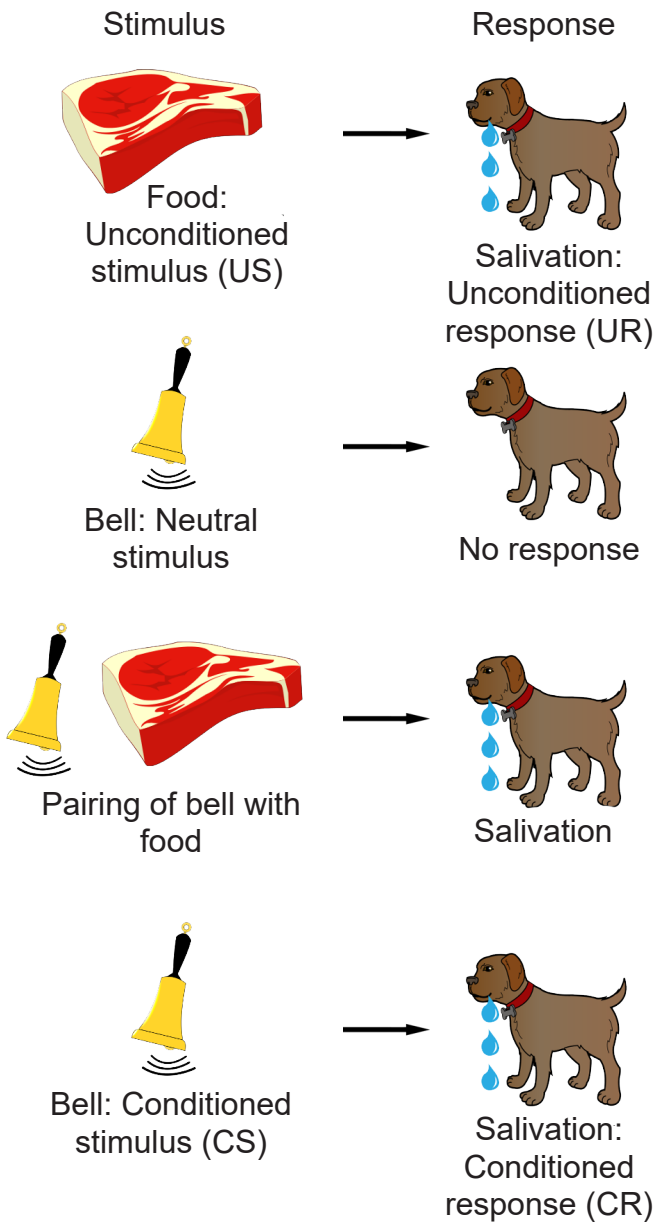


Figure 13.6 A CR after exposure to a CS, such as in classical Pavlovian conditioning, is an example of an associative memory, one type of procedural memory.

can be thought of as a “short-term memory on overdrive.” Although HM struggled with working memory immediately after his surgery, several years later HM performed as well as age-matched control patients on these tasks.

For example, a test of working memory is the **digit span test**, where a person is given a series of numbers to remember, then they are asked to repeat the numbers in reverse order. After successfully completing this task, a different series of numbers, this time one digit longer, is presented to the patient until they first start making errors in recall. A related task is called the **Corsi block tapping test**, where an experimenter sets up several blocks on a table. The experimenter then taps a series of blocks in a specific order, then the subject is asked to tap on the blocks in reverse order. As with the digit-span test, the experimenter then makes the series of blocks longer until they make mistakes in the tapping.

Patient HM died in 2008 at age 82 of respiratory failure. His name was Henry Molaison.

Trial 1:	Prompt	4 8 2 6
Expected response		6 2 8 4
Trial 2:	Prompt	2 8 7 9 1
Expected response		1 9 7 8 2
Trial 3:	Prompt	8 4 1 1 8 6
Expected response		6 8 1 1 4 8

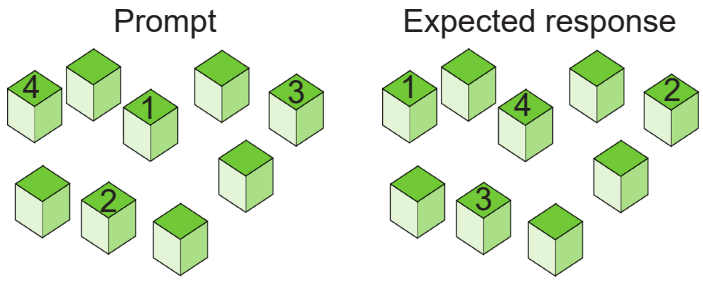


Figure 13.7 The digit span test (top) and the Corsi block tapping test (bottom) are measures of working memory.

13.2 Neural Structures Involved in

The Hippocampus (HPC)

The **hippocampus (HPC),** meaning "seahorse" in Greek, was named based on its morphology. The HPC is located along the ventral and medial surface of the brain. The HPC is one of the critical structures of the **limbic system**, a series of subcortical brain structures that are involved in several different complex behaviors, such as emotions and memory. The limbic system is an evolutionarily ancient brain network.

The synaptic connectivity of the hippocampus is very well characterized. Hippocampal synaptic connectivity was first described by Ramon y Cajal, and is made up of three main synaptic connections; sometimes called the **trisynaptic circuit**. First, the axonal outputs of layers 2 and 3 from the **entorhinal cortex** make up the inputs into the HPC. This white matter signaling tract is called the **perforant pathway**, and they synapse onto the granule cells of the **dentate gyrus**. These neurons send axons, called **mossy fibers**, to the pyramidal cells of the **Cornu ammonis (CA) 3** region of the HPC. The axonal projections from here, called **Schaffer collaterals**, project into **CA1**, which are the neurons that make up the output of the hippocampus. These outputs project out to layer 5 and 6 of entorhinal cortex. While the three main neuronal projections are glutamatergic, the trisynaptic circuit is modulated by GABA, acetylcholine, norepinephrine, and serotonin.

The HPC is involved in **spatial memories**, memories involved in navigation of our surroundings and the creation of a mental map of our world. Spatial memories are developed when we enter a new building for the first time, and we search for a new classroom. We also use our

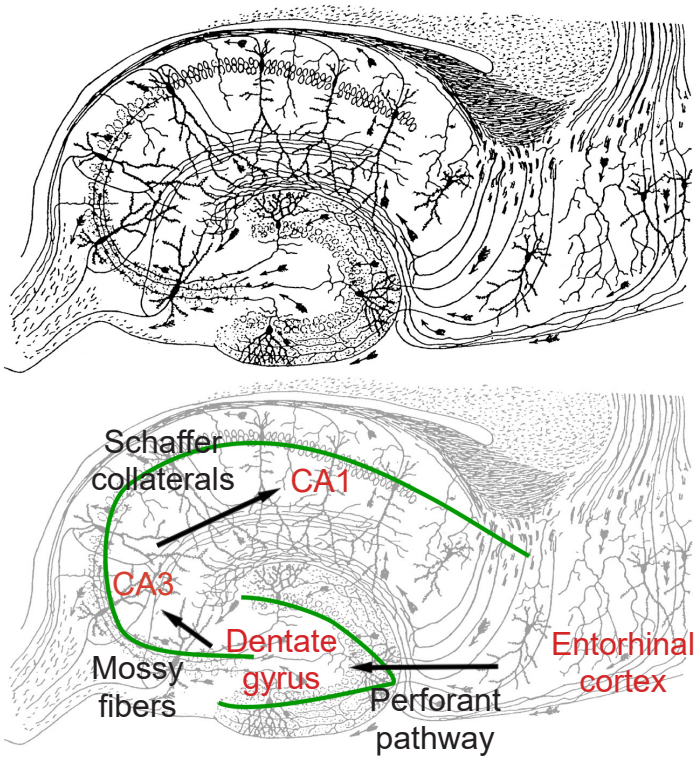


Figure 13.8 The circuitry of the hippocampus as illustrated by Ramon y Cajal (top) and as a schematic diagram (bottom). Structures are labeled in red and communication pathways are in black.

spatial memory whenever we are walking around campus, making our way from one building to another, thinking about the streets you'd need to cross or the buildings you can cut through. While the volume of the hippocampus is not a reliable indicator of the strength of a healthy person's spatial memory, injury to the hippocampus causes deficits in spatial memory.

To test spatial memory behaviorally in non-human animals, one test that is regularly used in rodents is called the **Morris water maze**. In this test, a shallow pool is filled with an opaque liquid, making it difficult to see through. Hidden somewhere in this pool is a clear plexiglass platform, surrounding the pool are different

environmental cues that can be seen from the surface of the water, such as different shapes or colors. The water is deep enough that when a rodent is put into the Morris water maze, they have to swim to stay afloat. The rodents swim around aimlessly until they find the platform, the time it takes for this to happen is recorded, and the trial ends. Over time, the animals learn that the platform is located near certain navigational cues, and on future trials, the animals spend more time near those cues, and the latency to find the platform decreases. When the HPC is surgically removed from rodents or inactivated, they perform poorly in the Morris water maze.

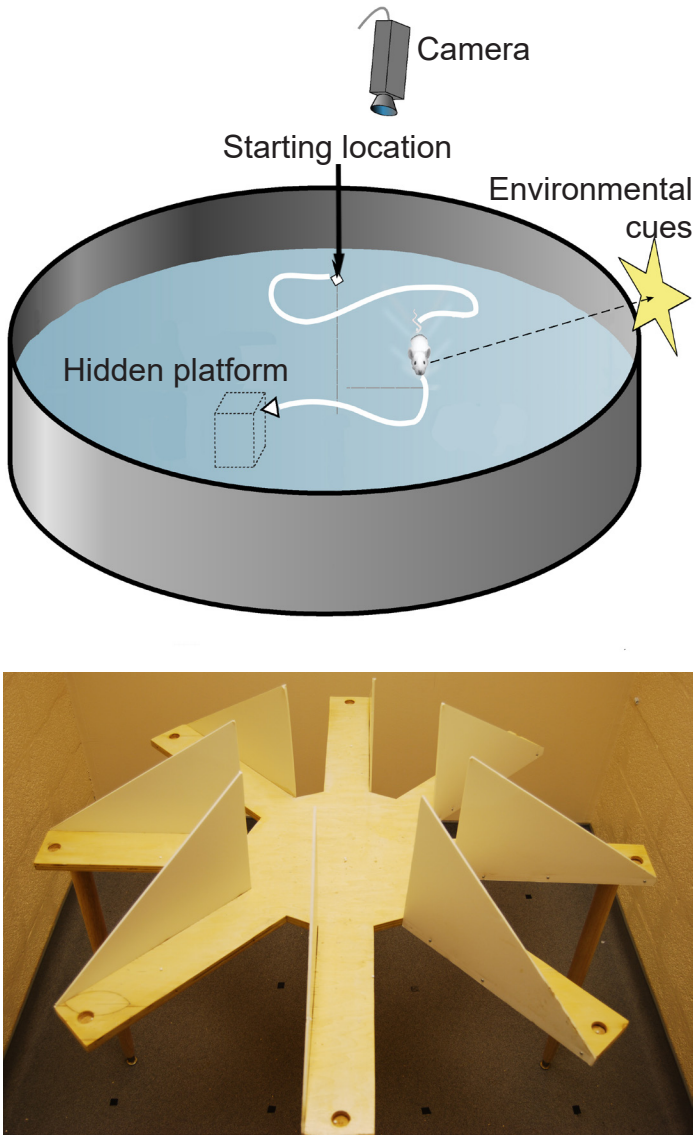


Figure 13.9 The Morris water maze (top) and the radial arm maze (bottom) are behavioral tests to assess spatial memory.

Another non-human behavioral test used to assess the capacity for learning navigational cues is the **radial arm maze**. In this test, a rodent is placed on a circular platform. Extending from this platform are eight or more "arms", at the end of each is a small dish. In one of the dishes is a morsel of food ("rewarded arm"), while the other dishes contain nothing ("non-rewarded arms"). The maze is designed so that the food cannot be seen from the end of each arm, so the animal must return to the starting platform before exploring another arm. The number of entries into a non-rewarded arms is counted as an error. Over time, the animals make fewer errors as they learn which arm is rewarded and which ones are not. Alzheimer's disease model organisms perform poorly on this task.

Based on the deficits seen in Patient HM and other experimental manipulations of the HPC, we conclude that the HPC is strongly implicated in the process of declarative memories and spatial navigation. Since some of HM's memory functions were still intact, such as procedural memories and working memory, it is believed that these functions are independent of HPC function.

The Amygdala

The **amygdala** is another limbic system structure found in the medial temporal lobe adjacent to the HPC. Amygdala comes from the Greek word meaning "almond," which roughly describes its shape. While the amygdala is often spoken of as a single structure, it is more accurately divided into several subnuclei, each with different cell populations and functions. One broad division distinguishes the basolateral amygdala (BLA) versus the central nucleus of

amygdala (CeA): The BLA contributes to both fear memories and reward processing, while the CeA contributes more to the physiological response in emotions as well the perception of emotion.

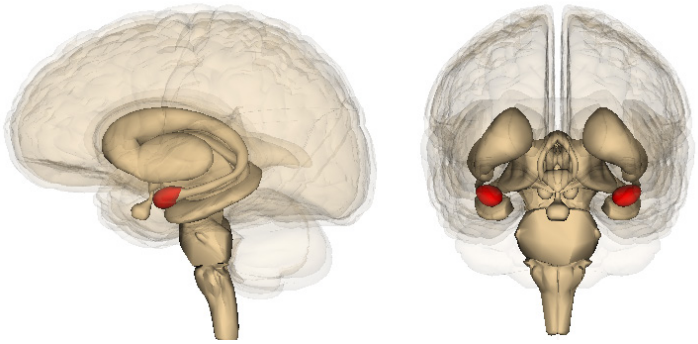


Figure 13.10 The amygdala are temporal lobe structures that contribute to the salience of emotional stimuli.

The amygdala is strongly involved with the formation and storage of emotional memories, memories or associations that have a strong emotional connection. Both positive and negative emotional states are represented here. For example, a whiff of grandmother's cooking may cause you to reminisce back to a fun childhood summer. Alternatively, the smell of vomit may elicit the unpleasant emotions and nausea associated with a nasty food poisoning incident.

One non-human test of emotional memory is the **foot-shock paradigm**, a form of **fear conditioning**. This test involves putting a rodent into a chamber with floors made of metal rods, which are connected to an electric current generator. The metal rods can deliver a non-lethal but painful electric shock to the rodent's foot. In this learning paradigm, a combination of sound and light cues is presented to the animal. Shortly after, the painful foot shock is delivered. If the animal learns that the cues are associated with the negative painful memory, they exhibit freezing after exposure to the cues. Amygdala lesions prevent the animal from freezing, while hippocampal lesions have no effect on this emotional learning. Changing cellular signaling in the amygdala alters the learning of fear conditioning. The foot-shock paradigm is often used as a non-human model of post-traumatic stress disorder.

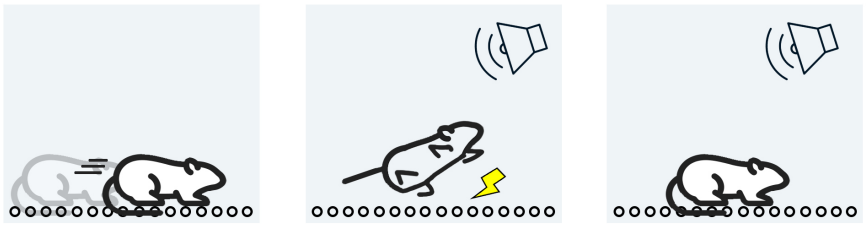


Figure 13.11 In the fear conditioning paradigm, a rodent is put into a room with medal rods as the floor (left). Then, a sound tone is repeatedly paired with a foot shock (middle). When the sound is played again, the rodent may exhibit freezing behavior (right).

Inferotemporal cortex (IT)

Structures of the **inferotemporal cortex (IT)** are part of the ventral stream of visual perception (chapter 7). The IT stores some components of visual memory. We use these functions when we see a classmate outside the classroom and recognize them from our Introduction to Neuroscience class, or when we see a parody of a famous painting and recognize the similarities to the original work. A simple behavioral task to assess visual memory would start by viewing a series of abstract shapes, and when a shape appears that you have already seen, you push on a button. The human capacity for visual memory is massive: After viewing 10,000 images for a few seconds apiece, people were able to identify a previously seen image successfully about 83% of the time.

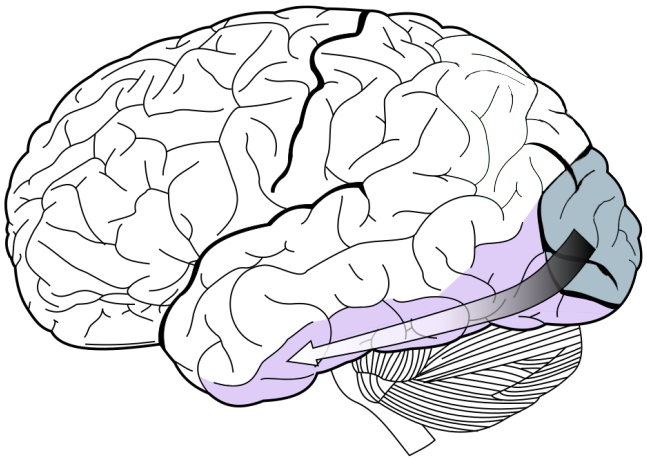


Figure 13.12 The inferotemporal cortex is one of the signaling pathways important for visual memories.

A bit more specifically, one part of the IT is the **fusiform gyrus**, which has been previously described in the context of facial recognition. People with prosopagnosia, a visual perceptual disorder affecting the fusiform gyrus, can perceive the different parts of a person's face, but have a difficult time putting the whole picture together and matching those features to a specific person. For facial recognition to be accurate, there must be some memory that allows for a person to match those facial features with someone they have seen before, which is a memory related process.

The **parahippocampal place area (PPA),** also found in IT, contributes to visual memories associated with locations and environmental scenes. Imaging studies have demonstrated that activity of the PPA increases specifically when people view place-related images, including scenic landscapes like mountains, man-made structures like campus buildings, or the interiors of rooms, both furnished and completely empty. To serve as control stimuli, viewing faces or objects does not increase the activity of the PPA.

Prefrontal Cortex (PFC)

As part of the frontal lobe, the PFC is involved in high order decision making and personality. In the context of memory, neural circuits in PFC are important for short-term and working memory. Patients with injuries to their prefrontal cortex after stroke, tumors or aneurysm, performed worse on a variety of working memory tasks such as the digit span test. Additionally, people with frontotemporal dementia, a neurodegenerative

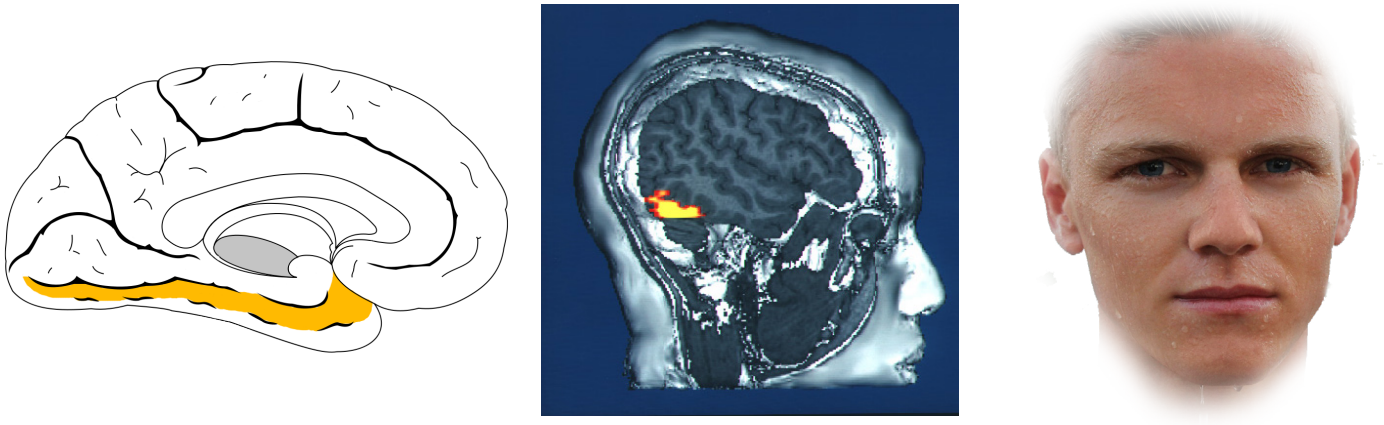


Figure 13.13 The inferotemporal cortex (left, sagittal view) is part of the ventral stream of visual perception, and is likely one site of visual information storage. Within the IT is the fusiform gyrus (middle), which is specifically activated strongly in imaging studies when a person is shown with a facial stimulus (right).

disorder characterized by a degradation of the frontal lobe, often have difficulty with working memory.

The PFC also has strong projections with the hippocampus, and these circuits are likely also involved in the formation of hippocampal-dependent memories.

Striatum

The striatum is a structure of the **basal ganglia**, a series of brain structures that contribute to behaviors such as motor activity (chapter 10) and procedural memories. The striatum likely holds memories involved in habits. Habitual behaviors help us preserve cognitive bandwidth, reducing the "mental energy" that is used during repetitive task performance. The downside of habits is that reliance on habitual responding can limit behavioral flexibility, and cause a person to act in a suboptimal manner, perhaps behaving in a way that led to a positive outcome in a previous set of circumstances without incorporating and evaluating the present circumstances.

Habitual action performance is likely related to a variety of neuropsychiatric disorders. **Obsessive compulsive disorder (OCD),** for example, is characterized by the presence of recurring, intrusive thoughts, which can lead to repetitive actions. Commonly observed is the thought that one's hands are unclean, which leads to repeated handwashing.

A rodent behavioral test of habitual activity is the observation of **self-grooming**, a natural and healthy series of stereotyped actions that consists of licking the paws and moving them through the fur of the nose, caudally down the body. Mouse models of OCD show excessive self-grooming to the point where they pull their fur out and paw their skin to the point of injury.

Drug addiction is also a striatal disorder. Compulsive drug use is often associated with a series of habitual motor actions that happen

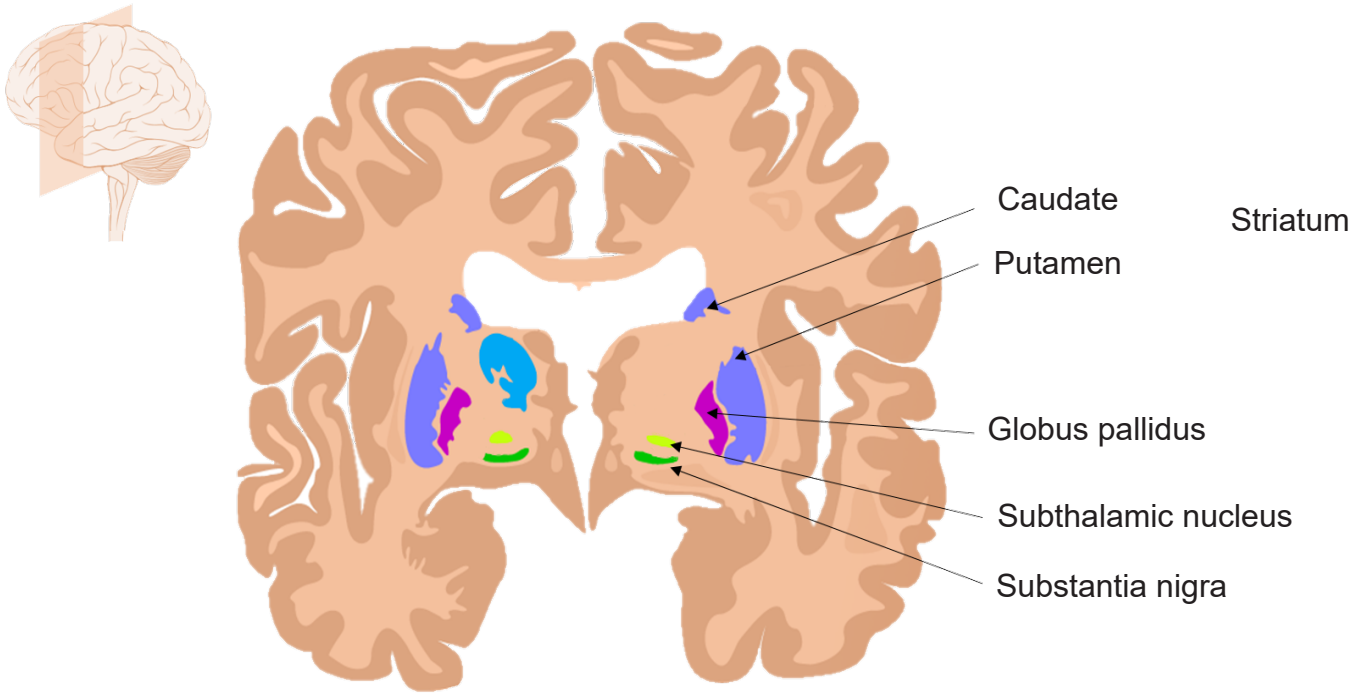


Figure 13.14 Several subcortical brain areas make up the basal ganglia.

before a person experiences the drug effect. For example, in tobacco use disorder, people will perform an orchestrated series of actions, including opening a pack of cigarettes, flicking the lighter, withdrawing the cigarette and taking a deep inhalation. Some of these behaviors are likely stored across striatal circuits (chapter 11).

Cerebellum

The cerebellum is the phylogenetically ancient structure found posterior and ventral to the cerebrum, and functions generally to help with motor functions (chapter 10). The cerebellum is involved in procedural memories, particularly the performance of motor abilities. Learning new motor skills likely requires changes in the circuit strength of cerebellar neurons.

This list of brain structures involved in memory is certainly not exclusive. For example, the orbitofrontal cortex plays a role in positive emotional memories, and sensory cortices are important for the memories related to the specific stimuli that are processed in those areas. A single memory could likely be stored in several brain areas, much like a mosaic.

Hypermnesia

Solomon Shereshevsky was one of a handful of rare, clinically documented cases of hypermnesia, the capacity to recall nearly any memory with perfect precision, even after several years. Remarkably, he could "easily remember any number of words and digits, equally easily he memorizes whole pages from books on any subject and in any language and for a quite long time at that. Shereshevsky can accurately quote anything he was told ten or twelve years ago." He received recognition in 1968 when his psychologist A. R. Luria published a case study in "The Mind of a Mnemonist."

Today, we would describe Shereshevsky as being autistic with strong multimodal synesthesia. Also of note, he had significant deficits in executive function, difficulty with recognizing faces, and could not interpret abstract ideas.

13.3 Cellular Mechanisms of Learning

In the late 1800s, around the time when Golgi and Ramon y Cajal were engaged in intense debate about the organization of the nervous system, many neuroscientists came to a strange observation: the weight of the brain increases dramatically over the first 10 years of life, but not much more after that. Even though we learn lots of new facts and make lots of new memories in adulthood, the brain itself doesn't grow in size. So how is it possible to store new knowledge if the brain is not making many new neurons?

Most likely, new pieces of information are held in the connections between cells, not just in the cells themselves. If our estimate of 150 trillion synapses per adult brain is correct, then it is possible that we could store all the knowledge and memories that we collect over our lifetime through some combination of activity across certain connections.

The activity at the cellular level is believed to take place on at least three different levels enabling us to build, store, and retrieve memories.

1. **Encoding** refers to the ability for brain circuits to store some piece of information. In real life, you are presented with countless stimuli simultaneously. Imagine walking down a busy street, and think of the number of different sights, smells, and tactile stimuli you experience. Storing memories is an energetically costly process, and we are limited in the fact that all of our sensory inputs cannot possibly get encoded. Instead, evolution has preferred to encode stimuli that are most salient pieces of information, such as perceptual cues associated with predators. Alternatively, information that we pay strong attention to can get encoded more strongly, like when we repeat a phone number to ourselves until we have a chance to write it down. It is also easier to encode novel information that "builds" on previous bits of knowledge, or information that is closely related to other well-established information, which is why analogies are such an effective way to learn new facts.

2. The process that enables memory storing is called **consolidation**, which makes the memory more permanent. In 1949, an early neuropsychologist, Donald O. Hebb, offered an explanation for how changes in synapses could possibly lead to a phenomenon as complex as learning. His theory, published in his text The Organization of Behavior, can be summed up in the phrase:

"Cells that fire together, wire together."

In Hebb's framework, repeated activity at a synapse within a circuit of neurons acts as a reinforcer signal that strengthens this synapse for future communication, making the next incoming signal more robust. Hebb also implies the inverse is true: when cells do not fire together, they weaken their connection. Through fine tuning of synaptic connections, some strengthening and others weakening, a lifetime of memories can be stored across a wide distribution of neurons. After a memory has been created, the specific circuit of neurons that represent that piece of information is called a **memory trace** or an **engram**.

A cellular process called **reverberation** is thought to be the mechanism that allows for consolidation. Reverberation is the process by which networks of neurons fire repeatedly. Each

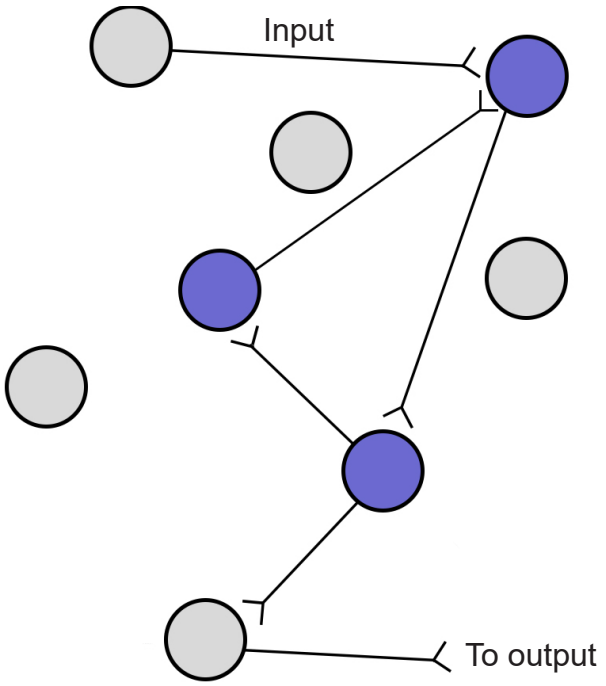


Figure 13.15 A reverberating circuit (purple) is a series of neurons that are activated repeatedly with the activity of a positive feedback circuit.

time that circuit is activated, the strength of the network is increased, meaning that it becomes easier for that circuit to be activated in the future.

Throughout the process of consolidation, memory traces are thought to become more represented in the neocortex and less in the subcortical structures like hippocampus or amygdala. Recall Patient HM's temporally graded retrograde amnesia, where he had lost declarative memories in the two years leading up to this surgery, and yet his memories from the long past were maintained. This finding suggests that some aspects of declarative memory consolidation depends at least partially on medial temporal lobe and HPC for a period of time, maybe up to two years, before those memories get stored in the cortex more permanently. However, we cannot conclude how long consolidation takes in healthy humans based on findings from HM, who had pervasive and frequent seizures, which might have negatively impacted memory consolidation before the surgery.

Consolidation seems to occur predominantly during sleep. Specifically, declarative memory is enhanced during non-REM sleep, while procedural memory is enhanced during REM sleep. Studies investigating sleep consolidation are often done by letting participants learn or perform behavioral tasks, after which they are deprived of a specific phase of sleep. To wake participants at the right moment, researchers commonly use EEG signatures, which are unique for different phases of sleep: sleep deprivation early in the night denies non-REM slow wave sleep, while late-sleep deprivation decreases time spent in REM sleep. Some speculate that the hallucinations we sometimes experience during dreaming are a consequence of consolidation processes, but it is inconclusive as to what role dreaming plays in memory formation.

3. Finally, for the stored memories to be recalled, a cellular process called **retrieval** happens which brings back the specific engram. Retrieval happens for both declarative and procedural memories.

The writing of memories can be differentiated from the retrieval of memories using specific priming-related behavioral tests. One example is a vocabulary recall test. Consider if you were given a list of 50 words to memorize, words that belong to a handful of different conceptual categories (such as cinnamon, pepper, or curry, which fall under the category "food flavors"). When asked to write down as many words as possible from the list using your memory, a test called **free-recall**, you may be able to successfully remember about a third of them. However, if you were prompted with the category titles, a related test called **cued-recall**,

you would perform much better at retrieving those memories, possibly recalling up to 75% of the words. The fact that cued-recall scores are often higher than free-recall scores indicate that there is a distinction between the encoding / consolidation of memories and the retrieval of memories.

Retrieval is not a passive function. When an engram is retrieved, it is **reconsolidated**, which is an act similar to replaying the activity of the circuit. During this reconsolidation, it is possible that some aspects of the memory are emphasized, while others are lost. This is likely the main reason why we experience **false memories**, memories that are not true to reality - one reason why eyewitness testimonies are notoriously unreliable. We may imagine our good memories as better than they actually were, while simultaneously dampening the negative aspects of those memories. A dysregulation of this reconsolidation process could lead to the symptoms seen in post-traumatic stress disorder, where the negative emotional components of a particular memory are exaggerated rather than being blunted.

Special populations of neurons

An individual memory is likely distributed widely across several different parts of the brain. However, there are a few special populations of neurons mainly in the MTL that contribute to highly specific types of memories.

Place cells

Place cells are a special population of pyramidal cells of the hippocampus. These neurons increase their firing activity when the animal is in a particular location in an environment, indicating that they contribute strongly to location and navigational memory. There is no apparent topographical arrangement of these place cells, meaning that adjacent areas of an environment do not necessarily activate adjacent hippocampal place cells. The place cells, when firing at the right times, help the animal create a spatial map of their surroundings.

Grid cells

Grid cells are located in the entorhinal cortex, the main input structure to the HPC. Closely related to the place cells described above, grid cells increase their firing properties

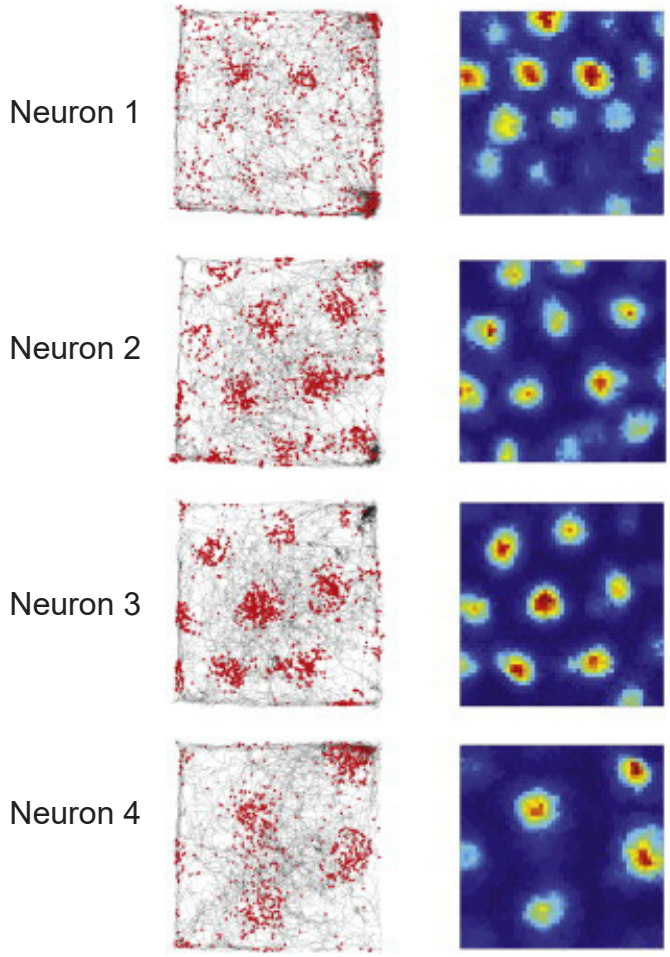


Figure 13.16 A rat is tracked (left, gray) as they move through an open field. Individual grid neurons spike when the rat passes through particular areas of the open field (left, red). Heat map (right) showing high neuronal activity in warm colors (red and yellow) with low activity in cool colors (blue).

periodically when an animal is at an intersection of a "grid" in a wide-open, previously-explored environment. The grid itself is roughly hexagonal, and spans the whole environment an animal is in. The overlap of multiple grids gives the animal an idea of the surroundings. The scientific description of grid cells earned three scientists a Nobel Prize in Physiology or Medicine in 2014.

Jennifer Aniston neurons

The "Jennifer Aniston" neurons, also called **concept cells**, are a series of cortical neurons in the temporal lobe that increase their firing exclusively in response to highly-specific stimuli, such as the idea of Jennifer Aniston, Halle Berry, or the Tower of Pisa. These concept cells respond to much more than just pictures: for example, a Luke Skywalker neuron that responded to a picture of young Mark Hamill (the actor who played Luke Skywalker) will also respond to text that reads "LUKE SKYWALKER" and the sound of a person saying "Luke Skywalker". Although this particular neuron probably won't fire in response to pictures of athlete Manu Ginobili or actress Marilyn Monroe, the neuron might fire in response to pictures of Yoda or Darth Vader, indicating that the neuron may encode an even broader concept, such as "Star Wars characters" or "Jedi", or is a part of a network that encodes concepts related to the Star Wars franchise.

Figure 13.17 Concept cells change their firing pattern in response to the presentation of highly specific stimuli, such as the character Luke Skywalker (portrayed by actor Mark Hamill; images 1, 3, and 5). Visually similar but conceptually different stimuli (male brunette actors appearing in film), like pictures of Leonardo diCaprio (image 2) or Keanu Reeves (image 4), fail to induce changes in firing. However, visually distinct but conceptually-related stimuli, like the picture of Yoda (a related character from the same series of films; image 6) may also drive the concept cells to fire.

13.4 Molecular Mechanisms of Learning

Zooming in beyond the level of anatomy, the substrates of learning can be found at the level of synapses. Synapses change in a phenomenon called **plasticity**. The word "plasticity" refers to a change in synaptic strength, which may be an increase or a decrease. This change may persist for minutes, hours, days, or in some cases, even a whole lifetime. When synaptic strength is increased and remains elevated, we call this **long-term potentiation (LTP)**. A prolonged weakness of a synapse is called **long-term depression (LTD)**. In our current limited understanding of plasticity, both phenomena are important for a healthy brain, and neither one is always good or always bad.

It is also important to clarify that both excitatory synapses and inhibitory synapses can be subject to either LTP or LTD.

Long-term potentiation

In 1973, Bliss and Lomo were the first to publish evidence of plasticity using electrophysiology. The experiment began at the hippocampal connections of an anesthetized rabbit. They put a stimulating electrode among the axons of the perforant pathway, at the entrance to the hippocampus. A second electrode, capable of detecting electrical charges in brain tissue, was placed among the cells of the dentate gyrus, the area where those axons release neurotransmitter. By stimulating the perforant pathway, Bliss and Lomo could record how neurons of the dentate gyrus respond. A single pulse caused the neurons to depolarize, a measurable observation called a **field excitatory post-synaptic potential (fEPSP).** The more neurons that depolarize, the larger the fEPSP would be.

Instead of simply giving a single electrical stimulation, however, Bliss and Lomo were interested in testing Hebb's theory about plasticity. If "cells that fire together, wire together," then perhaps they could experimentally drive those cells to fire in a pattern that would induce a rewiring of the connections, resulting in LTP. The duo delivered a very intense electrical stimulation, zapping the axons at 100 stimulations a second (100 Hz) for 3 seconds. This **high frequency stimulation (HFS)** led to an enhancement of the amplitude of the fEPSP in response to a single stimulus - this demonstrated that LTP was a measurable phenomenon. In a different experimental setup, this LTP was shown to persist up to one year later! In humans, we theorize that some synaptic connections may remain potentiated for our entire lifetime, however investigating this is in humans is ethically constrained.

Long-lasting changes in synaptic strength, such as the LTP that Bliss and Lomo demonstrated, are made possible through a series of molecular and cellular level changes. One form of LTP results from a change in the types of glutamatergic receptors. Of the three classes of ionotropic glutamate receptors, two are important for this form of LTP: the **AMPA** and the **NMDA** receptors. The AMPA receptors are the glutamate receptors that we generally imagine as contributing to excitation (more information in section 5.4). When a molecule of glutamate binds to the active site of this receptor, the ligand-gated ion channel changes and allows cations, mostly Na^+, to cross the cell membrane, leading to depolarization.

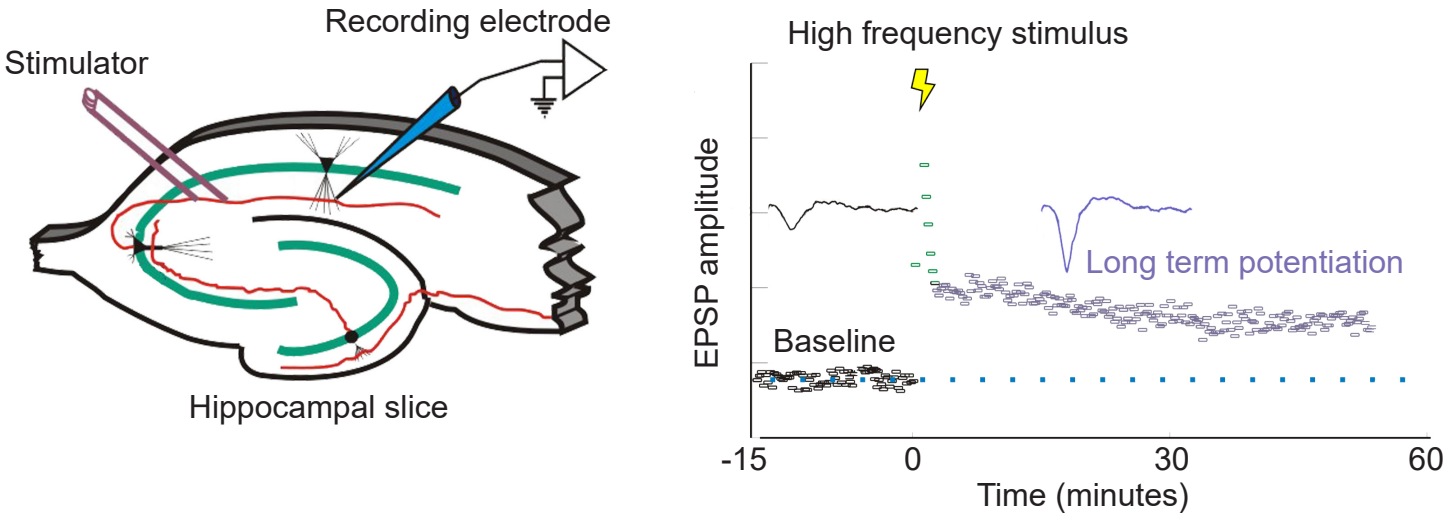


Figure 13.18 Schematic of the recording configuration of Bliss and Lomo's experiments (left) demonstrating that high frequency activation of the Schaffer collaterals while recording the field EPSP in the CA1 region leads to long term potentiation (right).

The NMDA receptors are somewhat more complex. NMDA receptors are like the ionotropic AMPA receptors because they are permeable to cations and therefore excitatory, but they have a few specific functional differences. For one, their molecular pore has space for a magnesium ion (Mg^{2+}) to sit in the middle of the ion channel. Mg^{2+}, like the other ions that we have discussed (chapter 4.2), responds similarly to the forces of the electrochemical gradient. Mg^{2+} is more highly concentrated outside the cell compared to the inside, and it has two positive charges, so these ions are drawn to the interior of the cell. But, the pore of the NMDA receptor is not large enough to allow the bulky Mg^{2+} ion to actually cross into or out of the cell membrane. Instead, it stays stuck inside the ion channel. Mg^{2+} physically takes up so much space that it occludes the movement of other ions across the cell membrane, basically blocking passage of ions through the NMDA receptor.

The other relevant feature of the NMDA receptor is that it is permeable to the Ca^{2+} ion. Increases in intracellular Ca^{2+} postsynaptically is the crucial trigger that leads to the cellular expression of LTP. Ca^{2+} ions activate an enzyme called **calcium / calmodulin-dependent protein kinase II (CaMKII)**. CaMKII itself has many molecular targets. As a kinase, it's main molecular action is to phosphorylate proteins. When CaMKII becomes activated, it phosphorylates amino acid residues on the AMPA receptor, which enhances their current passing properties, thereby increasing their response with glutamate present. Secondly, CaMKII also contributes to cellular mechanisms which result in increased trafficking of AMPA receptors to the cell surface. Thirdly, CaMKII interacts with the transcription factor **cAMP response element-binding protein (CREB)**, which can then move into the nucleus and instruct the nucleus to synthesize more of the mRNA that leads to increased synthesis of the AMPA receptors. Taken together, increases in intracellular Ca^{2+} postsynaptically leads to an enhancement of a signal that persists over the time course of hours: The definition of LTP.

But, these NMDA receptors are not activated by glutamate alone. Because of the

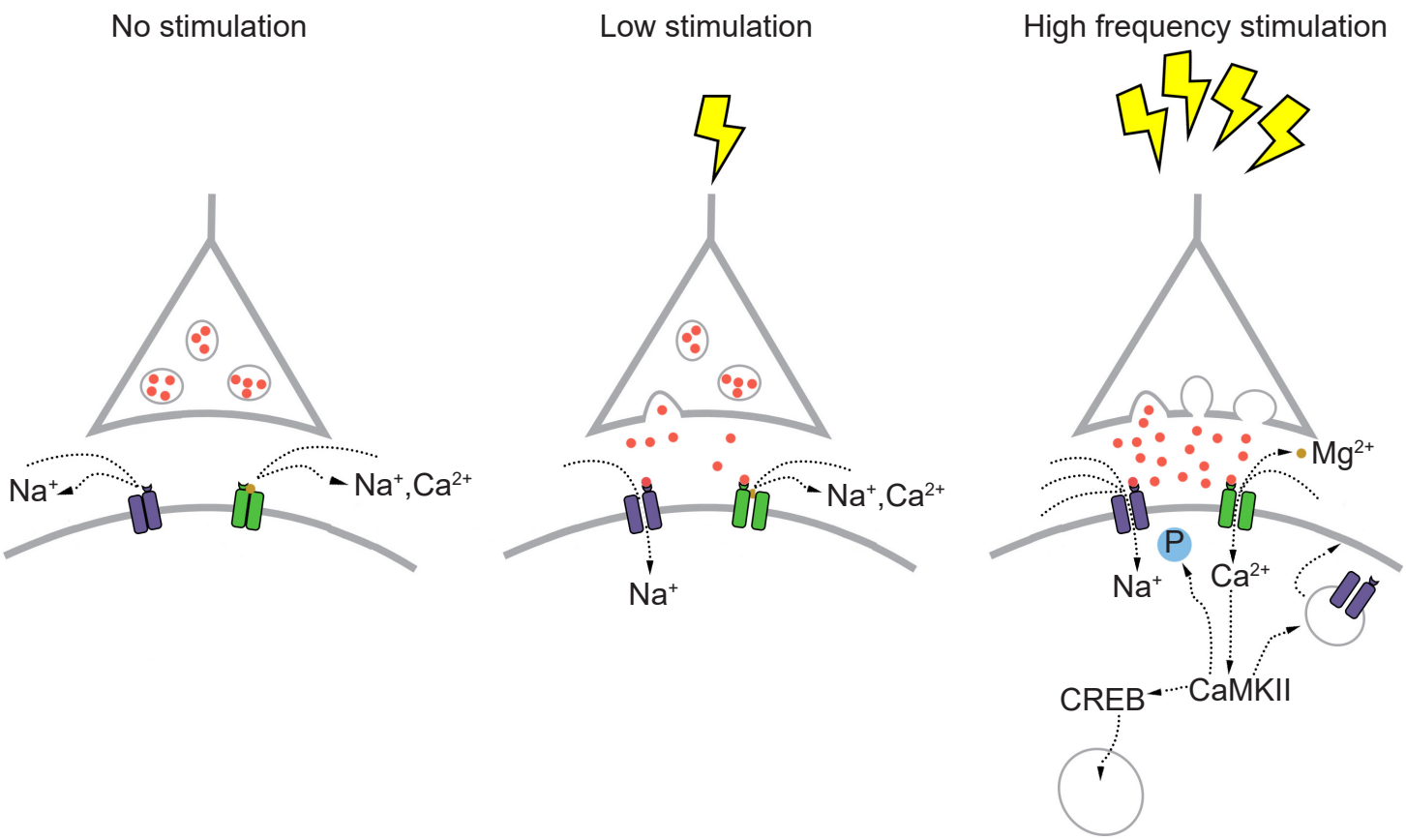


Figure 13.19 Molecular mechanisms explaining postsynaptic LTP. At no stimulation, glutamate (pink) does not strongly activate the AMPA receptors (purple; left). A single presynaptic depolarization causes some glutamate to be released, which activates AMPA receptors, causing postsynaptic depolarization (middle). At high frequency stimulation, significant glutamate release activates AMPA receptors, strongly depolarizing the postsynaptic cell, which causes the Mg^{2+} ion to leave from the NMDA receptor (green). Ca^{2+} enters through the NMDA receptor, and can trigger long term changes in the molecular components of the neuron (right).

molecular properties of the NMDA receptors, they need two conditions to be fulfilled before these receptors get activated and Ca^{2+} moves into the cell membrane:

1. A ligand like glutamate must activate the receptor. As with other receptors, there is no activation in the absence of an agonist.

2. The postsynaptic cell must also be depolarized. When the cell is at positive potentials, the electrical gradient causes the bulky Mg^{2+} ion that is stuck in the pore to be repelled by the cell's interior, which then frees the ion channel for movement of cations across the cell membrane.

Because both conditions must be met before Ca^{2+} can trigger plasticity through CaMKII activation, the NMDA receptor can be described as a **coincidence detector**. These properties can help explain why LTP was only observed after Bliss and Lomo activated the hippocampal slices robustly with their high frequency stimulus paradigm. Strong activation depolarized the axonal fibers, which caused a significant amount of glutamate to be released, activating many of the postsynaptic AMPA receptors. This strong

activation caused the postsynaptic neurons to depolarize, which expels the Mg^{2+} ion out of the NMDA receptor. At this stage, both conditions are fulfilled, and Ca^{2+} enters into the postsynaptic cell, which activates CaMKII, triggering LTP.

Some glutamatergic connections between neurons contain only NMDA receptors but no AMPA receptors. Because these postsynaptic cells do not depolarize in response to glutamate release, and no current passes through the NMDA receptor due to the Mg^{2+} block, these synapses do not change their activity even with glutamate release. These synapses are called **silent synapses**. As we grow, the number of silent synapses decreases, another aspect of brain development.

Long-term depression

Around the same time LTP was being characterized in the rabbit hippocampus, its cellular opposite, **long-term depression (LTD)**, was also being demonstrated in a different experimental preparation. Through the 60s, psychiatrist Eric Kandel and his colleagues worked with the marine mollusk ***Aplysia californica***. With a nervous system of only 20,000 cells, Aplysia is orders of magnitude simpler than the other model organisms used at the time. Additionally, some Aplysia neurons are huge, up to a millimeter in diameter, which took away the need for highly precise equipment.

Aplysia also has a relatively simple anatomy. It breathes using a half-circle of delicate tissue called the gill, which is guarded by the mantle shelf. They also have an organ called the siphon, a small tube that is used for moving water through the animal. Kandel and his colleagues began their exploration of memory by studying the **gill-withdrawal reflex**, a defensive motor response behavior. When a stimulus,

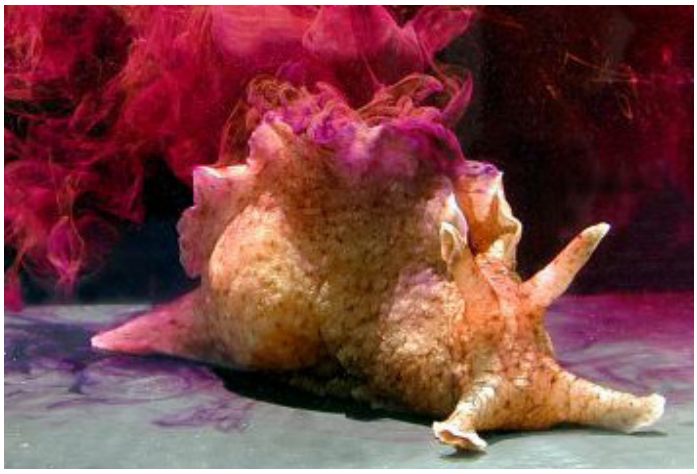


Figure 13.20 *Aplysia californica* was the model organism first used to demonstrate the neuronal level changes that underlie the habituation behavior.

such as a hungry predator (or an experimenter's paintbrush), grazed the siphon, the Aplysia would reflexively withdraw their gill, as if to protect this vital organ by shrinking away from the threat. However, after repeated brush strokes to the siphon, the sea slugs figured out that the stimulus was completely innocuous, and decreased the strength of gill withdrawal. Kandel and team suggested that this change in behavior was a form of learning.

Kandel discovered evidence of **habituation**, the suppression of a normal reflex behavior that is dependent on LTD. To further explore the cellular and molecular level changes behind this LTD, they conducted electrophysiological experiments on Aplysia. The gill-withdrawal reflex circuit relies heavily on two different populations of neurons: the sensory neurons that receive somatosensory information from the skin of the siphon, and the motor neurons that control the muscles of the gill. By using two different tiny glass pipettes, they could impale these neurons, inducing action potential firing in the sensory neuron, and observe changes in membrane potential of the motor

neurons (a depolarization of the membrane of a single neuron is called an **excitatory post-synaptic potential**, or **EPSP**). When the sensory neuron was activated, they observed an EPSP in response, since an action potential caused release of glutamate that activates post-synaptic receptors on the motor neuron. However, after the reflex had been habituated, the same sensory neuron activation caused a much smaller EPSP in the motor neuron.

The group went about seeing if they could modify this habituated response, curious if a stored memory can be modified by stimuli from the outside world. When they paired the mild siphon touch with a painful electric shock to the tail, the Aplysia began responding with a strong motor reaction, withdrawing the gill very intensely, indicating that the inhibited response disappeared. They called this observation **sensitization**. In electrophysiological studies, they observed that the EPSP at the motor neuron was much larger following the tail shock.

On a molecular level, presentation of sensitization is downstream of the action of a third population of neurons, interneurons that synapse onto the motor neurons. The noxious stimulus triggers these interneurons to release the neurotransmitter serotonin, which activates the excitatory signaling molecule cAMP found at the terminals of the motor neurons, thus increasing release probability and strengthening the gill withdrawal reflex.

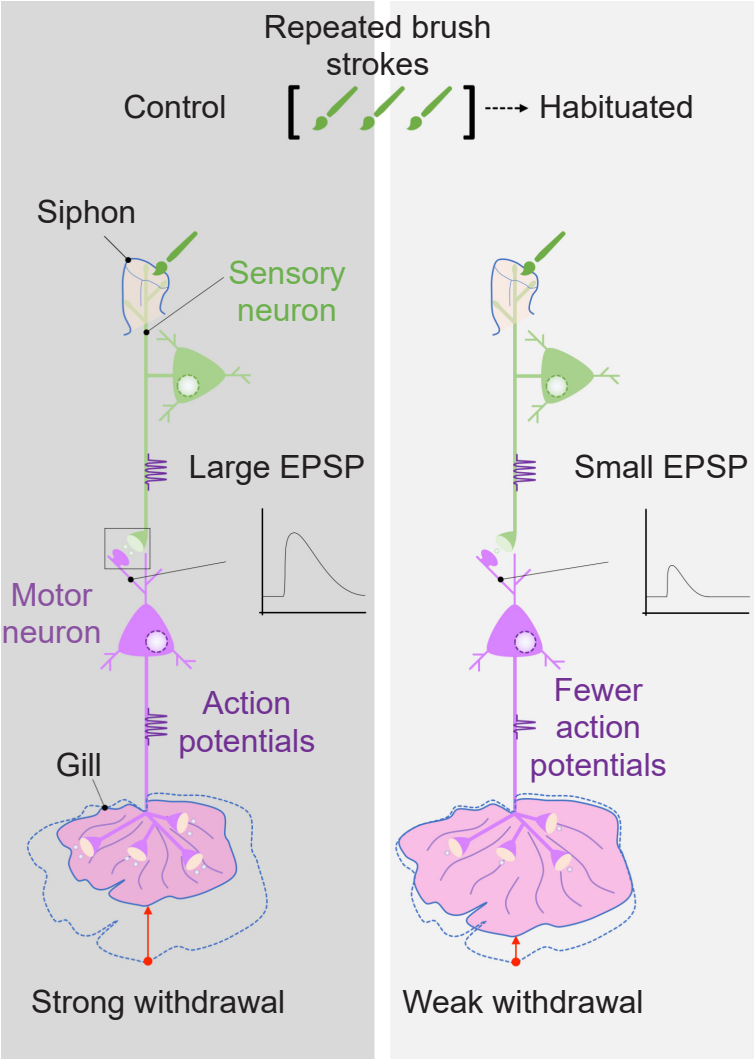


Figure 13.21 Habituation of the gill withdrawal reflex was modeled by Kandel using Aplysia. Repeated brush strokes to the siphon caused cellular level changes that led to decreased gill withdrawal.

Other forms of plasticity

There are many possible mechanisms that can lead to plasticity. Some activity patterns may induce a short-term plasticity that does not last on the scale of hours or years, but rather on the order of minutes or seconds.

One important form of plasticity is dependent on the activity of the endocannabinoid signaling pathway. During robust post-synaptic depolarization, enzymes such as phospholipase C or DAG lipase create signaling molecules called **endocannabinoids (eCBs)**, which are chemically similar to the primary psychoactive ingredient of *Cannabis* (chapter 11). These eCBs are neurotransmitters that act in a retrograde manner, meaning they are produced at dendrites and signal to the axons. Expressed at some populations of axons are inhibitory Gi/o coupled

receptors that respond to the presence of eCBs, called **cannabinoid receptors** (**CB1** is the main form that exists in the brain). These receptors inhibit cAMP signaling, which decreases release probability. This specific mechanism of plasticity is called **endocannabinoid-mediated LTD**.

Another atypical neurotransmitter, the gas **nitric oxide (NO)**, also contributes to a similar form of LTD. The enzyme nNOS synthesizes NO, which diffuses across cell membranes and can activate an intracellular receptor called **soluble guanylate cyclase (sGC)**. SGC changes the activity of the signaling molecule cGMP, which can either lead to enhancement or depression of synaptic strength.

There are also two directionally-opposite forms of plasticity that function at the millisecond timescale, called **paired-pulse facilitation (PPF)** and **paired-pulse depression (PPD)**. These forms of plasticity can be observed by stimulating a presynaptic area while recording simultaneously from the post-synaptic site. However, instead of delivering a single depolarizing pulse pre-synaptically, these forms of plasticity can only be observed by giving two pulses right after each other, usually separated in time by 10 milliseconds up to a second. In response to these depolarizations, you can also observe two EPSPs. Depending on the time between the pulses (called the **interevent interval**, or **IEI**) you may see that the second pulse is a different amplitude than the first. Dividing the amplitude of the second EPSP by the amplitude of the first EPSP gives us a value called the **paired pulse ratio (PPR)**. If the amplitude of the second pulse is increased relative to the first pulse (PPR > 1), this is described as a PPF, whereas if the amplitude in response to the second pulse is smaller than the first (PPR < 1), this is PPD. Both of these plasticity phenomena can be observed at the same synapse by varying the duration of the IEI.

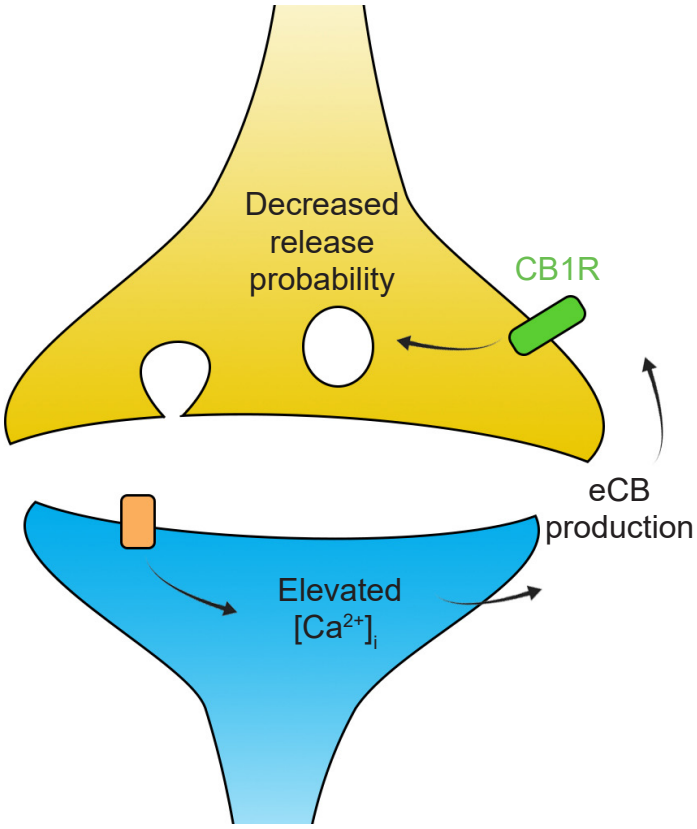


Figure 13.22 Endocannabinoid-mediated LTD is a result of a retrograde signal that decreases release probability.

By understanding the molecular mechanisms of synaptic release, we can explain how PPF and PPD can occur at the same synapse. Recall that increased Ca^{2+} concentration is an important pre-synaptic intracellular signal that allows for vesicular fusion and neurotransmitter release (see section 5.2 for more details). Generally, pre-synaptic Ca^{2+} enters through **voltage-gated calcium channels (VGCCs)** that open during depolarization. Delivering a single pulse causes some Ca^{2+} entry, but giving two pulses, one right after another, allows for more robust activation of the VGCCs thus increasing the intracellular Ca^{2+} greatly, leading to more neurotransmitter release.

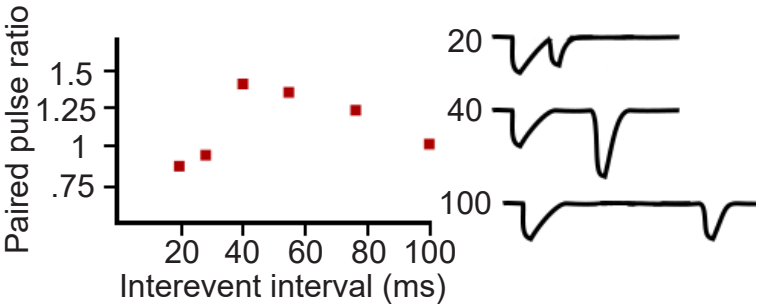


Figure 13.22 Paired pulse ratio an electrophysiological measure of short term plasticity.

PPD can also be accounted for by looking at the localization of vesicles in the axon terminal. The readily-releasable pool are the vesicles that are close to the inside of the cell membrane. During a depolarization of the terminal, these vesicles are the first ones to fuse. However, the readily-releasable pool of vesicles can get depleted. On the second pulse, there may be a small number of vesicles remaining because vesicles need to be refilled before they can be released, which makes the amplitude of the second pulse small compared to the first. This is an explanation for paired-pulse depression.

13.5 Disorders of memory

Alzheimer's disease

In 1907, Dr. Alois Alzheimer described one of his patients, a 50-year old woman named Auguste Deter, whose symptoms included profound cognitive impairment, memory deficits, and delusions. After Deter's death, post-mortem analysis of her brain revealed anomalies: degenerating neurons that contained atypical tangles and deposits scattered between cells. These reports were the first of a disease we now call **Alzheimer's disease (AD)**. It is an irreversible, slowly progressing neurodegenerative condition that leads to deficits in thinking, behavior, and memory loss. Specifically, declarative memory is the first form of memory loss observed in AD patients. As the disease progresses, procedural memory loss becomes more apparent.

AD is a devastating disease that accounts for 60-80% of all cases of dementia and is the sixth leading cause of death in the United States. As for prevalence, approximately 10% of people older than 65 and nearly a third older of people older than 85 has AD.

AD is divided into two categories, familial and sporadic. **Familial AD** is diagnosed when a person is in their 50s or 60s, and is strongly influenced by genetic risk factors. This form of AD only makes up about 10% of cases. **Sporadic AD** is by far more common than familial AD, and is believed to be caused by a combination of old age and environmental factors in addition to genetic risk factors.

Several genetic risk factors are linked to AD risk. **Apolipoprotein epsilon4 (ApoE4)** is the greatest genetic risk factor identified, where individuals who are homozygous for the e4 polymorphism have a 12-times higher risk of developing AD than people with the more common e3 variant. Additionally, mutations in genes like amyloid-precursor protein (APP), presenilin-1 (PSEN1) presenilin-2 (PSEN2), and triggering receptors expressed on myeloid cells

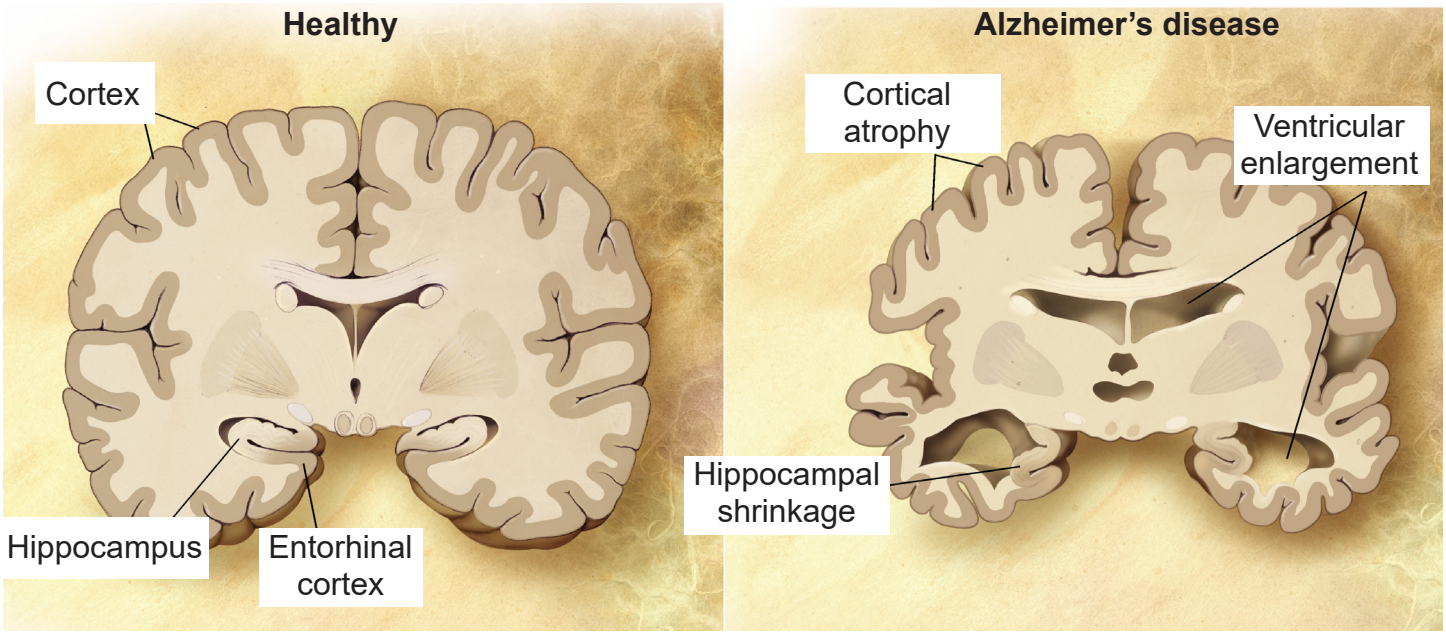


Figure 13.23 Gross anatomical changes observed in the brain of a patient with Alzheimer's disease.

2 (TREM2) are all associated with higher risk of developing AD.

To this day, only post-mortem analysis can conclusively diagnose someone with AD. As expected, this raises many issues in terms of treating these patients. Currently, physicians diagnose a patient with AD based on the symptoms they presented with. However, new techniques to make more accurate diagnoses are emerging, including identifying the presence of biomarkers from blood samples, certain functional characteristics as observed in PET or fMRI imaging, and possibly even through eye exams!

An early hypothesis proposed to explain the neuronal loss and memory deficits observed in AD is centered around changes in the signaling of acetylcholine. Early post-mortem analyses discovered that people with AD have profound atrophy of the basal forebrain, an area which contains a large density of cholinergic neurons. The **cholinergic hypothesis** suggests that it is this loss of cholinergic neurons and the loss of acetylcholine signaling that is the main pathological driver of AD. The theory is supported by the idea that acetylcholine plays an important role in learning and memory. Three of the four FDA approved drugs for AD are **acetylcholinesterase (AChE) inhibitors**, drugs that act to increase levels of acetylcholine. Unfortunately, these compounds show only short-term benefits for cognitive function, and these therapeutic effects subside over time.

Pathologically, AD is characterized by the extracellular accumulation of **amyloid-beta plaques (Aβ)** and intracellular hyperphosphorylated **neurofibrillary tau-tangles (NFT)**.

The **amyloid cascade hypothesis**, originally proposed in 1992, has become the leading theory in the field describing how AD develops. The hypothesis suggests that the main driving factor of AD is the deposition of Aβ in the brain, and this in turn leads to neurodegeneration via cell death, abnormal protein buildup, and neuroinflammation. Aβ is produced from the cleavage of **amyloid-precursor protein (APP)**, an integral membrane protein expressed by neurons. A class of secretase enzymes are responsible for the degradation of APP. When the alpha and gamma secretases degrade APP, the resulting Aβ protein is likely to clump together in unpredictable ways, leading to the formation of **Aβ plaques** (pronounced “A beta”). These plaques can cause neuronal death and lead to the cognitive deficits observed in AD patients. Two major genetic risk factors for AD, presenilin 1 and 2, are mutations of the gamma secretase

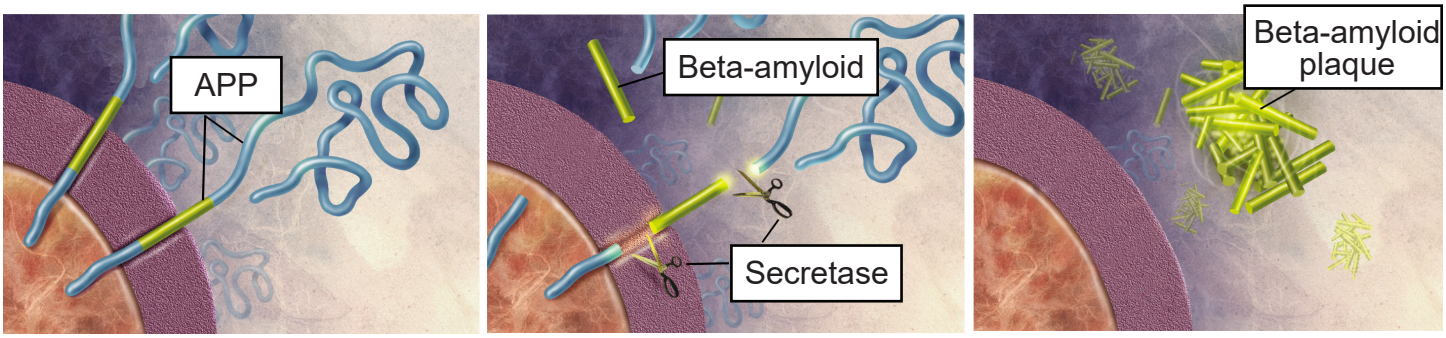


Figure 13.24 Amyloid precursor protein is processed by a secretase enzyme, resulting in beta amyoid, which may accumulate and form plaques in the extracellular space.

that leads to increased APP cleavage.

Therapies are being developed to target and control the Aβ protein to mitigate AD symptoms. Recently an antibody-based drug has been approved by the FDA. This groundbreaking drug is unique from those traditionally prescribed because it directly targets Aβ.

However, the amyloid-cascade hypothesis does not tell the whole story. For one, there are patients who carry a heavy Aβ load but present no clinical symptoms of AD. More so, in mouse models with APP mutations, they develop significant Aβ plaques, but have no accumulation of tau (see below) and no significant neurodegeneration. Additionally, there have been several drugs targeting Aβ that have failed in human trials.

While these amyloid plaques seem to be the primary driving factor of AD, **hyperphosphorylated tau** and **neurofibrillary tangles (NFTs)** are also potential culprits in disease progression. NFTs are an intracellular pathological marker of AD and correlate strongly with cognitive deficits in AD. Tau protein is the major microtubule-associate protein (MAP) of mature neurons, and helps to function to maintain cellular morphology. Excess phosphorylation of this protein causes tau to accumulate inside the cell, leading to neuronal dysfunction and cell death. The presence of Aβ increases the levels of the tau, and vice versa, adding to the complexity and difficulty of treating AD. Tau pathology is also observed in other neurodegenerative diseases such as frontotemporal lobe dementia and Parkinson's disease.

Although most discussion of AD revolves around plaques and tangles, it is well known that a

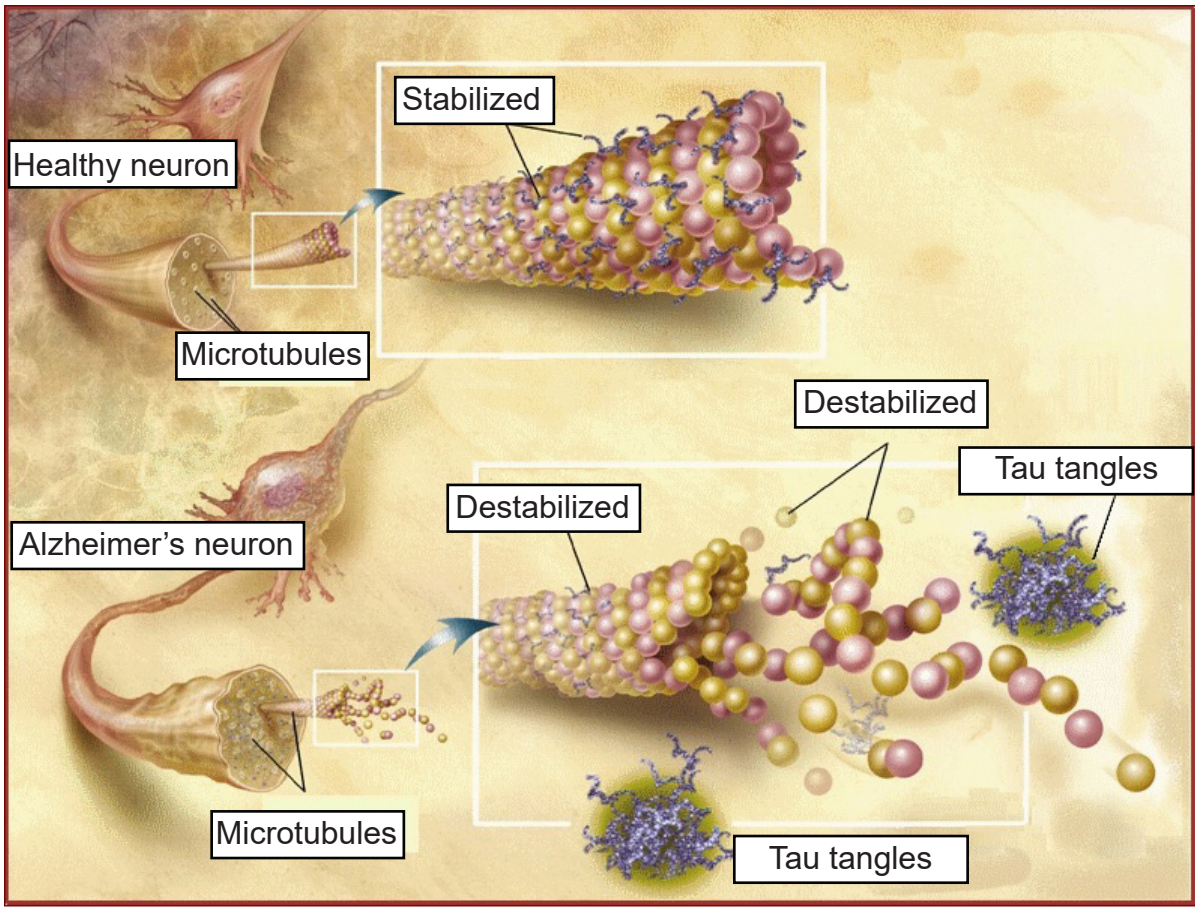


Figure 13.25 Tau protein is used in stabilization of microtubules.

cacophony of pathological markers are also seen in AD, including neuroinflammation, oxidative stress, blood-brain barrier dysfunction, heavy metal dysregulation, mitochondrial impairment, and many more.

Korsakoff's syndrome

Korsakoff's syndrome is a disorder resulting from a severe deficiency of **thiamine**, an essential vitamin that functions in metabolic processes. Dietary thiamine is found in whole grains, legumes such as beans and peas, as well as some meats and fishes. Healthy people with a well-balanced diet get sufficient thiamine, however gastrointestinal illnesses can cause an inability to absorb thiamine properly. Chronic alcohol misuse also impairs the body's ability to take up thiamine, and is often the cause of Korsakoff's syndrome.

People with Korsakoff's syndrome experience both retrograde and anterograde amnesia, as well as severely impaired short-term memory. The patients may experience a very strange behavior called **confabulation**, which is the fabrication of false memories ranging from subtle to wildly fantastical. People who confabulate do not consciously recognize that their statements are untrue, and are not intentionally trying to deceive others, which is why it is sometimes called "honest lying". Generally, confabulation only happens as a person is trying to recall recent autobiographical memories, while their semantic and procedural memories are less susceptible to confabulation. Scientists suggest that people confabulate as a compensatory mechanism to make up for their retrograde amnesia.

Destruction of both neurons and glia are seen in the brains of people with Korsakoff's syndrome. As a result of this cell loss, there is often shrinkage of the cortex, thalamus, the hippocampus, cerebellum, and the **mammillary bodies**, paired structures located at the ventral surface of the brain close to the brain stem, themselves part of the limbic system.

Korsakoff's syndrome can be treated by giving thiamine supplements and eliminating alcohol consumption. If treated within days after the onset of brain damage, people are expected to make a complete recovery. However, one of the major challenges is making a proper diagnosis, since the symptoms of Korsakoff's syndrome present similarly to other disorders.

Traumatic brain injury

A sudden blow to the head is likely to produce a **traumatic brain injury (TBI).** TBIs are the most common form of brain damage, with about 2.88 million Americans in 2014 with a TBI-related emergency department visit, hospitalization, or death. People are exposed to these injuries

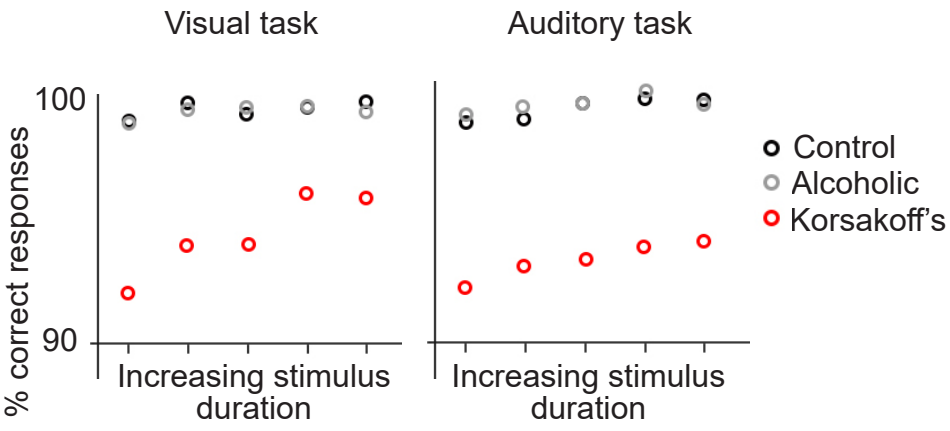


Figure 13.26 On a delayed-response task, a measure of short term memory, patients with Korsakoff's perform significantly worse than either a control population or a similar group of alcoholics.

regularly, such as in automobile accidents, falls, high-contact sports, occupational hazards such as in construction or military deployment, among many others. A **concussion** is a mild form of a TBI. Some of the common symptoms that people experience shortly after experiencing a TBI include double-vision, headache, dizziness, nausea and vomiting, and loss of consciousness, which are common symptoms in concussions as well.

Usually in a traumatic brain injury, there are two simultaneous insults: the **coup**, which is the injury when the brain hits the inside of the skull closest to the external force. Shortly after, the **contrecoup** happens, which is the injury when the brain recoils backwards and hits the interior surface of the skull opposite of the cause of the insult. Following these injuries, there is a widespread inflammatory response that changes cellular activity in numerous complex ways, such as microglial activation, axonal stretching, and elevated levels of immune signaling molecules.

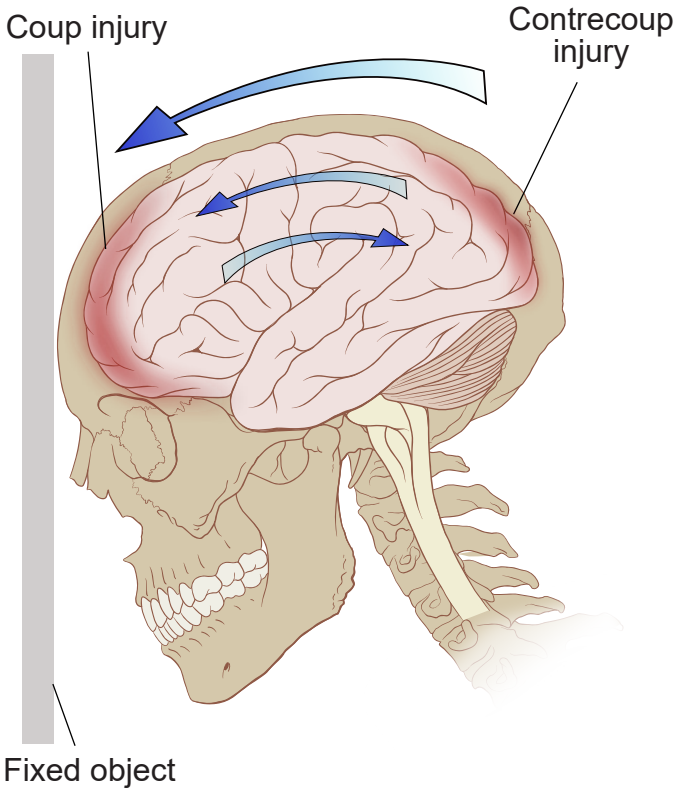


Figure 13.27 TBI is often characterized by a pair of injuries, the coup and the contrecoup injury.

People with repeat head injuries may also experience some memory loss. It is suggested that the damage to the brain creates an abnormal cellular environment that interferes with the reverberation processes that are normally important for solidifying memories. Additionally, TBI decreases hippocampal neurogenesis, which is likely to impair the formation of declarative memories.

Savant syndrome

Savants are people with "islands of genius," knowledge in extreme specializations, such as ability to calculate large prime numbers, identify the specific day of the week for any date, or recall sports trivia. Many savants have a passion for creative endeavors, such as in the visual arts or musical performance. About 20% of savants demonstrate extraordinary memory. The prevalence of savants in the general population is exceedingly low, but about 1 in 10 autistic people demonstrate some expression of savant skills.

While most savants are born with these skills, some acquire them as a result of a head injury. Orlando Serrell, after losing consciousness in a baseball accident, developed a complete autobiographical memory, able to remember where he was and exactly what he was doing on any given date after his injury. Following a diving accident, Derek Amato developed prodigious musical talents despite never having taken lessons, improvising on the piano as if playing through procedural memory.

One of the biggest difficulties is that TBI is very different from case to case, making development of new therapeutic strategies challenging. Currently, the best treatment for TBI is rest. There is still debate about the types of medications that can help best improve outcomes, psychostimulants and antidepressants being the most common classes of drugs used in TBI.

The Open Neuroscience Initiative is funded by a grant from the Vincentian Endowment Fund of DePaul University.

Cover art: https://pixabay.com/photos/handmade-birthday-cake-dessert-3989069/
13.1 https://en.wikipedia.org/wiki/File:Henry_Gustav_1.jpg
13.2 https://commons.wikimedia.org/wiki/File:Hippolobes.gif
https://commons.wikimedia.org/wiki/File:Hippocampus_and_seahorse_cropped.JPG
13.5 https://pixabay.com/vectors/pencil-sharp-school-supplies-153561/
13.6 https://commons.wikimedia.org/wiki/File:Classical_Conditioning_Diagram.png modified by Austin Lim
13.8 https://commons.wikimedia.org/wiki/File:CajalHippocampus.jpeg modified by Austin Lim
13.9 https://commons.wikimedia.org/wiki/File:MorrisWaterMaze.svg modified by Austin Lim
https://commons.wikimedia.org/wiki/File:Simple_Radial_Maze.JPG
13.10 https://commons.wikimedia.org/wiki/File:Amygdala.png
13.11 https://commons.wikimedia.org/wiki/File:202003_Model_animal_mouse_mono.svg
https://pixabay.com/vectors/sound-audio-music-icon-studio-2935370/
13.12 https://commons.wikimedia.org/wiki/File:Ventral-dorsal_streams.svg modified by Austin Lim
13.13 https://commons.wikimedia.org/wiki/File:Gray727_fusiform_gyrus.png modified by Austin Lim
https://commons.wikimedia.org/wiki/File:Fusiform_face_area_face_recognition.jpg
https://pixabay.com/photos/man-water-wet-male-face-swimming-984504/ modified by Austin Lim
13.14 https://commons.wikimedia.org/wiki/File:The_structures_of_the_basal_ganglia.png
13.16 https://commons.wikimedia.org/wiki/File:Examples_of_Grid_Cells_with_Different_Grid_Spacing_and_Field_Size.jpg
13.17 https://commons.wikimedia.org/wiki/File:Mark_Hamill_(right)_and_me_(left)_(210269249).jpg
https://commons.wikimedia.org/wiki/File:Mark_Hamill_(1980).jpg https://commons.wikimedia.org/wiki/File:Star_Wars_characters_at_Madame_Tussaud.jpg
https://commons.wikimedia.org/wiki/File:Retrato_del_Maestro_Yoda.jpg https://commons.wikimedia.org/wiki/File:Leonardo_DiCaprio_2014.jpg https://commons.wikimedia.org/wiki/File:Keanu_Reeves_2014.jpg
13.18 https://commons.wikimedia.org/wiki/File:LTP_exemplar.jpg modified by Austin Lim
13.19 https://commons.wikimedia.org/wiki/File:Potential_mechanisms_of_LTP_in_spinal_dorsal_horn_in_vivo.jpg modified by Austin Lim
13.20 https://commons.wikimedia.org/wiki/File:Aplysia_californica.jpg
13.21 https://commons.wikimedia.org/wiki/File:Habituation.png
13.23 https://commons.wikimedia.org/wiki/File:Alzheimer%27s_disease_brain_comparison.jpg
13.24 https://commons.wikimedia.org/wiki/File:Amyloid-plaque_formation-big.jpg modified by Austin Lim
13.25 https://commons.wikimedia.org/wiki/File:TauProtein.jpg modified by Austin Lim
13.26 https://upload.wikimedia.org/wikipedia/commons/a/a8/Alcohol-induced_brain_damage_%28IA_alcoholinducedbr00hunt%29.pdf modified by Austin Lim
13.27 https://commons.wikimedia.org/wiki/File:Contrecoup.svg

Chapter 14:

Lateralization and Language

Austin Lim, PhD (DePaul University)
Editor: Sandra Virtue, PhD (DePaul University)

A common misconception among non-scientists, popularized by the media and online quizzes, is that analytic people are “left brained” while the creatives among us are “right brained” (Chapter 1). Modern studies have concluded repeatedly that correlating brain function with behavior on this broad level is not this simple. Both hemispheres of the brain are capable of carrying out the same essential functions: processing sensory and perception information, motor communication to the body, and the storage and retrieval of memory.

However, there are some features that are slightly more focused in one hemisphere than the other. We describe these features as being **lateralized**. Many different functions have a slight preference in lateralization: for example, the right hemisphere seems slightly better at making judgments about the duration of visible stimuli or processing of low-frequency musical stimuli. Keep in mind, the left hemisphere can also perform these functions, just not quite as well as the right can.

One heavily-lateralized function is language: for most people, the production and comprehension of language is dominated by structures in the left hemisphere of the brain. This chapter deals with these particular functions.

Chapter 14 outline

14.1 Lateralization

Almost all mammals are bilaterally symmetrical, with a left half that is more or less a mirror image of the right half. The internal organs, however, often tell a different story. We have a single stomach, liver, and heart, none of which are symmetrical. Even paired organs like the lungs or kidneys, are slightly asymmetrical. The brain can most accurately be thought of as a pair of intimately-connected organs with subtle differences in function.

The brain’s two hemispheres are connected by white matter tracts which allow the two halves to communicate. The largest interhemispheric white matter tract is the **corpus callosum**, which is made up of 200-250 million axons. If you held a human brain and separated the two hemispheres dorsally along the longitudinal fissure, you would be able to see the fibers of the corpus callosum holding the two halves together. The corpus callosum is about 10 cm (~4 inches) long from

anterior to posterior, and the middle part of the structure forms the dorsal-most roof of the lateral ventricles.

In addition to the corpus callosum, there are a handful of other white matter tracts that allow the hemispheres to communicate. The much-smaller **anterior commissure** is a tenth of the thickness of the corpus callosum, connects the two temporal lobes, and conveys important limbic information such as memory and emotion. The **hippocampal commissure** is one of the outputs of the hippocampus that connects the structures in the left and right hemispheres. These small white matter tracts are often used as points of reference in imaging studies or surgical dissection.

A pair of researchers, Drs. Ronald Myers and Roger Sperry, were very curious about these pathways of communication between the two hemispheres. In the 1950's, they wanted to understand how information from one visual field gets conveyed into the opposite hemisphere of the brain. To answer the question of **interhemispheric transfer**, they conducted experiments in cats. One of their early experiments presented healthy cats with two different boxes, only one of which contained food. An eyepatch was placed over one of the cat's eyes, and the cat was free to paw at one of the boxes, which if chosen correctly, would yield the food reward. At first, as expected, the cat would choose from the boxes at random, obtaining the food reward 50% of the time. Over multiple trials, as the cat began to learn which box held the food, the success rate rose to picking the rewarded box 100% of the time. When the eyepatch was then moved to the other eye, the cat performed the task correctly 100% of the time, reliably picking the box associated with food.

Then, Myers and Sperry performed a two different surgical procedures on the cats. One severed the optic chiasm, which kept visual information in the ipsilateral hemisphere. This ensured that when wearing the eye patch, visual information does not cross into both hemispheres. The other procedure to severed their corpus callosum, a process called a **corpus callosotomy** (or **commissurotomy**), which limited interhemispheric transfer after visual cortex processing. Between these two interventions, there were four groups of cats: Fully intact, optic chiasm cut, corpus callosum cut, and the experimental group with both optic chiasm and corpus callosum cut..

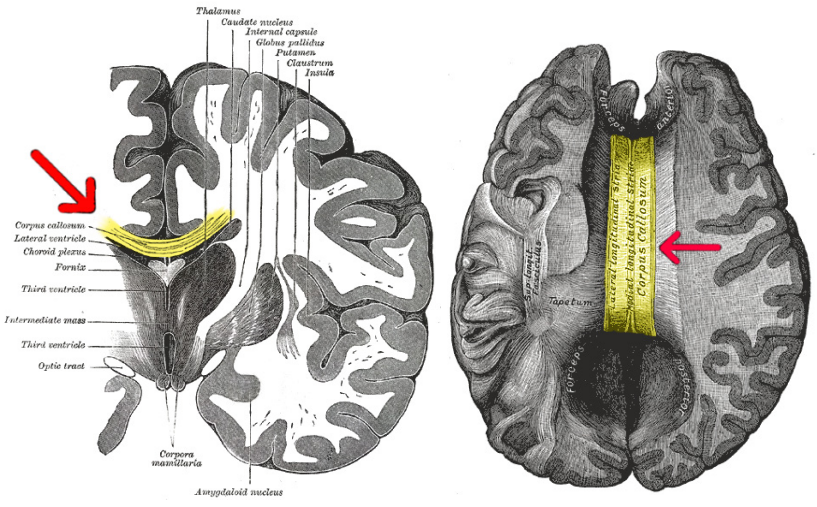


Figure 14.1 The corpus callosum, indicated in yellow with red arrow, in a coronal slice (left) and seen from the top when both hemispheres are gently pulled apart (right).

The box-selection behavioral experiment was then repeated. As with the intact cats, when the eyepatch was placed over one eye, the experimental cats (both chiasm and corpus callosum severed) initially guessed at the boxes, getting the reward 50% of the time. Again, as before, these animals improved their performance over repeated trials, eventually getting the

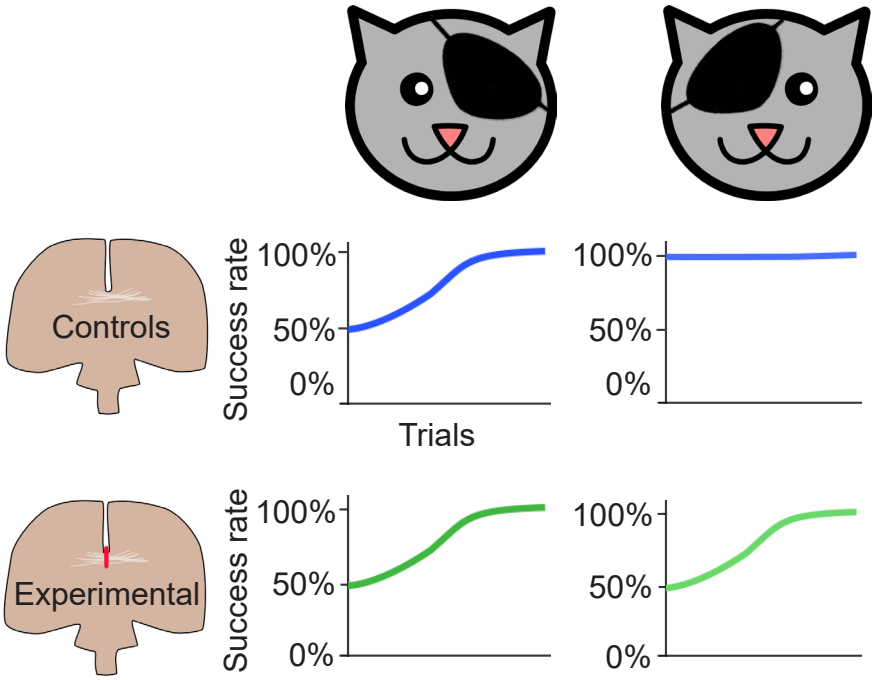


Figure 14.2 Myers and Sperry demonstrated that each hemisphere is capable of learning and storing memories independently.

reward every time. However, after the eyepatch was switched from one eye to the other, these cats essentially had to "start over" with their learning: they picked the rewarded box only 50% of the time, improving to 100% over trials. Because of the surgical procedures, the visual information and associated reward memory in one hemisphere never made it to the other half of the brain - a failure of interhemispheric transfer.

The two other control groups immediately performed at 100% after the eyepatch was switched over, just as well as the fully intact cats. When the optic chiasm was severed with the corpus callosum intact, the visual information remained in the ipsilateral hemisphere, but after processing in V1, that information passed over the corpus callosum to the contralateral hemisphere. When the corpus callosum was severed with the optic chiasm intact, the visual information made their way into both hemispheres through the optic nerve.

Myers and Sperry then extended their research to humans. Sometimes, commissurotomy is suggested for younger patients with drug-resistant epilepsy. Grand mal seizures are often characterized by uncontrolled electrical activity in one hemisphere, which then crosses the corpus callosum to the other hemisphere before "bouncing back" to the original hemisphere. During the procedure, the surgeon cuts the corpus callosum, and in doing so, keeps the atypical electrical activity isolated in one hemisphere. Patients have significantly fewer and less severe seizures following recovery from the operation.

People who have had this surgery are sometimes called **split-brain patients**, a population of patients who were extensively

Clinical connection: Agenesis of the corpus callosum

In a handful of rare cases, people can be born without a corpus callosum, a condition called **agenesis of the corpus callosum (ACC).** Some people with ACC develop atypically, experiencing seizures and poor motor control or coordination. An estimated one-fourth of people diagnosed with ACC after birth have some intellectual disability, but most have typical levels of intelligence. They may have subtle abnormal developmental traits, such as a difficulty with processing common social cues (as seen in autism). Notably, the real-life savant who served as the inspiration for the movie *Rain Man* was born with ACC.

studied by Dr. Michael Gazzaniga.. Overwhelmingly, split-brain patients are healthy with no significant changes in intelligence and no dramatic changes in personality. However, some of them do experience deficits in memory and concentration.

Among split-brain patients, very unique behavioral and cognitive deficits can be observed under specific experimental circumstances. The baseline test begins by briefly showing the patient some visual stimulus, such as a picture of a donut, only in their right visual field, which gets represented in the left visual cortex (refer back to chapter 7.2 for a reminder of the circuitry of the visual system). When asked what the patient had seen, they would report "a donut," just as any typical person would (because the left hemisphere is highly involved in language and enables the person to report the object verbally).

In a second experiment, both of the patient's hands are placed on a table hidden behind a screen. An object, such as an apple, is then placed in their right hand. As the patient feels that object, tactile information such as its hardness, diameter, and temperature, ascends contralaterally into the left somatosensory cortex (chapter 8). After the object is removed from their hand, the patient is asked to feel blindly through a collection of objects, all hidden behind the screen, and find a matching object. When doing this task with the right hand, they would be successful in selecting an apple (because the motor system is crossed). However, when the left hand was now tasked with reaching behind the screen to select a matching object, they would not be able to know which object to pick up because this information goes to the right somatosensory cortex (which has no knowledge of the apple). From these data, the researchers concluded that each hemisphere is independently capable of receiving their own sets of somatosensory inputs

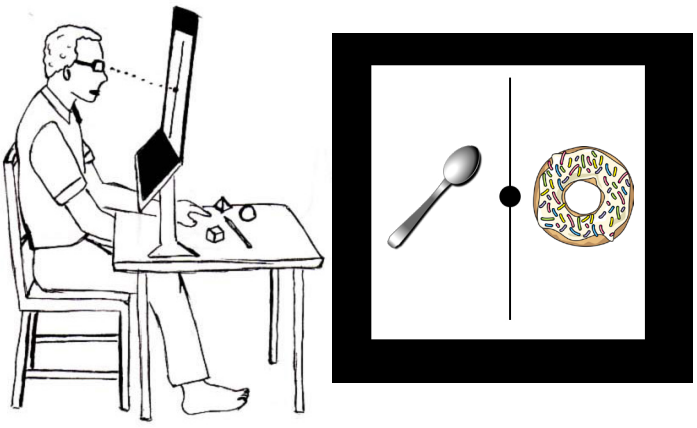


Figure 14.3 Experimental setup for studying interhemispheric transfer of visual information in humans (left). After fixation on the dot in the center of the visual field, the stimuli (right) are flashed briefly, and the patient is asked the either name the item observed, or reach behind the screen and select a matching object.

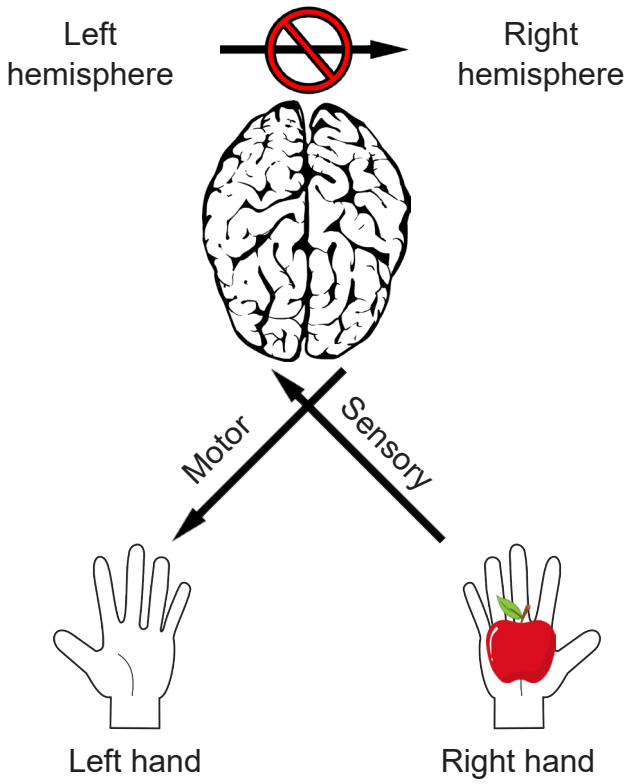


Figure 14.4 In split-brain patients, when an apple is placed in the right hand, that information ascends contralaterally but cannot cross hemispheres.

and storing their own memories. Without an intact corpus callosum, the two hemispheres are unable to share that knowledge, so the sensory and memory information that reaches the left hemisphere isn't capable of reaching the right hemisphere, which controls the left hand - so the left hand is clueless to the object placed in the right hand.

In the next step of the experiment, a different visual stimulus, like a picture of a spoon, is presented to the left visual field, which is initially sent to the right half of the brain. When asked what they saw, they might say "nothing" or "I don't know." (because the right hemisphere is not specialized for language and the person is not able to report the object verbally). But, when the patient is asked to reach behind the screen with their left hand, they could successfully select a spoon! (Left hand is controlled by the right brain, which has knowledge of the spoon.) Their right hand, however, couldn't correctly pick a matching object (since the left brain does not have the information about the spoon). Again, these results demonstrate that each hemisphere is capable of receiving their own contralateral sensory information and storing their own sets of memory.

For most people, who have their corpus callosum intact, information is transferred rapidly between hemispheres. So, when a spoon is shown to our right brain, the left brain learns that information as well, which is why we would be able to select a matching object with our right hand.

Myers and Sperry's human studies noted an interesting difference in the ability of split-brain patients to respond verbally. When the stimulus was sent into the left hemisphere, either a visual stimulus in the right visual field or an object placed in the right hand, the patients were able to verbalize what they either saw or felt. But, when the stimulus was represented in the right hemisphere, they couldn't. Their conclusion was that the left hemisphere is much better equipped for language-related functions compared to the right hemisphere. As it turns out, language comprehension and production is heavily lateralized to the left hemisphere. For his work regarding the "effects of disconnecting the cerebral hemispheres", Dr. Sperry earned the 1981 Nobel Prize.

14.2 Language

Language is one of Homo sapiens' greatest intellectual evolutionary accomplishments. Using language, we are able to communicate very complex concepts, such as survival instructions (Don't eat those berries because they taste weird and you'll get sick) or a shared belief in the existence of complex stories (Hang stockings by the chimney and you'll get presents if you were good). Language, when used in these ways, has a powerful influence on behavior, and modern humans rely heavily on language in every aspect of society.

Components of language

Speech pathology experts have identified at least four distinct components for describing different aspects of language. The most granular unit is the **phoneme**, which is an individual sound that generally has no meaning on its own. For example, the word map can be split into three phonemes, "mm", the short "/ă/", and "p" sound.

The next larger unit of language is the **morpheme**, which is a combination of phonemes. Morphemes are capable of conveying an idea, such as "cat". Suffixes such as "-s" and "-ing" also convey ideas (plural and verbs in action, respectively) are also considered morphemes.

The **syntax** represents the next higher level of language, which is the information conveyed when words are combined in order to produce meaning at the level of phrases and sentences. For example, a statement such as "He gave a gift to his brother" contains syntactic information identical to "He gave his brother a gift", even though the organization is different. The grammatical rules of many languages tell us the order of nouns, verbs, and objects, and inappropriate deviation from these rules can change the meaning of the sentence dramatically.

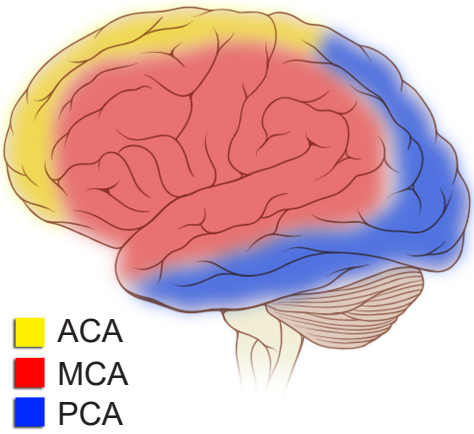
Semantics refers to the understanding of meaning, especially the meaning of words in relationship to one another in a phrase, sentence, or paragraph. Extracting meaning from statements not meant to be taken literally (such as a hungry person exclaiming "I'm so hungry, I could eat a horse!") and identification of the meaning of a word under two different contexts (such as in the sentence "I held a nail between my fingers, but when I swung the hammer, I hit my nail instead.") fall under the category of semantics.

Brain structures involved with language

Whereas the left and right hemispheres of the brain are mostly symmetrical, one of the biggest asymmetries is related to the structures responsible for language. Myers and Sperry observed that split-brain people can verbally report observations made with the left brain, while having difficulty when information is stored by their right brain. This suggests that the left hemisphere is dominant for language functions. It is estimated that about 90% of right hand-dominant people and about 50% of left hand-dominant people use their left hemisphere for language related functions.

However, this does not mean that the other hemisphere does not contribute to language. The right hemisphere, for example, shows activation during the use of nonliteral language, such as in metaphor production or irony comprehension.

In addition to the split-brain patient case studies, there are several other significant pieces of evidence to support left hemispheric dominance for language.



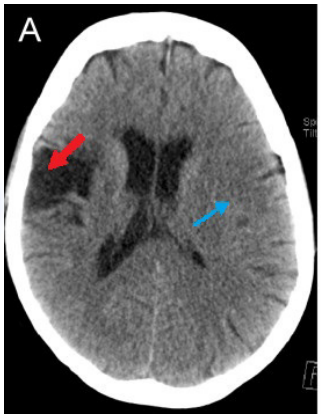


Figure 14.5 Areas of the cortex that receive blood flow from specific arteries (left). The middle cerebral artery (MCA) provides perfusion to frontal, parietal, and temporal areas that are important for language. CT scan of a patient after a stroke of the MCA, showing loss of brain tissue (red arrow, right).

People with left hemisphere lesions may lose their language capacities. A stroke of the left middle cerebral artery often leads to a variety of language related deficits. Unfortunately, similar injuries sometimes happen after brain surgery, traumatic brain injury, or brain infections, also resulting in language deficits when localized to the left hemisphere.

Experimental methods have allowed researchers to study the lateralization of language without causing any permanent damage. The **Wada test** is the most reliable method by which hemispheric lateralization of language can be determined. Named for the Japanese-born neurosurgeon Jun Wada, the test is a presurgical assessment to minimize the risk of a person losing their language capacity in the process of brain surgery. The protocol begins with the surgical team asking the patient to hold up both hands, wiggling their fingers, while counting. The patient then receives an intravenous infusion of **sodium amytal**, a GABA receptor positive allosteric modulator that acts as an anesthetic. When infused into the internal carotid artery, the drug gets delivered into just one hemisphere of the brain with little leakage into the other. When the anesthesia perfuses through the left brain, their right hand loses muscle tone and their fingers will stop moving (remember the contralateral organization of the motor control system, chapter 10.) And, if language is lateralized in this hemisphere as it is for most people, they will also be unable to count during this time. Within seconds, the anesthesia is cleared from the brain, and the wiggling and counting resume. If the patient is right hemisphere dominant for language, then they will be able to count, even though the fingers stop moving. The procedure is then repeated while the anesthetic is perfused into the other hemisphere.

The Wada test, because of its invasive nature and occasional side effects (pain, infection, and seizure or stroke in very rare cases), is used less frequently as functional brain imaging methods have become cheaper and more available through the 2000s. The fMRI is a preferred test of hemispheric dominance. To conduct these tests, a person is put into the imaging machine, then asked to perform a series of language tests, such as listing several items of a given category, or listening to a conversation in preparation for follow-up questions. During this process, the fMRI informs the medical team about which half of the brain shows greater activity during the language tests. These behavioral tests have been found to be as accurate as the Wada test in determining lateralization of language functions.

Across the language dominant hemisphere, there are a few brain regions that

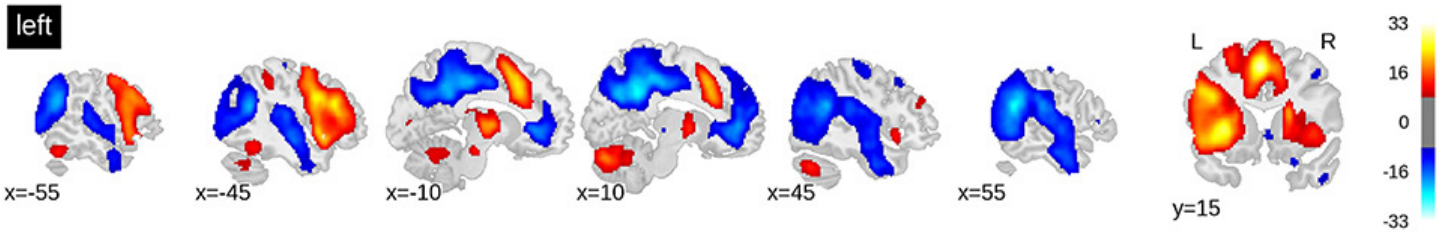


Figure 14.6 Non-invasive fMRI scans demonstrate that left hemisphere (negative x) brain areas increase in blood flow compared to right hemisphere (positive x) during the performance of language tasks. Warmer colors indicate increases in blood flow, while cooler colors represent decreases.

contribute significantly to language functions. When something goes wrong with these areas, a person may develop **aphasia**, a language disorder. It is estimated that about 180,000 new cases of aphasia are diagnosed in the United States annually. Stroke is a common cause of aphasia, but other neurological insults such as head trauma, traumatic brain injury, or subdural hematoma can induce aphasia. Just like nearly everything in biology, there is a wide range of severity, with some cases being very minor and other cases being much more severe. Speech therapy can help a patient recover from aphasia, and this progressive restoration of function is a demonstration of the brain's capacity for plasticity and remodeling.

Expressive (or non-fluent; or Broca's) aphasia

One of the first language-related cortical structures identified is the posterior **inferior frontal gyrus (IFG)**. Deficits in this area lead to a difficulty with the production of language.

In the 1860's, a patient named Louis Victor Lebourgne had a very unusual condition: he could only speak one syllable. For Lebourgne, the syllable "tan" meant everything, from "yes" to "no" to "hat" to "thirty-four". Lebourgne would say "tan" while gesturing emphatically, scream "TAN TAN!!" when angry, and whisper "tan" when telling secrets. Because of this, the staff at the hospital called him **Patient Tan**.

When Patient Tan died, the French physician Paul Broca performed an autopsy on the brain. Broca discovered a huge lesion about the size of a "chicken's egg" in the left hemisphere, just dorsal of the lateral fissure in the frontal lobe. Soon after, Broca performed autopsies on the brains of seven other patients with similar language difficulties, all with the same prominent injury to this portion of their frontal lobe. Because of the work that Broca did in correlating structure with function, the posterior IFG came to be called Broca's area.

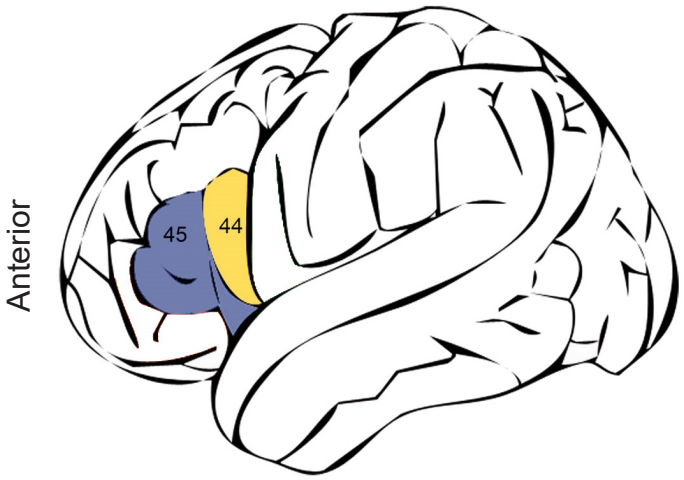


Figure 14.7 The posterior IFG, or Broca's area (labeled as 45 and 44; purple and yellow) contribute to language production.

Today, we understand that a localized injury to the IFG produces a form of aphasia called **expressive aphasia** (also called **non-fluent aphasia** or **Broca's aphasia**). These patients have difficulty expressing themselves, only speaking in short, effortful phrases, using just nouns and verbs while omitting tenses, conjunctions, and prepositions. They speak haltingly, sometimes filling the silences in their sentences with filler phrases. The patients are profoundly aware of their deficit, leading to overwhelming frustration with their inability to communicate. They know what they want to say, but often can't get it out. Interestingly, these patients do not have any significant impairment of comprehension.

Patients with IFG injury show similar expressive deficits regardless of the modality of their language. For example, when asked to write, they write slowly, using mostly nouns and verbs. Alternatively, patients who use American Sign Language also lose grammatical syntax and communicate slowly when signing!

Receptive (or fluent; or Wernicke's) aphasia

A different brain structure, called the **superior temporal gyrus** is linked to language comprehension. This area is sometimes also called **Wernicke's area**, named for the German physician named Carl Wernicke, who studied a group of patients with a different form of aphasia than Broca's. These patients had no deficits in the production of speech, but the words they used were very disorganized. They could speak complete sentences fluently, but their speech contained almost no substantial semantic content. Unlike Broca's patients, Wernicke's patients had

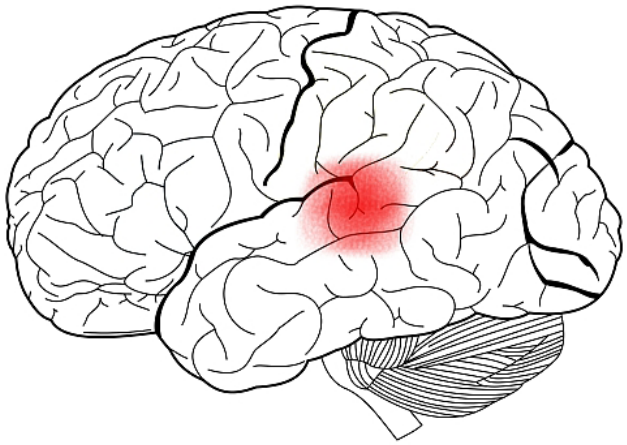


Figure 14.8 The superior temporal gyrus, or Wernicke's area (red), contributes to language comprehension.

dramatic impairments in comprehension. This language disorder is **receptive aphasia** (or **fluent aphasia**, or **Wernicke's aphasia**.)

While talking, people with receptive aphasia may create new meaningless words they are unaware of, a symptom called **paraphasia**. These words could be a mispronunciation of a word, perhaps sounding like the jumbling of syllables. They can happen at the level of the phoneme or morpheme, such as in nonwords such as "emchurch" or "plehzd". They also appear at the level of syntax, when a person substitutes a word incorrectly for another, as in the sentence "But I seem to be table you correctly, sir."

Sometimes, they experience a difficulty with recalling words, a symptom called **anomia**. This happens in the middle of a sentence, and may be difficult to catch in casual conversation, since they will often use vague language ("stuff" or "things") or use several words in a roundabout fashion to describe what they are trying to say, a behavior called **circumlocution** ("red, it's green, and yellow means be cautious, to keep people safe" instead of "traffic light.")

Conduction aphasia

Early theories suggested that communication between the STG and the IFG is important for healthy language production and comprehension. Anatomically, a band of white matter called the **arcuate fasciculus** spans these areas, originating in STG and terminating in the IFG. When this structure is injured, people develop some difficulty with repeating language they hear, a disorder called **conduction aphasia**. Generally, these patients display paraphasias when asked to repeat multisyllabic words, often switching phonemes around in a single word.

These patients have no significant deficits in language production or comprehension, presumably because their IFG and STG are still intact and healthy. Conduction aphasia is less severe than expressive or receptive aphasia.

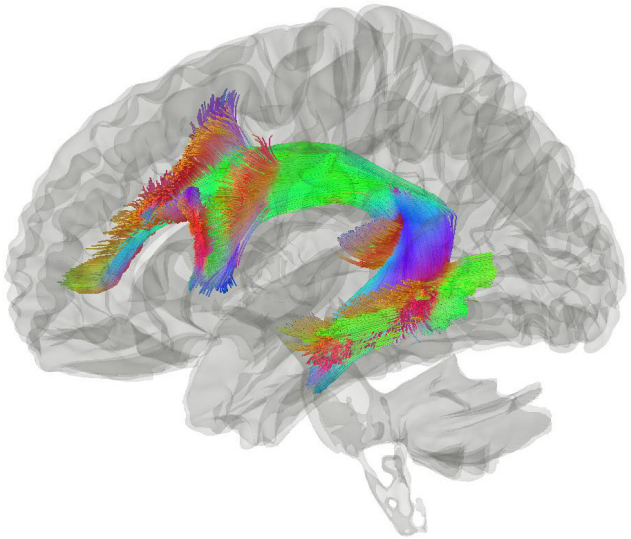


Figure 14.9 The arcuate fasciculus (colored) is a large white matter band that connects the two major language-related cortical structures.

Global aphasia

Extensive brain damage to the left IFG, STG, and arcuate fasciculus may cause the most severe form of aphasia, **global aphasia**. Patients experience both expressive and receptive deficits, usually only being able to communicate using only single words or grunts. They also struggle with repeating words spoken to them. Following a major stroke to the left middle cerebral artery, global aphasia may first present, possibly lessening in severity as the brain heals.

If their other hemisphere is spared, patients with global aphasia can learn to communicate using pantomime or facial expressions.

The Wernicke-Geschwind model

From case studies of injuries leading to aphasia, a few cortical structures emerge as being major contributors to language: the IFG, STG, and the arcuate fasciculus that connects the two. Two neurologists, Carl Wernicke and later Norman Geschwind, proposed the **Wernicke-Geschwind model**, which suggests that information is passed along through language structures in a linear pathway, and each section is responsible for a different aspect of language.

The model begins with the simple scenario: An interviewer is asking you a question, and you answer. First, the sound information arrives into A1, the primary auditor y cortex (see chapter 8 for more details). From there, that information is processed by the STG (Wernicke's area), which then takes meaning out of those sounds. Then, that information travels across the arcuate fasciculus. Then, that information arrives at the IFG (Broca's area), where neurons carry information related to the planning of language, such as coordinating the muscle movements that create the verbal response. Finally, those signals arrive at the motor cortex, which is then responsible for sending the descending signals to control the muscles required for speech (chapter 10).

Another component of the model proposes an explanation for the following situation: You read

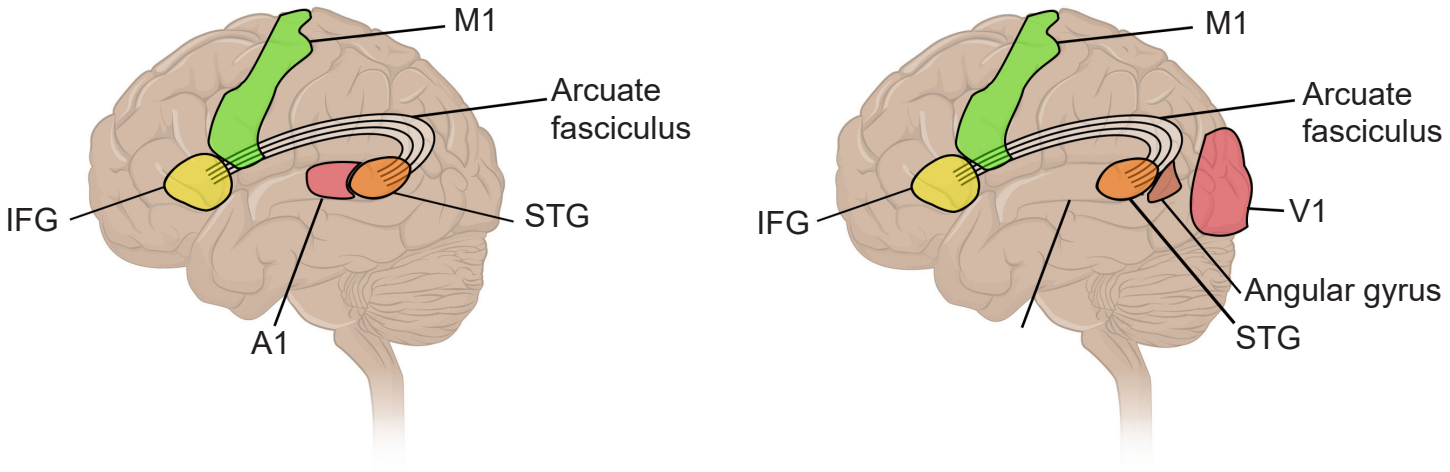


Figure 14.10 The Wernicke-Geschwind model in auditory processing and responding suggests that information signaling arrives into cortex through A1, travels through STG, IFG, then M1 (left). In a reading and responding task, the model suggests that information signaling arrives into cortex through V1, passes through circuits in the angular gyrus, then through STG, IFG, then M1 (right).

a question on a piece of paper, and answer the question verbally. Visual information arrives into the V1, the primary visual cortex. The output of the visual cortices arrives at the angular gyrus, a parietal lobe structure just posterior to the STG. From here, the signal travels through STG and continues through motor cortex, following the same pathway described above.

This Wernicke-Geschwind model was initially helpful for providing a framework for understanding language. But in modern times, we regard it as an oversimplified and outdated explanation of a complex behavior. Sometimes the model fails to accurately predict the nature of a patient's aphasia even if the locus of a lesion has been precisely identified. Furthermore, some injuries to brain areas outside of those structures identified in the model produce aphasia.

Modern research indicates that language functions are not strictly localized as described by the Wernicke-Geschwind model. Instead, language is such a complex behavior, that the interactions between these areas and more are used in language.

Clinical connection: Dyslexia

Affecting an estimated 7-20% of the population, **developmental dyslexia** is a pronounced difficulty with identification of phonemes in printed words and a related difficulty with reading unfamiliar words. Challenges appear in preschool, when learning to decode phonemes is an expected developmental milestone. These difficulties are not a result of intellectual disability. However, dyslexia is not explicitly a language disorder, since patients generally have no difficulties with comprehension of spoken word.

Genetic factors contribute to risk, but a definitive neural mechanism behind dyslexia is still unknown. There are differences in the anatomy and activity of the cerebellum and some atypical lateralization in temporal, parietal, and occipital lobes, suggesting that perhaps some atypical communication from V1 to the language areas of the brain or memory of previously-learned words contribute to symptoms.

Cover art: https://pixabay.com/photos/call-calling-cell-cellphone-15828/
14.1 https://upload.wikimedia.org/wikipedia/commons/d/df/Gray_718-emphasizing-corpus-callosum.png modified by Austin Lim
https://upload.wikimedia.org/wikipedia/commons/3/3b/Gray_733-emphasizing-corpus-callosum.png
14.2 https://pixabay.com/vectors/baby-cat-face-cute-pet-smile-47809/ modified by Austin Lim
https://commons.wikimedia.org/wiki/File:Pirate_patch.jpg modified by Austin Lim
14.3 https://upload.wikimedia.org/wikipedia/commons/3/3b/SplitBrainExperiments.jpg modified by Austin Lim
https://pixabay.com/vectors/bud-donuts-the-cake-cakes-2013344/
https://pixabay.com/vectors/spoon-dishes-food-breakfast-1300507/
14.4 https://pixabay.com/vectors/hand-counting-fingers-one-two-162127/ modified by Austin Lim
https://pixabay.com/vectors/brain-human-anatomy-head-1531009/
https://pixabay.com/illustrations/apple-fruit-red-ripe-fresh-5967971/
14.5 https://commons.wikimedia.org/wiki/File:Cerebral_vascular_territories.jpg modified by Austin Lim
https://upload.wikimedia.org/wikipedia/commons/1/11/1-s2.0-S0967586810002766-gr2.jpg
14.6 Wegrzyn M, Mertens M, Bien CG, Woermann FG and Labudda K (2019) Quantifying the Confidence in fMRI-Based Language Lateralisation Through Laterality Index Deconstruction. Front. Neurol. 10:655. doi: 10.3389/fneur.2019.00655
14.7 https://upload.wikimedia.org/wikipedia/commons/c/ca/Broca%E2%80%99s_area_-_BA44_and_BA45.png
14.8 https://commons.wikimedia.org/wiki/File:Areawerinicke.jpg
14.9 https://upload.wikimedia.org/wikipedia/commons/2/2a/Arcuate_Fasciculus.jpg
14.10 https://commons.wikimedia.org/wiki/File:1605_Brocas_and_Wernickes_Areas-02.jpg modified by Austin Lim

The Open Neuroscience Initiative is funded by a grant from the Vincentian Endowment Fund of DePaul University.

Chapter 15:

Emotion

Austin Lim, PhD (DePaul University)

Editor: Helena Ledmyr, PhD, INCF (Karolinska Institutet)

At the airport, you can observe people experiencing a wide range of emotions: Sadness at seeing family members off, fear and anxiety for those about to fly for the first time, love when a long-distance relationship is reunited, and anger over unpredictable cancellations.

Emotions are complex neurophysiological states that contribute to an internal feeling and guide behavior. Some emotions are pleasant (joy), some are negative (disgust), and some are a mix of both (nostalgia). Some are short-lasting (surprise) while others may persist over years (vengefulness).

However, beyond this statement, it becomes very difficult to put a clear-cut definition on "emotion". The difficulty with defining emotion arises because of the fluid nature of emotions: they exist on a spectrum, multiple emotions are experienced simultaneously, each emotion is perceived by different people in unique ways, and everyone has a slightly different interpretation and understanding of an emotion.

The field of **affective neuroscience** seeks to understand the neural mechanisms that underlie emotion. The field has expanded with the help of functional imaging methods like EEG and fMRI, where changes in brain activity can be measured and quantified as a person experiences different emotion-provoking stimuli. Affective neuroscientists work to develop biology-supported therapies for disorders such as depression, PTSD, and addiction, which are dysregulations of the emotions of sadness, fear, and desire, respectively.

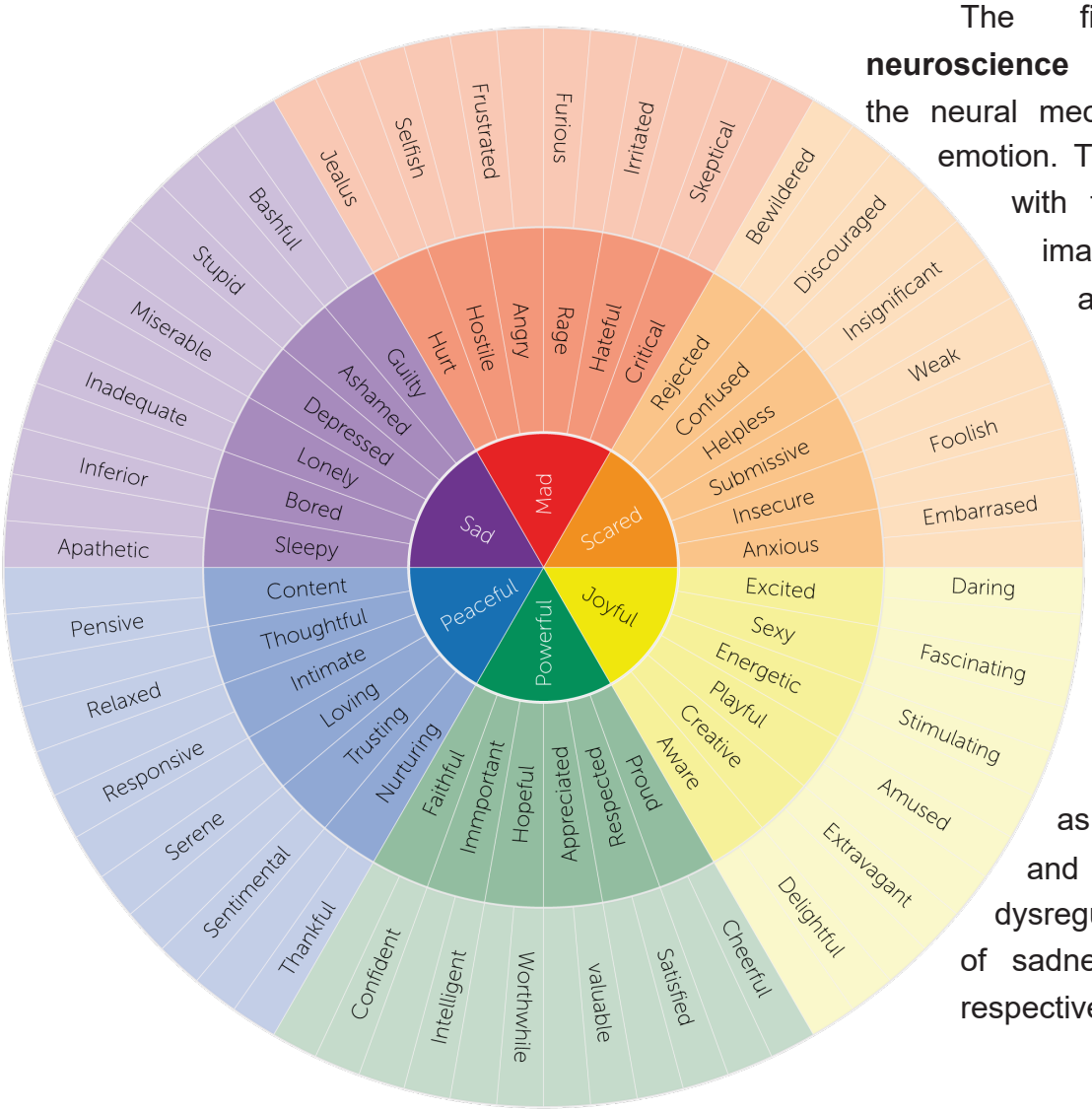


Figure 15.1 An emotion wheel listing several related subtly different feelings.

Chapter 15 outline

15.1 A History of Emotion Research

Origins of emotion

One of the older theories about the origin of emotion is based on the most common sense interpretation of cause and effect. For example, imagine some noticeable emotional stimulus, such as encountering a hungry lion on the sidewalk. The logical cause and effect explanation suggests that seeing the lion prompts the emotion of fear, which then causes the sympathetic nervous system "fight-or-flight" response (elevated heart rate and blood pressure, increased respiration, and cellular mobilization of energy).

In the 1880s, psychologist William James and physician Carl Lange independently developed a new theory about the origin of emotion. According to the **James-Lange theory of emotion**, and contrary to a common sense understanding of the origin of emotion, the body's physiological changes precede the onset of an emotional response. For example, imagine encountering that same hungry lion on the sidewalk. The James-Lange theory tells us that the perception of the threat of being eaten causes the sympathetic nervous system response, and that these physiological changes trigger the onset of fear.

Soon after, in the 1920s and 30s, two physiologists named Walter Cannon and his doctoral student Philip Bard criticized the James-

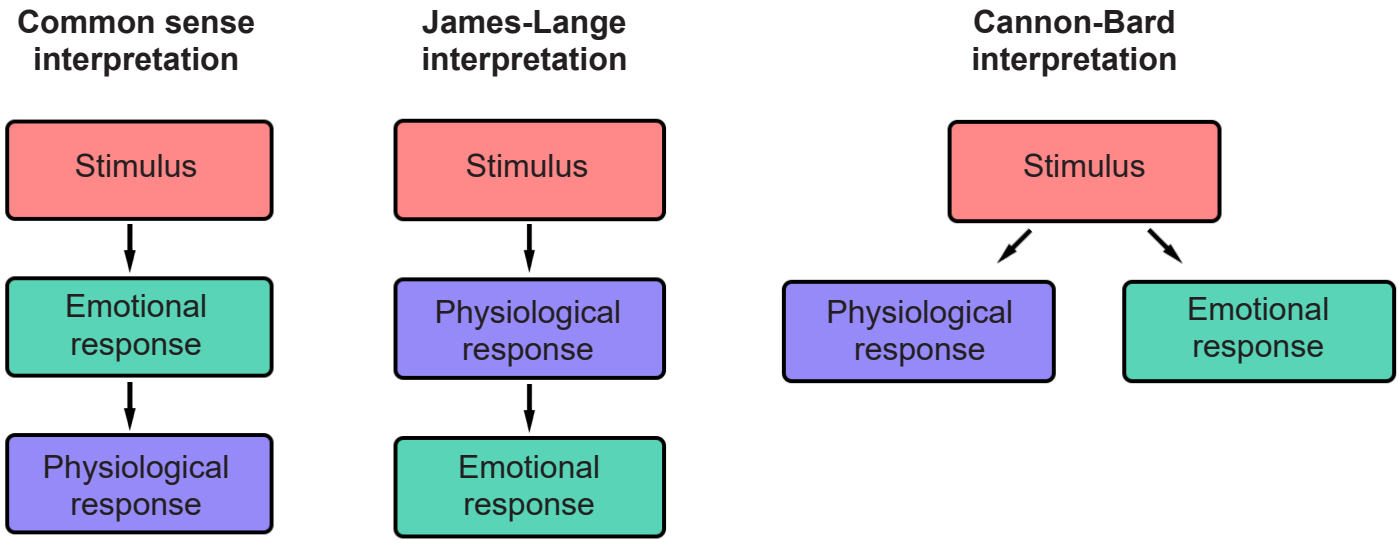


Figure 15.2 Three different theories about how an organism responds to some stimulus.

Lange theory. In one experiment, they surgically removed the entire sympathetic nervous system from cats, destroying the nerves that regulate vascular dilation, the activity of liver enzymes, and the reaction that causes the hair standing on end. These cats were then put before a threatening aggressor. If the James-Lange theory was true, then the physiological changes should precede the emotive response. However, the cats exhibited the fear / aggression response (such as posturing, hissing, and clawing) even without an intact sympathetic nervous system. Relatedly, patients with spinal cord injuries have a similar lack of autonomic inputs to the brain, but their capacity for emotional responding is still intact.

In a second criticism, Cannon and Bard proposed that the physiological changes seen in sympathetic nervous system activity may arise for a variety of reasons, not always for emotionally salient reasons. For example, intense exercise causes strong cardiorespiratory changes; however, we do not necessarily feel a strong emotional state after this physiological perturbation. Likewise, exogenous administration of epinephrine, onset of fever, or being in cold temperatures may also trigger some physiological changes without causing a strong emotional response.

Figure 15.3 Contrary to the James-Lange theory of emotion, physical exercise produces a high physiological state without always inducing an emotional response.

Based on their evidence opposing the James-Lange theory, Cannon and Bard developed an alternative explanation for the origin of emotions. According to the Cannon-Bard theory of emotion, the perception of an emotionally charged stimulus prompts simultaneous but independent activation of both the autonomic nervous system and the emotional response.

The Cannon-Bard theory also draws attention to the neuroanatomical structures that trigger the autonomic and emotional responses, with a particular focus on diencephalon structures such as **hypothalamus** and **thalamus**. To localize the anatomy behind emotional signaling, researchers performed a series of lesion experiments, systematically injuring different parts of the brain. To their surprise, when the cortex of the cat was surgically separated from the rest of the nervous system, a procedure

Figure 15.4 Decorticated cats express sham rage in response to harmless stimuli, suggesting that anger is kept under control by cortical inhibition.

called the **decorticate preparation**, the cats would exhibit a hyper-aggressive response to stimuli. For example, an innocuous touch of the tail would trigger violent clawing, biting, and hissing, behaviors which were described as **sham rage**. However, when they lesioned the thalamus, sham rage was no longer observed. They concluded that rage (and other powerful emotions) is normally under inhibitory control by the cortex.

Just years later, in 1937, American neuroanatomist James Papez (pronounced payps) ascribed emotional behavior to a particular series of brain structures. These structures, collectively called the **Papez circuit**, consist of the hypothalamus, cingulate gyrus, thalamus, hippocampus, and more. Papez observed unusual aggression among animals with injury to these structures, suggesting that emotional responding is distributed across many areas, rather than localized. (Notably missing from the Papez circuit structures is the amygdala, which was later added in a future revision in the 1950's.)

In 1939, two researchers named Heinrich Kluver and Paul Bucy described a unique set of emotional deficits in monkeys with bilateral temporal lobe removal, further providing evidence of the neuroanatomy of emotion. These animals fail to display fear or anger, even in the face of life-threatening stimuli such as a large snake. They also display **visual agnosia** (the inability to recognize faces or objects visually), psychic blindness, hypersexuality, and **hyperorality** (an inappropriate fixation with using the mouth to interact with surroundings, such as licking or eating nonfoods). Collectively, these sets of symptoms are called **Kluver-Bucy syndrome**. Although it is primarily an experimental manipulation observed in monkeys, some human patients may develop the condition, such in after brain surgery, stroke, viral encephalitis, traumatic brain injury, and many others.

A new angle to the origin of emotion was proposed by Stanley Schachter and Jerome E. Singer in 1962. They were interested in how the same physiological response can be attached to

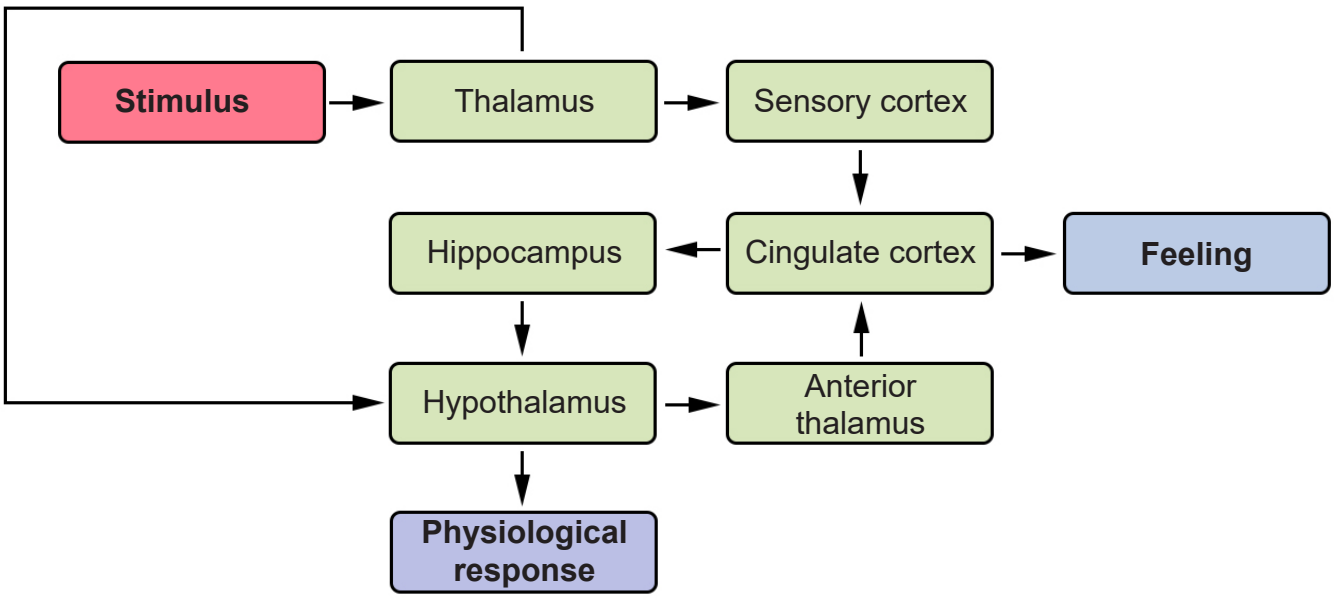


Figure 15.5 The Papez circuit was an early neuroanatomical description of some structures (green) involved in emotional processing and responding (purple).

Figure 15.6 In addition to a loss of fear, a macaque monkey with Kluver-Bucy syndrome would display hyperorality.

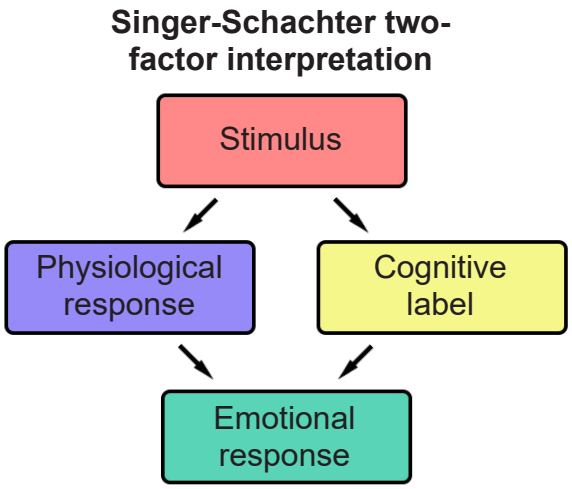


Figure 15.7 The Singer-Schachter two-factor theory of emotion suggests that both the physiological response and a cognitive label contributes to the emotional response.

two wildly different emotions - consider the rise in heart rate and respiration that are associated with encountering that lion (fear), or when you receive wonderful news (elation), or when you make eye contact with your romantic partner (love). According to their **two-factor theory of emotion**, people use a combination of the physiological response and a cognitive label to determine the emotion that is most appropriate for a given circumstance. The cognitive label comes from two parts. One is prior knowledge, such as the thought processes "lions eat meat if they are hungry, so I should be afraid." The other part is observing environmental cues, such as "everyone around me is running away and screaming, so I should also be afraid."

To test their theory, Schachter and Singer administered epinephrine to patients. Epinephrine is a hormone that produces the physiological signs very similar to those observed in a sympathetic nervous system response: elevation of heart rate, respiration, blood flow to muscles, and energy utilization. Then, the patient is put into a waiting room with another "patient" - in actuality an undercover member of the research team (the confederate). For one set of patients, the confederate would act "euphorically," jokingly playing trash-can basketball, making paper airplanes, and hula-hooping, all the while exclaiming how much fun they are having, inviting the patient to play along. For another set of patients, the confederate would act in anger, expressing irritation before eventually ripping up the survey and storming out of the study.

When the patients were not told what to expect from the epinephrine, they were more sensitive to the emotional responses of the confederate. This demonstrated that environmental cues play a significant role in determining emotion regardless of physiological state.

Faces in Emotion

Although best known for his theory on evolution, naturalist Charles Darwin published prolifically about other topics in biology, ranging from botany, coral reefs, and even a treatise on psychology. In this 1872 text, called *The Expression of Emotions in Man and Animals*,

Figure 15.8 Charles Darwin suggested that emotional facial expressions are useful for conveying survival cues, such as evoking prosocial behaviors towards sad people (top) or avoiding angry animals (bottom).

Darwin suggested that similar emotional responding is found across different cultures, and to some extent, even in nonhumans. In his view, the main purpose of emotive expression is to communicate survival cues between individuals: a relaxed expression conveys safety, while a fearful expression promotes alertness, since danger may be nearby. Darwin also suggested that we gain survival information from non-human behaviors, for example, a hissing snake or a snarling lion is an immediate threat that should cause fear or other avoidance behaviors.

Moving forward into more modern times, the American psychologist Paul Ekman expanded on Darwin's theory, distilling down the range of human emotions to belonging in one of seven basic categories of emotions: Anger, contempt, disgust, fear, happiness, sadness, and surprise. If the purpose of emotion was to communicate pro-survival cues, then Ekman theorized that all humans, regardless of culture, would use similar facial expressions. To test this hypothesis, Ekman visited a remote village in Papua New Guinea, where he studied a population that is isolated from any other known cultures. As predicted, these people made the same facial responses in reaction to various emotional circumstances. In 1972, Ekman published his theory of **universal facial expressions**.

Part of Ekman's research led their team to develop a series of photographs of actors portraying six major emotions. This **Ekman 60 faces (EK-60F) test** has since been used to assess facial emotional with fascinating results: People with major depressive disorder or borderline personality disorder have a lessened ability to detect happiness in others, seeing emotional faces results in a similar emotion in the viewer, and people with dementia or Parkinson's disease identify emotions as being less intense.

Ekman's research led them to develop the **Facial Action Coding System (FACS)**, a system that uses facial anatomy to differentiate the features that are characteristic of different expressions. For example, a feature of a happy face is the flexing of the zygomaticus major and orbicularis oculi muscles, which produces an upward turn of the corners of the mouth and a rising of the cheeks. Other facial features, such as head movement, eye movement, and larger physical

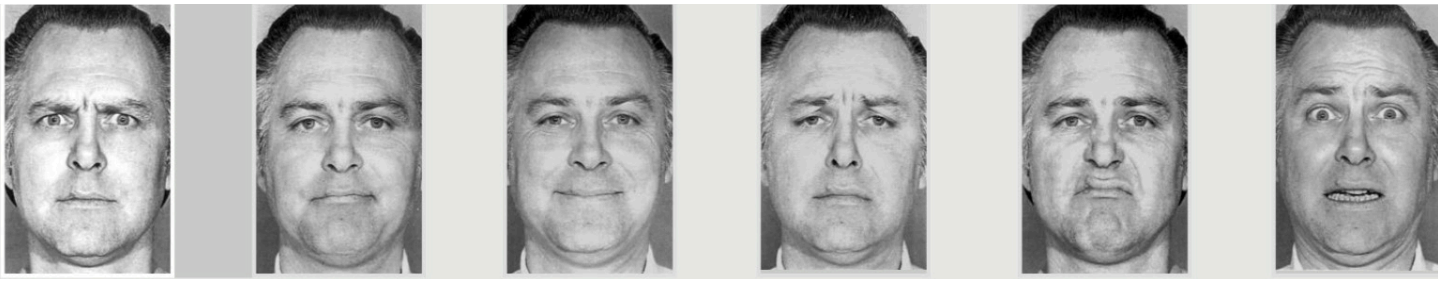


Figure 15.9 Paul Ekman's research in Papua New Guinea suggests that across cultures, humans use similar facial muscle activity patterns to convey a universal set of emotions. Ekman also developed anatomical definitions for describing specific emotions.

movements are also scored, and are also used to help identify emotions. The FACS can be used to formally describe why some smiles appear as genuine (a Duchenne smile, with simultaneous muscle action) while others look fake or forced (a non-Duchenne smile, characterized by the turn of the corners of the mouth without much change to the top part of the face).

15.2 Structures Involved in Emotion

Emotions are one of the most holistic functions of the brain, incorporating neural circuitry from across several different structures, ranging from the phylogenetically older to the most advanced frontal cortical areas. The structures of the Papez circuit are not a comprehensive list of brain structures involved in emotional processing but offer a good starting place for describing the anatomy of emotion.

Amygdala

A limbic structure called the **amygdala** contributes heavily to processing the valence of emotional experiences. Roughly the shape and size of an almond, the amygdala is part of the temporal lobe. Generally, the amygdala is subdivided into several nuclei, including the basolateral amygdala, central nucleus, and cortical nucleus.

As a temporal lobe structure, amygdala is strongly implicated in emotional memory formation. Emotional memories can be either positive valence (such as the happiness you may have experienced at a birthday party when younger) or negative valence (such as childhood trauma or being teased as a child).

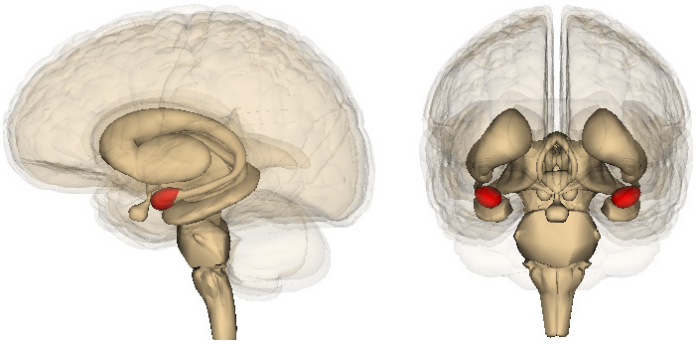


Figure 15.10 The amygdala (red) contribute to emotional processing.

Of the forms of declarative memory (chapter 13), autobiographical memories more often have emotional content compared to semantic memories.

Amygdala lesions have been used as a last resort treatment for patients with temporal lobe epilepsy or psychiatric conditions with pathological and dangerous aggressiveness. These psychosurgery strategies have been variably successful, but they often have high complication rates and upwards of a 4% mortality rate. These treatments are rarely used today. Deep brain stimulation may offer a less intrusive and therefore less risky therapeutic approach.

Monkeys with amygdala lesions made up the earliest descriptions of the emotional disruption condition later described as Kluver-Bucy syndrome. These monkeys did not exhibit the traditional fear response when presented with the hissing of a large snake. They also showed an absence of anger, such as when the animal is exposed to provocative novel stimuli.

A useful nonhuman model for studying emotional learning is the **fear conditioning paradigm**. In this test, a rodent is put into a cage. Occasionally, an innocuous tone and light would activate. Shortly after, an unpleasant foot-shock would be delivered to the animal. Over time, the healthy animal learns that the tone and light precede the shock to the foot, and upon future presentation of these stimuli, they will freeze. If the amygdala is lesioned, however, they freeze less, meaning they do not acquire the emotional memory associated with the stimuli. This paradigm is used as a model for **post-traumatic stress disorder (PTSD),** a psychiatric condition associated with the recurring recall of negative memories.

Hypothalamus

One of the major output signaling pathways of the amygdala is the **hypothalamus**, an almond-shaped nervous system structure found at the base of the brain. The hypothalamus is often seen as the neural structure that initiates endocrine responses in the rest of the body, such as hormone production. Several different behaviors, ranging from homeostasis, hunger, and circadian regulation are modulated by hypothalamic signals.

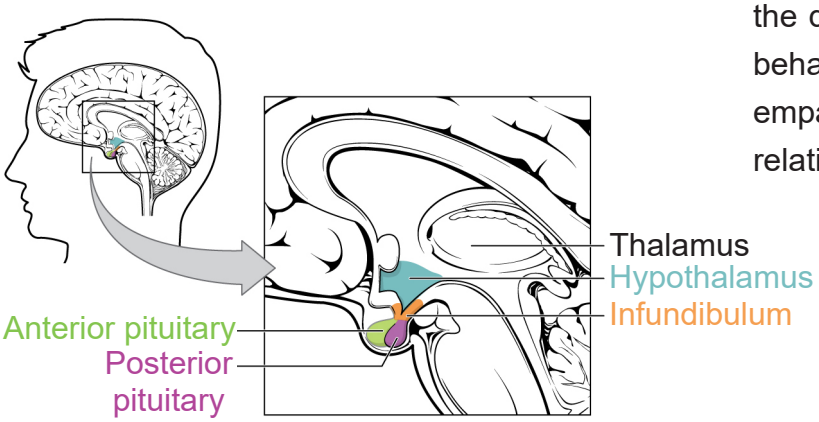


Figure 15.11 The hypothalamus is a brain structure that initiates hormonal changes through influencing the pituitary gland.

Pituitary gland

Downstream from the hypothalamus is the **pituitary gland**, a pea-sized endocrine organ that protrudes from the base of the brain. It is strongly involved in the production and release of **neurohormones**, signaling molecules produced by nerve cells that travel throughout the bloodstream to influence the activity of several organs throughout the entire body. The pituitary gland is subdivided into two regions with distinct anatomical and functional differences.

Posterior pituitary gland

The **posterior pituitary gland** (also called the **neurohypophesis**) does not synthesize any neurohormones. Instead, the axonal projections of the magnocellular neurosecretory cells of the hypothalamus run through the posterior pituitary, where the hormones are secreted into a special anatomical feature called the **hypophyseal portal system**, a series of leaky capillaries densely wrapped around the area, allowing for the hormones to easily diffuse into the bloodstream. There are predominantly two main hormones released at the posterior pituitary.

Oxytocin (OT) plays a significant role in the development and maintenance of **prosocial** behaviors, acts such as trust, compassion, and empathy, all actions that enhance interpersonal relationships. For example, increased blood levels of OT is seen in new couples compared to unattached singles, and OT release happens during orgasm, which may contribute to romantic attachment. OT signaling increases dramatically during childbirth and triggers milk letdown in lactation, which strengthens the mother-child relationship. Interestingly, while OT generally strengthens the social bonds between people, it promotes antisocial behaviors against those not perceived to be within one's own social group.

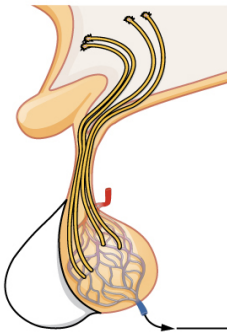
Disorders of the OT system are believed to contribute to autism spectrum disorder and psychopathy, two complex conditions characterized partly by social impairment. Some studies have examined the therapeutic use of nasal OT for a variety of psychiatric conditions, but the studies have been unable to demonstrate strong clinical effects despite success in non-human animal models.

Vasopressin (**AVP**; or **antidiuretic hormone, ADH**) like OT, also contributes to social behaviors. Biochemically, AVP is very similar to OT, as they are both nine amino acid peptides

differing by only two residues. AVP additionally regulates osmolarity, increasing water retention by the kidneys, which returns the body from a hypertonic state back to homeostasis. AVP also constricts blood vessels, raising blood pressure. AVP is also used therapeutically to raise blood pressure for patients in shock and to treat **diabetes insipidus**, a fluid dysregulation disorder (unrelated to diabetes mellitus, the much more common blood glucose disease).

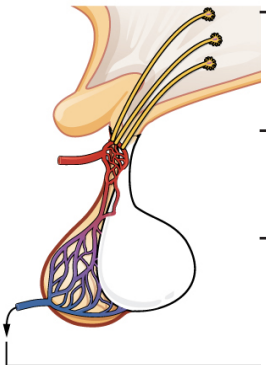
Anterior pituitary

The **anterior pituitary** (or the **adenohypophesis**) is capable of both synthesizing and secreting a variety of neurohormones, which contribute to several functions such as the stress response, growth, sexual development, circadian rhythms, and more. These hormones are collectively called the **trophic hormones**. The neurosecretory cells of the anterior pituitary are sensitive to signals from the hypothalamus.



Posterior pituitary hormones

Releasing hormone (from hypothalamus)	Pituitary hormone	Target	Effect
	OT	Breasts Uterus	Milk letdown Contractions
ADH		Kidneys Circulatory system	Increased water retention Constricts blood vessels



Anterior pituitary hormones

GHRH	GH	Somatic tissue	Promotes anabolism
CRH	ACTH	Adrenal glands	Glucocorticoid production (stress response)
GnRH	LH	Reproductive system	Sex hormone production
	FSH	Reproductive system	Production of sperm and eggs
Thyroid releasing hormone	Thyroid stimulating hormone	Thyroid gland	Promotes metabolism
	Prolactin	Mammary glands	Milk production

Figure 15.12 The hypothalamus sends a variety of hormonal signals, which influence several different physiological features.

Within the anterior pituitary are cells called somatotrophs, which produce and secrete **growth hormone (GH)**. GH is a signal that promotes **anabolism**, the buildup of larger molecules through biochemical reactions. Anabolic processes increase cellular repair and protein synthesis, which promote growth. Most of the time in the day, GH production is at a steady low level, but it increases at specific events, such as after meals or during slow wave sleep. The biochemical signal from the hypothalamus is **growth-hormone-releasing hormone (GHRH)**.

The anterior pituitary releases **adrenocorticotropic hormone** (or **ACTH**). Once ACTH is released into the bloodstream, it travels to the **adrenal cortices**, a pair of organs that sit on the anterior surface of the kidneys. Once there, ACTH triggers production of **cortisol**, a glucocorticoid hormone. Cortisol is best known for its initiation of the stress response, which is characterized by mild sympathetic nervous system activity. ACTH is synthesized downstream from the hypothalamic signal **corticotropin releasing hormone (CRH)**. The series of organs that result in the stress response is called the **hypothalamic-pituitary-adrenal axis**, or the **HPA axis** for short.

The anterior pituitary also synthesizes and releases the **gonadotropins**, hormones that are important for regulation of puberty, sperm / egg production, the release of sex hormones by the testes / ovaries, and menopause. The two main gonadotropins in humans are **luteinizing hormone (LH)** and **follicle-stimulating hormone (FSH)**. These are produced in response to the hypothalamic signal **gonadotropin-releasing hormone**. This signaling cascade is called the **hypothalamic-pituitary-gonadal axis**, or the **HPG axis**.

Insula

The **insula** (or **insular cortex**) is the lobe of the cortex buried deep within the lateral fissure. Although not visible from a side view, the insula is often considered to be the fifth lobe of the telencephalon. Early studies by Wilder Penfield where he directly stimulated the brain during open brain surgery led researchers to suggest that the insula contributes to **interoception**, detecting the internal state of the body and conveying that information for processing.

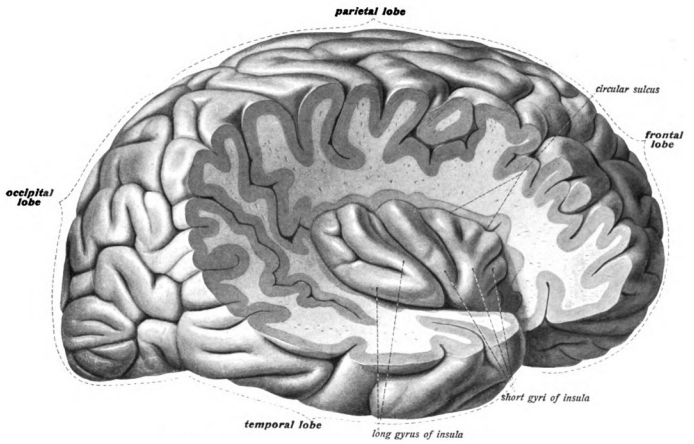


Figure 15.13 The insular cortex is only visible in a cut-away view of the brain.

In functional imaging studies, the insula is involved in the recall or many different emotional stimuli, especially those emotions that have a sensory component.

Notably, the insula is strongly implicated in the emotion disgust. For example, a patient is placed in an fMRI scanner while breathing through a mask, which allows the experimenters to change the smells that are perceived. Patients are then given pleasant smells (such as passion fruit, pear, or mint), a neutral smell, or unpleasant smells (like ethyl-mercaptan or isovaleric acid, which smells like skunk or body odor, respectively). There is increased activity of the anterior insula in response to the unpleasant smells but not the pleasant smells.

The insula also responds to social cues related to disgust as well. When a patient in an fMRI sees a video of a person smelling something unpleasant and reacting with a “repulsed” face (the closing of the nostrils and curling of the upper lip), their anterior insula likewise increases in activity just as if they had smelled it themselves.

In addition to sensory stimuli, feelings of social repugnance (unwarranted violence, murder) or moral disgust (incest) also increase insula activity.

Atypical insula activity is implicated in behavioral disorders. For example, insensitivity to disgust can lead to squalor-dwelling conditions (sometimes seen in excessive hoarding or late cognitive decline), which puts those people at heightened health risks due to regular exposure to unsanitary conditions. Substance use disorders, PTSD, and suicide attempts have all been associated with atypical insula activity.

15.3 Specific Emotions

Fear / Anger

Nearly everyone has experienced the prototypical fear response: Imagine reading a book when you see a spider skittering along the wall. Suddenly, you'll feel your heart racing, your breathing increase, dryness of your mouth, and your palms sweating. You probably won't notice the dilation of your pupils or the change in liver activity and digestion. This sympathetic nervous system activity is downstream of being presented with a fearful stimulus.

But upon closer inspection, it was no spider at all - just an errant piece of brown fuzz picked up by a draft. Within minutes, your body's physiology returns back to normal.

This anecdote points out a few important features about the fear response. First, the onset of the fear response is quick, and so is the dissipation of the fear. Second, it is triggered by exposure to a perceived threat, regardless of whether the stimulus is a genuine threat or not (the overwhelming majority of spiders are clinically harmless to humans!). Third, the fear response is greatly modified by knowledge and experience - an entomologist would recognize that the spider is a harmless house spider and would, instead of fear, display curiosity, interest, boredom, or other emotions. On the other hand, someone who has been bitten by a dangerous spider and sent to the hospital when younger would have a much stronger physiological response.

Fear is likely the most evolutionarily ancient emotion, and is highly protective. When encountering a hungry mountain lion, faces displaying the traits of fear (enlarged eyes, flared nostrils, and a slightly open mouth accompanying a gasp) would signal to others nearby that a threat is nearby, which helps initiate heightened alertness and the appropriate fight-or-flight response.

Figure 15.14 The fear response is likely an evolutionarily protective behavior, since it provides social cues about dangers.

Patient SM

Patient SM is a notable case study of a person who does not experience the fear response. Born with an extremely rare genetic condition called **Urbach-Wiethe disease**, Patient SM progressively developed calcification in her amygdala bilaterally, causing destruction and cell death. Like other people with Urbach-Wiethe disease, she had no severe significant cognitive deficits, except for the inability to experience fear.

In one test, she was shown a variety of emotionally charged videos, then asked to rate the intensity of each clip with respect to different emotions. SM found clips from America's Funniest Home Videos to be just as funny as

the control patients, and she found the clips of disgusting toilets to be just as repulsive. But, when presented with clips depicting ghost hauntings or suspenseful serial killers on the loose, SM rated these stimuli as being non fearful.

Other studies challenged Patient SM with more concrete threats. Researchers brought SM to a pet store, where she asked to handle the snakes. She was stopped by an employee before she put her hand into a tarantula cage out of curiosity. The researchers also took her to the Waverly Hills Sanatorium haunted house,

Figure 15.15 Patient SM exhibit no physiological fear response following presentation of typically scary stimuli, such as snakes or a haunted house.

Clinical connection: Plasmatoxmosis

Plasmatoxmosis is a disease caused by infection of the parasite *Toxoplasma gondii*. Plasmatoxmosis may cause symptoms characteristic of a generic immune response, such as fatigue, body aches, and fever. Some people may be completely asymptomatic; however, others experience severe side effects like seizures or birth defects when an infection happens during pregnancy.

T. gondii is only capable of reproducing within cats, but the parasite has developed a very unusual evolutionary adaptation that permits propagation of the species, called the **manipulation hypothesis**. Cats excrete *T. gondii* via feces, which is eaten by rats and mice, where the parasites take up residence in their temporary host. *T. gondii* causes modifications in amygdala neurons, changing the rodent fear response, in turn making them attracted to cat-related odors rather than fearful of them. The fearless rodents are therefore less likely to avoid a hungry cat, increasing their risk of being eaten, which begins the cycle again.

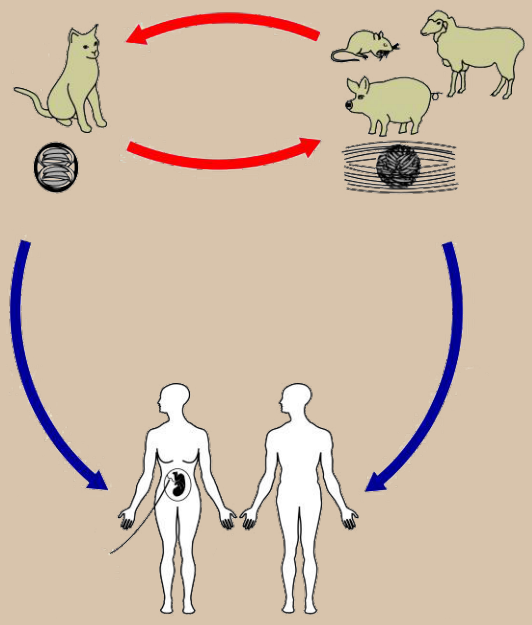


Figure 15.16 A toxoplasma infection changes the rodent fear response, causing them to be less afraid of predators. At either point, plasmatoxmosis can infect humans with potentially severe outcomes during pregnancy.

Some research suggests that a *T. gondii* infection subtly affects human behavior in cat owners, hinting that *T. gondii* infection is associated in suicides and the onset of schizophrenia.

where she bravely led a group of strangers through the house populated by actors dressed as monsters and ghosts. Although she did not display any fearful behaviors, such as hesitation to walk through the darkened corridors, patient SM reported the sensation of exhilaration and enthusiasm, akin to riding a roller coaster.

She also was asked to recall some of her past real-life, fear-provoking experiences, such as when she was attacked in a domestic violence incident or was held up by a stranger at knife point in a public park. In none of these cases did she ever report feeling fear, although she was upset and angered at the situation. The destruction of her amygdala seemed to make her resilient against PTSD: the day after being threatened with a knife to her throat, she walked past the very same park bench.

Fear and anger are two very closely related emotions. As with many other emotions, anger manifests in complex ways. The anger spectrum runs from low (irritation) to high (rage), and from quick (lashing out) to persistent (vengeful). Strong anger can provoke anti-social behavior such as violence.

Since anger may precede violence, angry faces function as a warning during evolutionary survival situations or during interpersonal conflict. Seeing angry faces may prompt awareness, initiating the sympathetic nervous system response in case “fight or flight” becomes necessary. As with fear, the emotional response to angry faces or body postures is partially guided by the amygdala, resulting in communication with the rest of the body through cortisol release downstream of the HPA axis.

Environmental factors such as childhood maltreatment or expectations are partially responsible for how a person responds to a particular anger-provoking stimulus. Internal homeostatic conditions also influence the anger response, just as how low blood sugar increases aggression and drives negative or hateful feelings in the face of a challenge (the portmanteau “hangry” was added to the Oxford Dictionary in 2018).

Neurobiological factors also contribute to the anger response. In addition to amygdala circuits, regions in the frontal cortex decrease activity during acts of aggression, suggesting that

Clinical connection: Borderline Personality Disorder (BPD)

Borderline Personality Disorder (BPD) is an emotional dysregulation disorder characterized by disproportionately intense emotional reactivity to environmental triggers. For instance, they may experience anger or hostility in response to minor inconveniences. These emotional changes may persist for unusually long durations, up to hours or days, which often puts strain on interpersonal relationships. They also experience a fear of abandonment, an unstable sense of self, suicidal ideation or attempts, and dissociation (an emotional detachment from events).

It is estimated that 1.4% of people have BPD. It is often comorbid with other psychiatric disorders, such as anxiety, depression, or substance use disorders. The main treatment approach is psychotherapy; medications are not always effective at reversing the symptoms.

Celebrities such as Saturday Night Live actor Pete Davidson, musician Amy Winehouse, and Princess Diana likely had BPD. Arguably, the fictional character Anakin Skywalker from the Star Wars prequel films also displays many diagnostic traits of BPD.

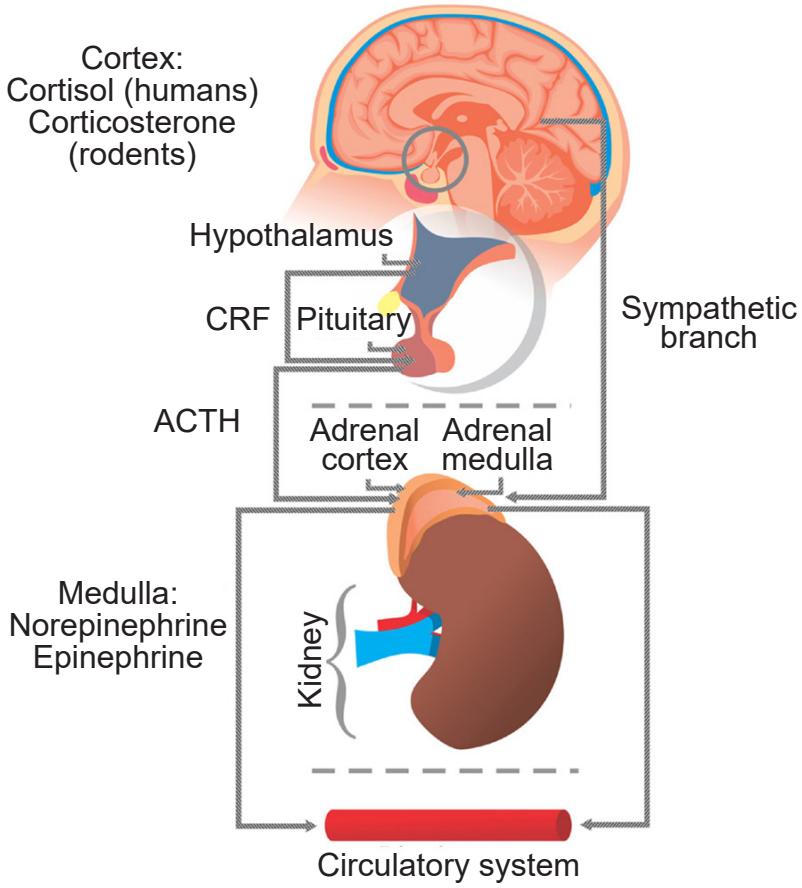


Figure 15.17 The HPA axis regulates both sympathetic nervous system activity and the stress response.

frontal circuits actively inhibit the limbic system, which drive our more "primitive" responses. Altered frontal cortical action may therefore account for one reason why two different people would react to the same anger-provoking stimulus in different ways.

Stress

Stress is a natural and healthy bodily response that is similar to the prototypical "fight-or-flight" response: elevated heart rate and blood pressure, increased respiration rate (shortness of breath), energy usage, and perspiration, among others. Physiologically, stress also causes a weakening of the immune response, a slowing of growth, and decreases in kidney filtration.

Stress can either result from purely physical stimuli (exposure to constant cold; chronic dehydration), a purely psychological threat (resentment; religious guilt), or a combination of both (living paycheck to paycheck).

The human stress hormone cortisol triggers the physiological response to stress. **Cortisol** is a signaling molecule synthesized from cholesterol which readily passes through cell membranes where it can activate intracellular **glucocorticoid receptors (GRs)**. Once the ligand-receptor complex is formed, it translocates into the nucleus, where it functions as a transcriptional regulator. Nearly every cell contains GRs, meaning that an elevation of systemic cortisol can influence several organs to change their activity.

A small amount of stress, sometimes called **eustress**, is beneficial and generally performance-enhancing, such as in the moments leading up to the beginning of an athletic competition. The opposite, **distress**, is characterized by the inability to cope with rising demands, leading to increased anxiety and other maladaptive responses. Distress is the kind of stress that people experience when they are facing multiple rapidly-approaching deadlines, difficulties in paying rent, or relationship issues.

Biochemically and physiologically, eustress is the same as distress. The major difference is the psychological interpretation or perception of the stressor. Eustress is seen as a challenge which someone can work to overcome and is often short lived or temporary. Chronic long-term distress causes all variety of negative health outcomes, such as cardiovascular

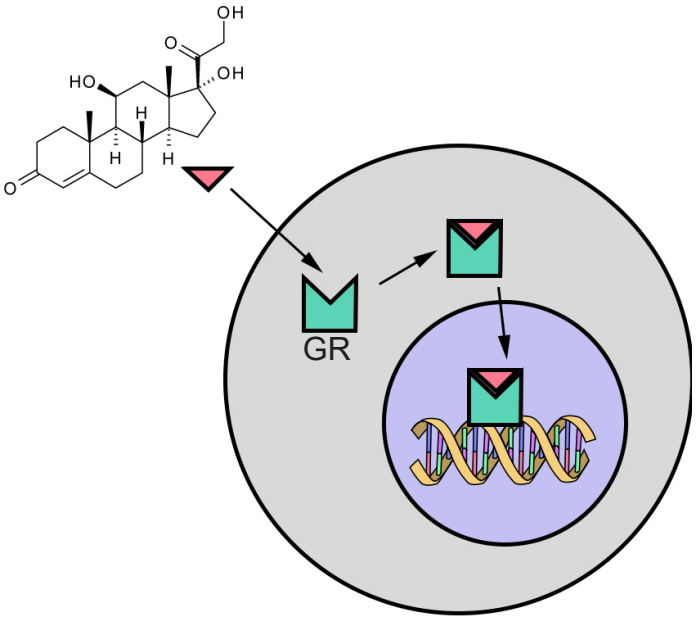


Figure 15.18 A molecule of cortisol (red triangle) diffuses through the cell membrane and binds to intracellular glucocorticoid receptors (GR, green).

disease, diabetes, and obesity, all risk factors for premature death.

The early medical descriptions of stress were put forth by Hans Selye in the 1930's. As a medical student, he observed that chronically-ill humans, regardless of their specific condition, often experienced a similar set of late-stage negative health outcomes, such as gastric ulcers, high blood pressure, and heart attacks. Collectively, he called this response the **general adaptation syndrome**.

Selye observed that extracts from different internal organs led to these health changes in rats, eventually leading to premature death. As a good scientist, Selye also injected a second group of rats with a harmless saline solution as a control. Surprisingly, Selye observed that

Clinical connection: Ulcers

Ulcers are injuries to the mucous membranes of the stomach or intestines resulting in abdominal pain, nausea, and changes in appetite (although some patients report no symptoms.) In severe cases, ulcers can cause bleeding or perforation of the stomach, which could be fatal. An estimated 4% of people have ulcers, although this is likely an underestimation.

For several years, ulcers were believed to be caused solely by stress. However, later research suggested that colonies of the gut bacterium *Helicobacter pylori* was found to contribute to the formation of 80% of ulcers, a discovery that earned Drs. Barry Marshall and Robin Warren a Nobel Prize in 2005. Today, the evidence suggests that neither factor exclusively leads to ulcer formation: Nearly half of all people have the *H. pylori* colonies

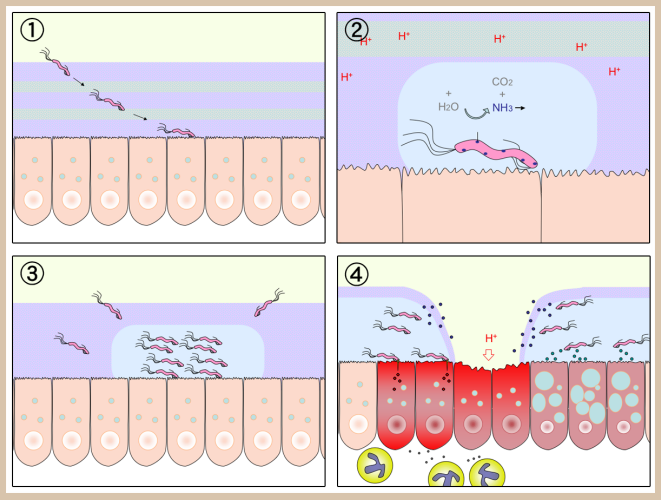


Figure 15.19 H. pylori interacts with stress and may result in stomach ulcers.

in their gut, but not all these people develop ulcers. Also, about 30% of people with ulcers do not harbor *H. pylori*, and people taking oral antibiotics that have eliminated their *H. pylori* colonies can still develop ulcers.

these control animals also exhibited the general adaptation syndrome.

It turns out that even minor stressors, such as the movement of the rat cage, grabbing of the animal, and administering an injection led to an eightfold increase in plasma levels of epinephrine, a hormone upstream of the stress response. Selye concluded that the general adaptation syndrome was not a result of the injections of organ extract, but rather because he induced a tremendous stressor to all the rats.

Like Selye, we use nonhuman models to study how the nervous system changes in response to stress. Rodents can be exposed to **chronic-restraint stress**, where they are put into a narrow cylinder for part of their day. **Food restriction stress** is a rodent model of starvation or dietary malnutrition. The **forced-swim test** is best known as an assessment of depressive behavior, but it can also be used to induce stress. Food restriction and the forced swim test increases plasma levels of the neurohormone **corticosterone** (the rodent analog of cortisol).

Psychological stressors also induce the stress response. Common stress paradigms include exposure to socially-dominant bully rats or predator-associated cues like fox urine, which contains a group of chemicals called kairomones. **Kairomones** cause behavioral changes in other species, whereas pheromones refer to interspecies chemical communication. Social isolation is also a stressor for rodents, and for this reason, rats and mice are normally group housed.

Love

Love can be thought of as an intensely strong attachment towards a person (romantic love, lust), a thing (passion project), or concept (patriotism as a love for country, or altruism as a love for fellow humans). However, it is very difficult to put a strict biological definition on these varied concepts of love.

This section focuses on interpersonal love, of which there are several unique forms, each resulting in different behavioral outcomes. For example, romantic love drives physical attraction, lust, and sexual activity. Parental love, on the other hand, encourages self-sacrifice and hyper-attentiveness towards a newborn. Some behaviors are shared between the two forms of love, such as respect.

Romantic love

Romantic love drives much of human behavior and has been documented thoroughly in the arts all the way from Homer's Iliad through Shakespeare's Romeo and Juliet up to modern Taylor Swift.

The majority of human societies have embraced **social monogamy**, the romantic relationship characterized by a pair of people who share resources, parenting duties, and exhibit preferential mating. Outside of humans however, only 9% of mammalian species form socially monogamous pairs, while at least 75%

Figure 15.20 Unlike mammals, most bird species form socially monogamous bonds after mating.

of bird species maintain socially monogamous relationships (which may explain the origin of the phrase “love birds”).

Dr. Helen Fisher, an anthropologist and a leader in the field of love research, suggests that love can be divided into three closely interconnected components. These three are guided in part by different signaling pathways, and lead to somewhat different behavioral outcomes.

1. Lust

Lust (or **libido**) refers to a very strong desire for sexual gratification. These behaviors are largely driven by the actions of the sex hormones **testosterone**, **estradiol**, and **progesterone**, released downstream of activation of the HPG axis. As hormones, they are synthesized from cholesterol and circulate through the bloodstream to influence the body in many ways.

Both testosterone and estradiol contribute to sex-seeking behaviors in men and women, where increasing testosterone levels drive up sexual desire.

In the brain, the sex hormones strongly influence the **medial preoptic area (mPOA)** of the anterior hypothalamus. The mPOA contains a **sexually dimorphic area**, the part of the brain that exhibits the biggest morphological difference between males and females: in humans, it is about twice as large in males with double the number of neurons throughout childhood and early adulthood. In the rat, it is up to 8 times larger in males than females, and if this area is lesioned, rats exhibit decreased motivation to engage in sex.

The amygdala also plays a significant role in mediating lust, and lesions may either result in hypersexuality (Kluver-Bucy syndrome) or a decrease in responding to socially-derived sex cues.

In addition to changing libido, the sex hormones contribute to sexual differentiation and maturation (mascularization with testosterone, such as muscle mass increases and hair growth), differences in immune system activity (decreases with higher testosterone), and pain tolerance (higher tolerance with higher testosterone) for example. Progesterone is less of a behavioral driver and functions more as a signal that prepares female bodies for changes in preparation for childbirth, such as in the uterus and breasts.

2. Attraction

Attraction is characterized by high energy investment and preoccupation towards a small number of people. From an evolutionary perspective, attraction may have developed to discriminate between multiple reproductive partners, allowing the focusing of limited resources towards fewer partners.

In Fisher’s theory, attraction is strongly related to the action of dopamine and norepinephrine. In humans, the reward circuitry (Section 11.2) is involved in feelings of love. Using fMRI, Fisher presented pictures of a patient’s romantic partner to them and identified increases in the blood flow to dopaminergic midbrain areas such as the ventral tegmental area and the striatum. This finding compares with imaging studies that observed increases in blood flow to insula, premotor, and hypothalamus as well as striatum in response to highly erotic, sexual imagery. These studies suggest that romantic love and lust have different driving neural structures underlying these behaviors.

Norepinephrine functions to increase attention, alertness, and energy, which accounts for the exhilarated feeling you may feel when spending time with a potential partner.

3. Attachment

Attachment is the long-term accompanied by feelings of comfort and emotional stability. Attachment also contributes to behaviors that maximize offspring survivability, such as sharing parenthood responsibilities and protectiveness towards offspring. The major neurochemical drivers of this form of love are oxytocin and vasopressin.

We have learned about mammalian pair bonding through experiments with **prairie voles** (*Microtus ochrogaster*), rodents that are indigenous to central North America. Atypical of mammals, prairie voles exhibit social monogamy after mating, where the two voles will cohabitate and share parental duties, such as mutual resource collection, nest building, and caring for their young.

Figure 15.21 Prairie voles, rodents found in the wild across central north America, form monogamous relationships after mating.

An experimental test called the **partner preference paradigm** can be used to assess vole monogamy. Here, a three-chamber testing apparatus is used where a test vole is placed in a central chamber. In the other two chambers are voles who have been harnessed into their rooms, unable to leave. The test vole is then free to move between any of the three chambers. When the test vole has mated with one of the harnessed voles, they will choose to spend more time with their "partner" vole compared to the novel "stranger" vole.

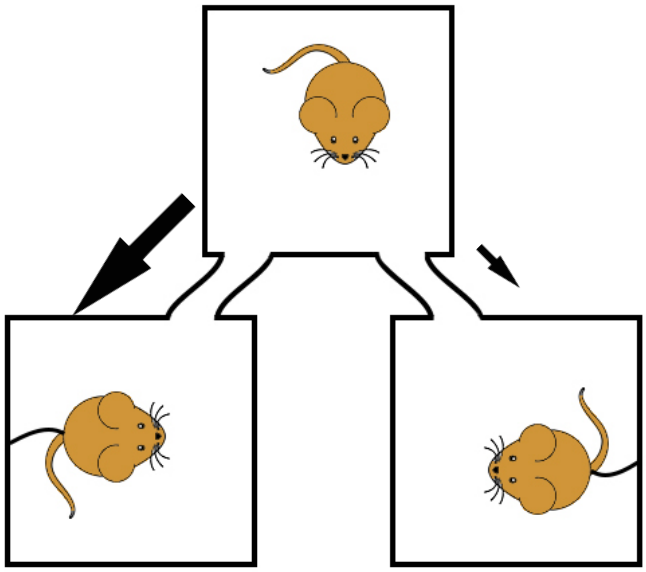


Figure 15.22 In a partner preference test, a mated prairie vole will choose to spend more time with their partner compared to the stranger, both of which are tethered to their chamber.

Not only do mated prairie voles prefer the company of their partner, they also demonstrate behaviors similar to some human romantic relationships. For example, mated prairie voles living together display selective aggression against a stranger, "intruder" voles, a behavior called "mate-guarding". They also spend significant time mutually caring for their young, their pair bonding can be modified by psychoactive substance use,

and they show increased anxiety when they are separated from their partner vole.

Other nonhuman studies have compared the prairie voles with physiologically comparable animal models, such as the montane vole or meadow voles. While these other rodents are similar in size, they have slightly different geographic distributions. Notably, they have radically different mating habits, as they exhibit social promiscuity rather than monogamy.

Comparing between monogamous and promiscuous species has allowed researchers to identify several neurotransmitter signals that are implicated in pair bonding. Oxytocin and vasopressin seem to play important roles in vole pair bonding, as inhibition of either of these signals decrease partner preference. Additionally, differences in dopamine receptor levels and corticotropin releasing factors are observed between vole species.

Parental love

Parental love refers to instinctive affection towards one's offspring. Parental love behaviors include nurturing (collecting and sharing resources), protecting (promoting aggression against "intruders"), and preparing one's young for their adult life (risk assessment training).

In evolutionary theory, parental love serves to improve the odds of passing of one's genes through the following generation. Most large mammals (humans included) are **K-selected** species, which benefit most from a small number of high-quality offspring that require substantial parental investment, as opposed to large numbers of offspring with little investment.

An extension of parental love is **familial love** or **kinship**, the protection and preferential support of one's extended genetic relations. Because blood relatives share common genes, increasing the survivability of these family members has benefits for passing genes to the next generation. In the words of the geneticist Jack Haldane, "I'd gladly give my life for two brothers or eight cousins."

Many behaviors related to mammalian motherhood are accompanied by changes in neural activity. Nursing, for instance, is feeding behavior that is regulated through a positive feedback cycle. Offspring suckling activates the mother's somatosensory afferents. Through a series of oxytocin-dependent circuits across the hypothalamus, suckling ultimately increases lactation through the **milk letdown reflex**. Auditory

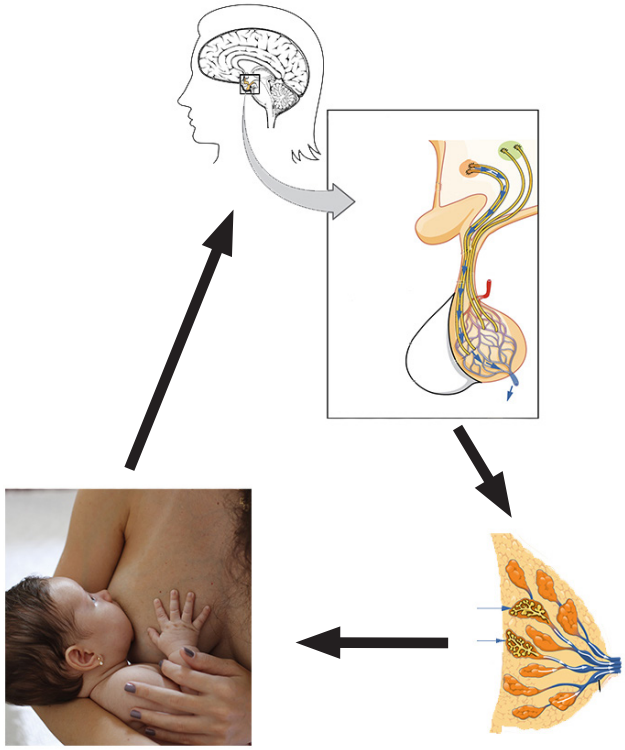


Figure 15.23 Suckling (bottom left) triggers somatosensory inputs, which send afferents into hypothalamus (top), increasing blood levels of oxytocin. Oxytocin increases accumulation of milk in the mammary glands (bottom right), which encourages increased suckling.

sensory inputs such as the sounds of a crying baby can also trigger this reflex. Sometimes, just thinking about the baby can induce letdown.

Some changes are dependent on neural plasticity. For example, after childbirth, the auditory areas of rodents rewire to become more sensitive to high frequency sounds. This adaptation allows the mothers to better detect the **ultrasonic vocalizations** that are emitted by offspring when they are distressed or hungry. Olfactory areas also change in order to become more sensitive to the particular odorants given off by their young, allowing them to better identify their offspring. In humans, these olfactory changes result in decreased aversion towards traditionally aversive stimuli (urine or fecal matter) when they originate from their children.

Disgust

Disgust is an evolutionarily protective emotion that helps protect an organism from being exposed to toxic or dangerous substances. It also leads to disease avoidance, which is why we dislike the smell of rotten foods or fecal matter, and why we may be repulsed by blood. As in the previous examples, disgust often happens in response to specific sensory stimuli (chapter 9).

The most common physiological response to disgusting stimuli is nausea, which can promote vomiting. Additionally, nausea decreases appetite, which prevents further intake of potentially toxic compounds.

Disgust is not inherently ingrained and has a strong learned cultural component that can be modified by experience. Americans may be turned off by potent moldy cheeses, which are delicacies

Figure 15.24 Typical face produced in response to a disgusting stimulus.

in many European countries, or fermented fish products, commonly used as a flavor enhancer in South East Asian cuisine.

Disgust extends beyond physical or chemical stimuli, and also manifests itself in thought. For instance, disgust is associated with sexual contact leading to low reproductive success, such as with animals or incestuous relationships. Antisocial behaviors can also provoke the disgust behavior, such as feeling physically nauseated when reading about serial killers.

The canonical “disgust face” is observed across several cultures. The curling of the upper lip and wrinkling of the nose decrease air flow through the nasal cavity, and squinting minimizes eye surface exposure. An opening of the mouth is a precursor for spitting out the disgusting flavors.

The insula is involved in mediating perception of disgust and displays changes in activity in fMRI activity in response to disgust-related cues.

Clinical connection: Obsessive-compulsive disorder (OCD)

OCD is a common psychiatric condition affecting an estimated 3% of the population, characterized by persistent intrusive thoughts or the need to perform some ritualistic series of actions or motions. OCD exists on a spectrum, and in severe cases, can severely impair the quality of a patient's life. OCD is often treated with psychiatric intervention or SSRIs like fluoxetine (more information in chapter 16).

Some studies implicate abnormal disgust sensitivity as contributing to OCD. Patients were shown a variety of disgust-provoking stimuli, such as unsanitary toilets or parasitic infections. If the threshold of disgust sensation is particularly low a patient may engage in their perseverative behavior more frequently.

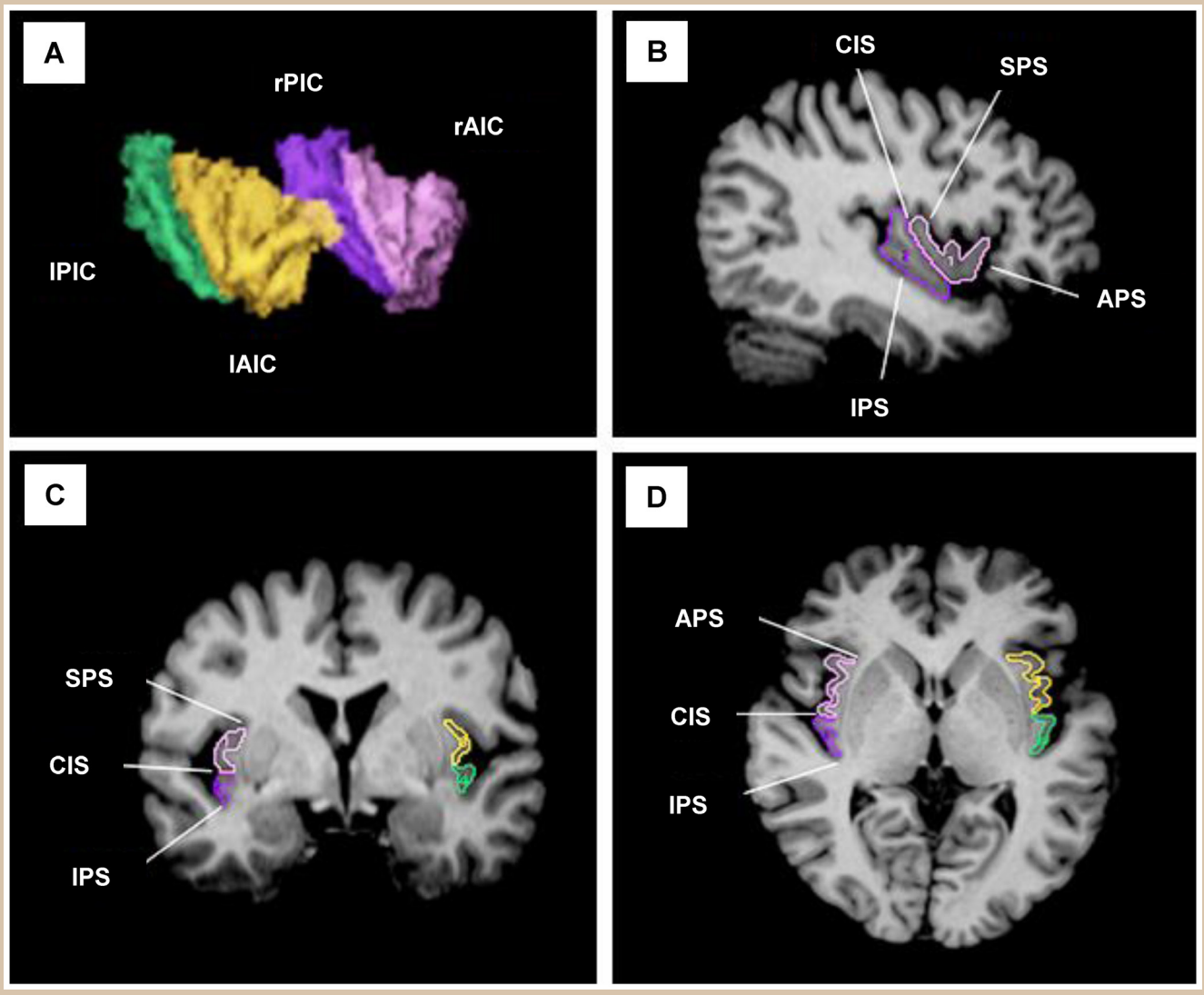


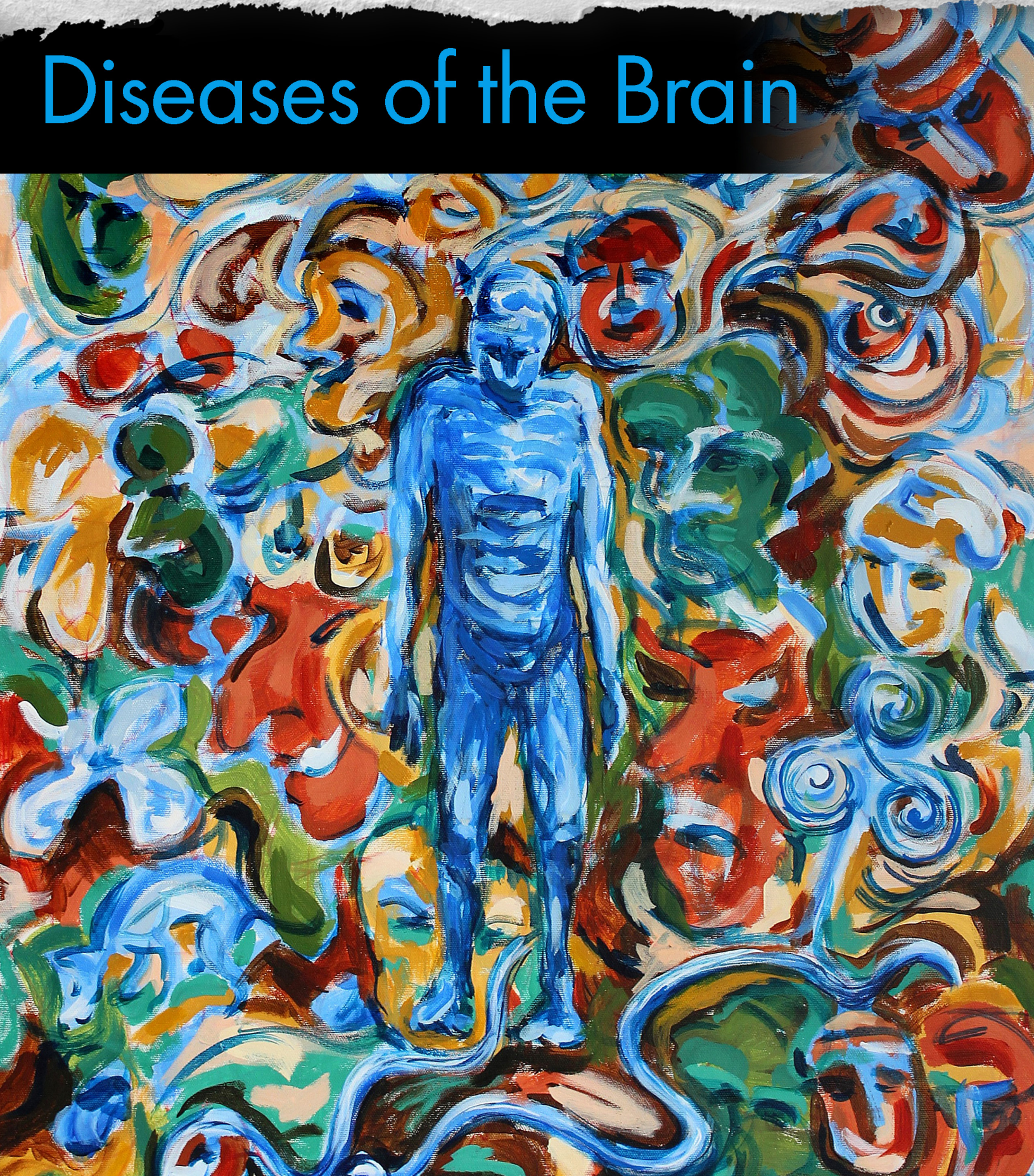
Figure 15.25 Morphological differences are observed in the volume of the insular cortex between people with OCD and neurotypical patients.

Image credits
Cover: https://pixabay.com/photos/airport-tourism-flying-air-traffic-1515431/
15.1 https://commons.wikimedia.org/wiki/File:The_Feeling_Wheel.png
15.3 https://pixabay.com/photos/bicycle-bike-biking-sport-cycle-384566/
15.4 https://pixabay.com/photos/cat-cat-hissing-angry-pet-5534007/
15.5 https://commons.wikimedia.org/wiki/File:Papez.jpg modified by Austin Lim
15.6 https://pixabay.com/photos/lion-tailed-macaque-monkey-rare-216227/
15.8 https://upload.wikimedia.org/wikipedia/commons/5/5a/Expression_Of_The_Emotions_In_Man_And_Animals_%28IA_TheExpressionOfTheEmotionsInManAndAnimals%29.pdf
15.9 https://commons.wikimedia.org/wiki/File:Facial_Action_Coding_System_-_Paul_Ekman_%26_Wallace_Friesen.png
15.10 https://commons.wikimedia.org/wiki/File:Amygdala.png
15.11 https://commons.wikimedia.org/wiki/File:1806_The_Hypothalamus-Pituitary_Complex.jpg
15.12 https://upload.wikimedia.org/wikipedia/commons/4/4a/1810_Major_Pituitary_Hormones.jpg modified by Austin Lim
15.13 https://upload.wikimedia.org/wikipedia/commons/b/b4/Sobo_1909_633.png
15.14 https://commons.wikimedia.org/wiki/File:Expression_of_the_Emotions_Plate_VII_(fear1).jpg 15.15 https://commons.wikimedia.org/wiki/File:Gigi_a_motley_yellow_corn_snake.jpg https://upload.wikimedia.org/wikipedia/commons/e/eb/Waverlyhillssanatorium.jpg
15.16 https://phil.cdc.gov/Details.aspx?pid=3421 modified by Austin Lim
15.17 https://commons.wikimedia.org/wiki/File:Response_to_stress.jpg modified by Austin Lim
15.18 https://commons.wikimedia.org/wiki/File:DNA_simple2-bn.svg modified by Austin Lim
https://commons.wikimedia.org/wiki/File:Cortisol.png
15.19 https://commons.wikimedia.org/wiki/File:H_pylori_ulcer_diagram.png
15.20 https://commons.wikimedia.org/wiki/File:Two_swans_in_a_lake_(23746714432).jpg
15.21 https://commons.wikimedia.org/wiki/File:Prairie_vole.gif
15.22 https://pixabay.com/vectors/mouse-rodent-gnawer-white-animal-146416/ modified by Austin Lim
15.23 https://commons.wikimedia.org/wiki/File:2922_Let_Down_Reflex-new.jpg modified by Austin Lim https://pixabay.com/photos/breastfeeding-mother-mother-s-love-1570695/ modified by Austin Lim
15.24 https://commons.wikimedia.org/wiki/File:How_to_express_disgust_(33491312650).jpg
15.25 Disproportionate Alterations in the Anterior and Posterior Insular Cortices in Obsessive–Compulsive Disorder. Song A, Jung WH, Jang JH, Kim E, Shim G, et al. (2011) Disproportionate Alterations in the Anterior and Posterior Insular Cortices in Obsessive–Compulsive Disorder. PLOS ONE 6(7): e22361. https://doi.org/10.1371/journal.pone.0022361

The Open Neuroscience Initiative is funded by a grant from the Vincentian Endowment Fund of DePaul University.

Chapter 16:

Diseases of the Brain



Austin Lim, PhD (DePaul University)
Editor: Monica Javidnia, PhD (Rochester University)

The brain can be thought of as a well-oiled machine made up of hundreds of billions of moving parts, all cooperating in tandem for healthy behavior and function. There is redundancy in the way these parts are organized, allowing for occasional mishaps without any significant loss of function. But when these parts interact in unusual or atypical ways, a person may develop some psychiatric disorder. The conditions described in this section likely involve dysregulations in molecules, cells, or circuits, and are therefore complex.

As far as we know, the likelihood of developing these conditions is not exclusively determined by either genes or influence from the environment. Instead, there is probably some influence from both. That is to say, none of these conditions are 100% **penetrant**; none of them are dictated exclusively by genetics. Having two parents or an identical twin with the condition may indicate an elevated baseline risk over the population at large, but it is not a guarantee that the disease will manifest. Environmental triggers and other exposures may lead to a sudden onset of the condition; on the other hand, certain factors in the environment may be protective against these diseases.

One of the major challenges with understanding these brain diseases is related to the difficulty of making an accurate diagnosis - as almost everything in biology exists on a spectrum, so do these brain disorders. The symptoms of these disorders frequently overlap, adding another layer of complexity. To help establish a diagnosis, the American Psychiatric Association (APA) has put together a series of criteria for psychiatrists to diagnose these complex conditions. The guidelines are compiled in a book called the **Diagnostic and Statistical Manual of Mental Disorders**. They are currently on the fifth revision of the text, referred to as the **DSM-5**. It's an imperfect set of criteria, but it is a start towards understanding these remarkably complicated conditions.

Figure 16.1 Environment and genetics both contribute to these diseases. Studies comparing twins can be helpful in identifying the influence of genetics.

Many of the treatments we currently have for neuropsychiatric disorders aren't always effective. Our ability to treat these conditions depends on our understanding of the disease. The better we understand the causes of these conditions, the wider variety of new therapies we can test. Therapeutic strategies for these conditions are often first tested using cells and non-human animal models, but these have shortcomings. Most of the time, animal models of disease incompletely mimic the symptoms of

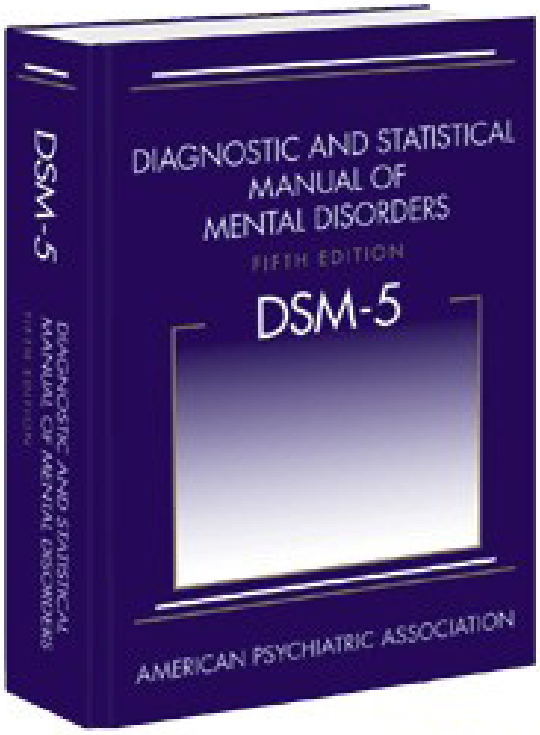


Figure 16.2 The DSM-5 is the manual that is used by psychiatrists to diagnose various psychiatric conditions.

the disease. Animal research concerns itself with three different forms of validity.

1. Face validity. If an animal model of a disease looks similar to the human condition, whether behaviorally or in physical appearance, we say that the model has good face validity. The animal exhibits the same set of symptoms that you would see in a human affected by the same condition, such as a rat model of post-traumatic stress disorder (PTSD) where the rat is exposed to a predator. Future exposure to predator-associated cues causes the rat to exhibit anxiety and avoidance, which are symptoms of human PTSD. In this case, the model has good face validity.

2. Construct validity. Sometimes, an animal model of a disease starts with the same pathological changes in the brain that are observed in human patients. We say that these models have good construct validity. The risk of developing Huntington's disease (Chapter 10) in humans is associated with the number of poly-glutamine repeats in the Huntingtin protein. Developing a genetically modified animal model that has several poly-glutamine repeats is an example of a model with good construct validity. Because humans and non-humans are very different animals, having the same origin of disease does not always produce the disease symptoms.

3. Predictive validity. An animal model has good predictive validity if the animal model can be used to predict whether a therapy would be effective in treating humans with that same condition. For example, if there were a genetic mouse model that showed symptoms of depression, and an experimental antidepressant reverses depression in both the mouse and humans with depression, the model would be said to have good predictive validity.

Unfortunately, we don't have any animal models that reproduce the symptoms of the most complex human conditions - it is almost impossible to create a mouse model of

Figure 16.3 A predator exposure paradigm has strong face validity because it causes a mouse to become anxious, just like a human when they are exposed to a predator.

dissociative identity disorder or dyslexia, and even if we could, scientists would struggle to quantify the behaviors that we use as diagnostic criteria, which are too subtle to be observed or quantified in non-humans. And for the animal models that we do have, they are often imperfect or incomplete, modeling only some of the deficits seen in humans. Furthermore, most human diseases have many symptoms, and only a few can be assessed with behavioral tests. We can only study disorders of the brain that have a clearly and easily quantifiable behavioral component.

16.1 Schizophrenia (SZ)

Schizophrenia is a psychiatric condition affecting just under 1% of people. SZ affects men slightly more often than women and affects people of all races. There is a strong association between low socioeconomic status and the risk of developing schizophrenia, indicating that stresses such as neonatal nutritional deficiency or food insecurity may be risk factors. Other risk factors that contribute to increased SZ risk include prenatal drug exposure, heavy drug use during early adolescence, and childhood adversity.

A diagnosis is generally made while a person is in their late adolescent years through their thirties. During this phase of life, the brain is still undergoing subtle maturation processes, which may account for why a person is more vulnerable in these years. After this age, the risk of developing SZ decreases significantly. Also, the later in life that SZ symptoms appear, the better the health outcomes are.

It is worth noting that people with SZ have the neurotypical range of intelligence, with the occasional outliers: John Nash, the real-life Nobel prize-winning economist depicted in the movie A Beautiful Mind, was first diagnosed with SZ in 1959.

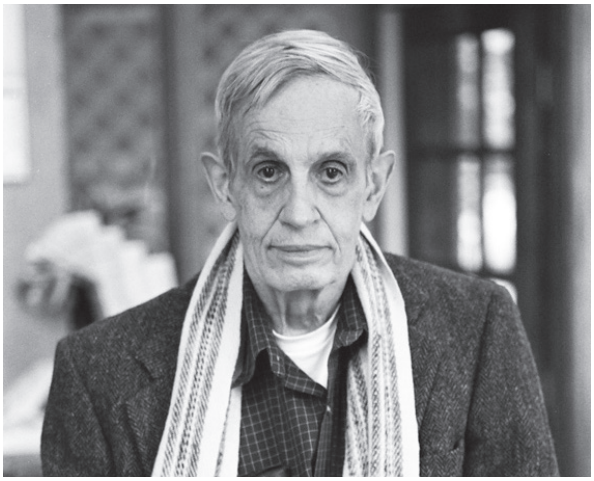


Figure 16.4 Dr. John Forbes Nash was diagnosed with schizophrenia before receiving the Nobel Prize in Economics in 1994.

Symptoms of schizophrenia

The symptoms of SZ can be roughly classified into two categories, positive symptoms and negative symptoms. These phrases are used to describe whether there is an excess of some function (**positive symptoms**) or a deficit of a function (**negative symptoms**). The symptoms do not appear uniformly across patients, so not all patients develop every symptom.

The most well-known positive symptom of schizophrenia is **hallucinations**, perceiving something that is not there (as opposed to illusions, which are misinterpretations of things that are there). Usually, patients experience auditory hallucinations, but they more rarely have visual hallucinations. The voices that these people hear may be consistent or can change over time. Interestingly, the nature of these hallucinations is influenced by society. In cultures with strong ancestor reverence, they may hear the voices of their grandparents, whereas people in religious cultures may hear the voices of deities.

Figure 16.5 Auditory (and sometimes visual) hallucinations are a common positive symptom of SZ.

Relatedly, people with SZ may experience a variety of **delusions**, untrue beliefs that cannot be changed despite overwhelming evidence. The delusions can come and go spontaneously. Delusions exist in many forms. A paranoid delusion is when a person believes that they are being spied on, maybe by the government or by aliens. A persecutory delusion is a persistent thought that the world is out to get them or to do them harm. Delusions of grandeur are when a person has a tremendously high sense of self-esteem, believing that they are royalty or are the reincarnation of God.

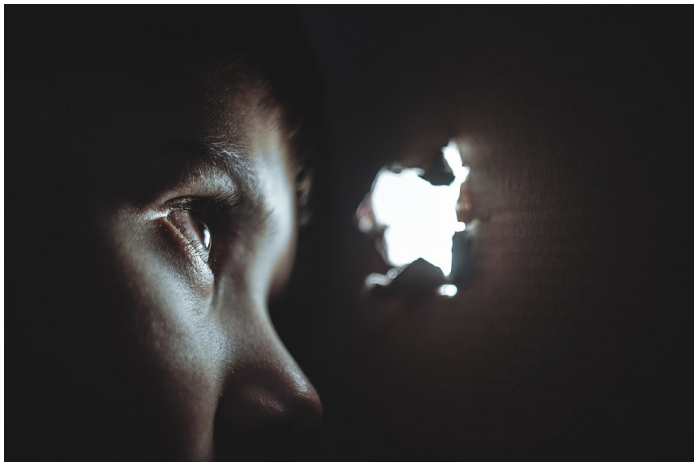


Figure 16.6 Persecutory delusions, a symptom of SZ, is the persistent belief that someone or something is constantly watching you.

The negative symptoms of SZ may include deficits in expression. One common symptom is a **flat affect**, where a patient does not show or express emotion in situations where you would expect to see them. A related negative symptom is **alogia**, a decrease in the use of language. People with alogia often use vague language that is lacking in content or repetitive.

Negative symptoms also include deficits in motivation or interest. Two closely-related negative symptoms include **anhedonia**, a loss of a sensation of pleasure and the inability to expect upcoming pleasure. Furthermore, patients with SZ may also exhibit **avolition**, a decrease in goal-directed activity, which can cause a person to stop seeing their friends and cease displaying interest in social gatherings, leading to worsened interpersonal relationships.

Other negative symptoms that can present in SZ are a variety of motor disturbances. Basal ganglia and cerebellar structural deficits are found in SZ, two brain structures involved in motor control (see chapter 10), which may explain why deficits are observed. One motor difficulty is **catatonia**, where a person can hold their body in a highly unusual position for a prolonged period of

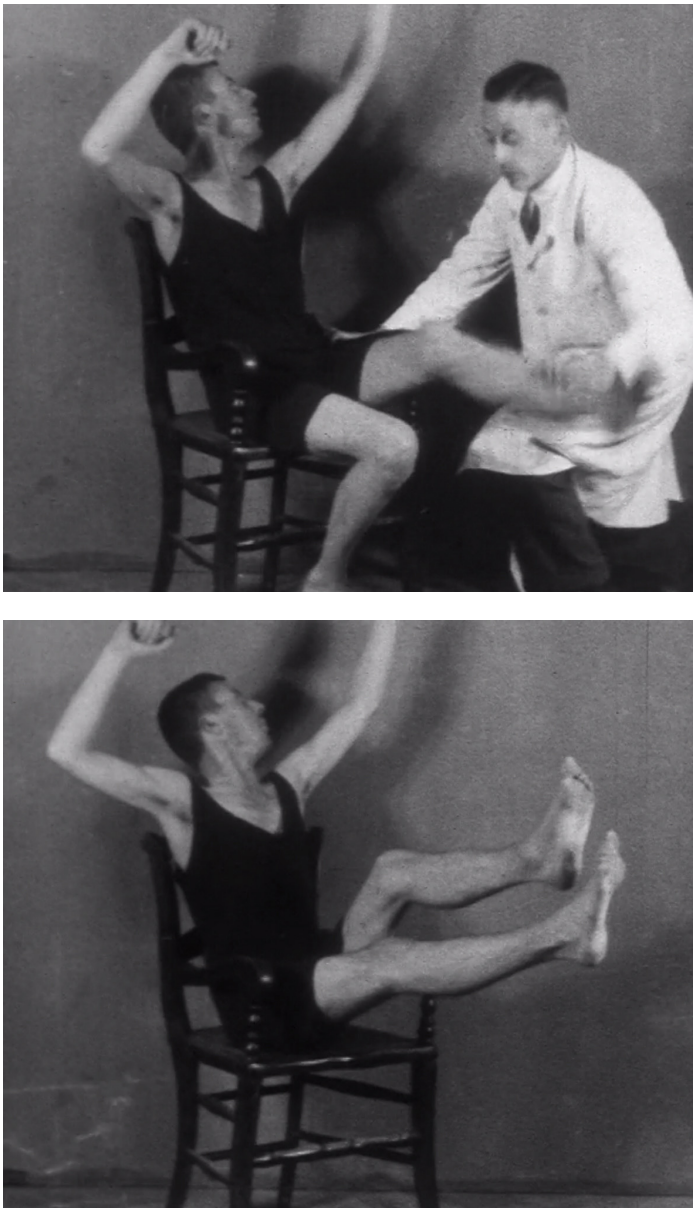


Figure 16.7 After being moved gently into an unusual body position, a person with catatonia may stay in that position for a prolonged time.

time. They may also display **stereotypy**, a series of repetitive, purposeless behaviors, such as the persistent rocking of the body or self-caressing.

Negative symptoms may also manifest as a deficit of a patient's cognitive abilities, particularly shortcomings in episodic memory (Chapter 13). They may also present with difficulty in performing attention-related behavioral tasks.

Potential causes of schizophrenia

In a healthy person, dopamine is important for motor control and motivation, two behaviors that are changed in patients with SZ. Therefore, scientists have suggested that abnormal dopamine signaling may be an underlying root cause. While the **dopamine hypothesis** is one of the earliest theories of SZ, modern genetics studies have shown that polymorphisms in the dopamine D2 receptor are risk factors.

Atypical cortical neuron network development is also likely to be present in SZ. In the healthy brain, networks of cortical neurons produce cyclic patterns of activity in the 40 Hz range, a pattern called a gamma oscillation. These gamma oscillations result from a combination of excitatory and inhibitory neurons. In SZ, there is a decrease in the density of dendritic spines on the excitatory neurons with a simultaneous decrease in GABA-ergic signaling, which leads to unpredictable gamma oscillations.

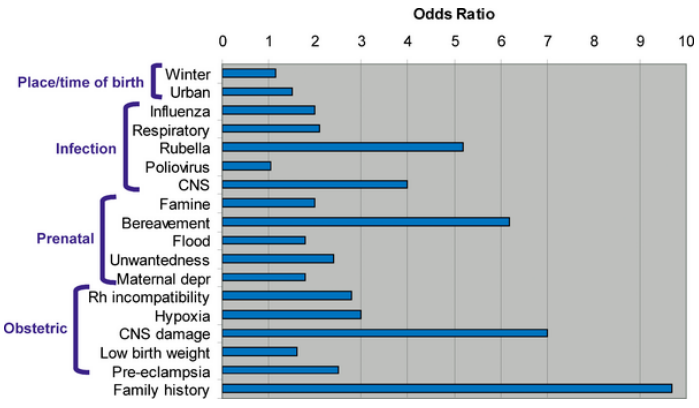


Figure 16.8 Many environmental factors contribute to the risk of developing SZ.

Animal models of schizophrenia

One animal model for SZ is based on the hypothesis that excess dopamine leads to the disorder. Introducing high doses of the drug amphetamine, which increases dopaminergic signaling, induces a temporary schizophrenic-like state in non-human animals. The hyperdopaminergic model of SZ produces cognitive deficits with no changes in memory or other negative symptoms. Alternatively, administration of NMDA glutamate receptor antagonists, like ketamine or PCP, is also used as a behavioral model of SZ.

Other non-human models of SZ are neurodevelopmental models. In these models, a pregnant dam is exposed to the compound MAM, which causes the newborns to develop atypically and display behavioral deficits similar to SZ. Inducing an unusually strong immune response in the pregnant mother can also cause atypical development in utero, which causes the animals to experience behavioral deficits after birth.

The biggest limiting factor to developing an animal model is that many symptoms in human SZ, like paranoid delusions or auditory hallucinations, are impossible to detect and quantify in non-humans. The PCP model can cause changes in rodent social behaviors, but it is hard to tell if this model causes any of the positive symptoms that you might see in a patient with SZ. Despite the limitations of these non-human models of SZ, they have been helpful in testing the therapeutic efficacy of anti-schizophrenia drugs.

Treatments of SZ

The dopamine theory of SZ has led to a few novel therapeutic strategies, especially in the context of the dopamine D2 receptor. D2 antagonists decrease hallucinations and delusions in some patients with SZ, and the effectiveness of the antagonist is correlated with the ability of that drug to block the D2 receptor. Clozapine, an atypical antipsychotic that functions as a dopamine receptor antagonist, can decrease SZ symptoms as well.

Unfortunately, pharmacological therapies are not always effective in humans. Around a third of patients discontinue their treatment regimen, and around a fifth of them report adverse side effects such as extrapyramidal motor symptoms, sedation, and weight gain.

A potential new therapy is based on transcranial magnetic stimulation (Chapter 6). Some evidence suggests that targeted activation of the cortex can decrease the severity of auditory hallucinations. There may also be some mild improvements in the negative symptoms.

Approximately 65% of North Americans with SZ are smokers, compared to 25% in the population at large. If they are smoking as self-medication to activate dopamine- or acetylcholine-sensitive networks in the brain, this observation may lead to a new therapeutic strategy. Alternatively, they may smoke to get pleasure, which acts to reverse the anhedonia that is one of the main symptoms of the disease.

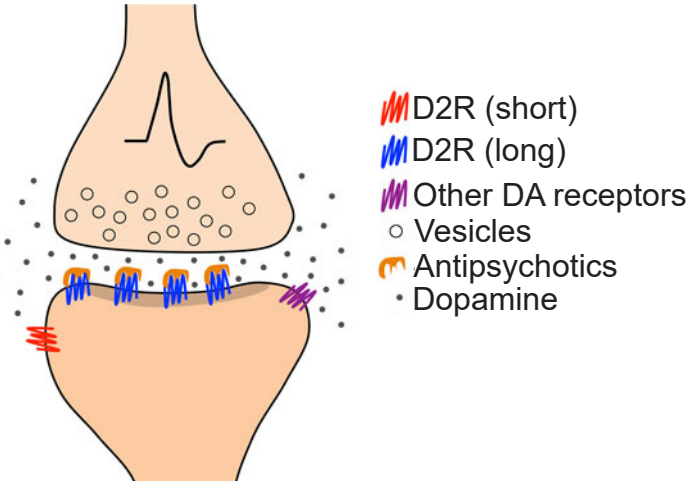


Figure 16.9 Atypical dopamine signaling is believed to contribute to schizophrenia.

16.2 Major depressive disorder (MDD)

Depression is a highly prevalent condition with a lifetime risk of about 18%. Depression gets diagnosed more frequently in women than in men, affecting about 5% of women and 2.5% of men. Even with treatment, there is a high rate of relapse: an estimated 80% of people with depression have more than one episode in their lifetime. The prevalence of depression is similar across both high-income and low-income countries, indicating that biological factors contribute significantly to the disease.

Figure 16.10 Spanish artist Pablo Picasso likely experienced some form of depression, and the paintings made during his "Blue Period," like *The Old Guitarist* (1903), reflect his emotional state.

MDD is also called **unipolar depression** to differentiate it from the depression that represents a phase seen in bipolar disorder (See section 16.3).

Clinical depression is often associated with another medical or psychiatric condition. For example, rates of depression are higher among people with terminal diagnoses, like cancer. In this case, we say that the two are **comorbid**. Additionally, a set of particularly challenging circumstances, like the death of a loved one, could trigger a depressive episode.

MDD represents a severe health risk across all ages. About 18% of adolescents report at least one instance of non-suicidal self-injury, and the lifetime risk for suicide among people with MDD is estimated to be about 10%.

Symptoms of MDD

The fundamental criteria that are used to diagnose people with MDD is a depressed mood/ self-esteem, low energy, and anhedonia, the decrease in sensitivity to pleasure. Because of the decrease in pleasure, they have a lessened desire to engage in activities that once produced happiness, thus leading them to become withdrawn from their friends and family.

Short-term changes in mood are completely normal and not clinical. The main diagnostic criteria for MDD is the severity and duration of the symptoms. When the depression begins to affect other aspects of life, including feelings of worthlessness, changes in sleep or appetite, difficulty concentrating, or suicidal ideation that persist daily for two weeks or longer, then a person may be diagnosed as clinically depressed.

To diagnose MDD, a trained psychiatrist or psychologist would use a combination of interview with a self-report questionnaire such as the **Hamilton Rating Scale for Depression (HAM-D scale)** or the **Beck Depression Index**. Items that appear on these sorts of tests include:

"Feels like life is not worth living"
"Experience frequent weeping"
"I blame myself all the time for my faults"

To date, there is no biomarker for depression.

Treatments for MDD

There is currently no completely effective treatment for MDD that reliably works for everyone. The currently accepted strategies can be divided into behavioral treatments and chemical treatments.

Cognitive behavioral therapy (CBT) is a therapist-guided form of talk therapy that may help a person manage their depression. In CBT, a patient's behavior, including their coping mechanisms, erroneous thoughts, and emotional responses, is analyzed through careful clinical examination and patient self-reflection. Then, the patient is taught mechanisms to counteract those maladaptive behaviors and replace them with adaptive behaviors. For example, a person undergoing CBT for MDD might learn to identify the moments when they dwell on something negative in their lives, and then learn to tell themselves "that thought does not function to make my day better. Let's start the day by getting out of bed and see what happens next." It should be known that CBT is not exclusively for the treatment of MDD; CBT can also be effective for anxiety, OCD, PTSD, insomnia, substance use disorders, behavioral addictions, and many others.

Clinical connection: Seasonal affective disorder (SAD)

Seasonal affective disorder (SAD) is one type of depression that has a dependence on daytime sunlight, increasing in prevalence in the winter and decreasing in the summer. The prevalence of SAD is heavily correlated with distance from the equator: people living closer to the poles experience longer nights in the winter, which increases SAD risk. In the US, for example, SAD affects about 1% of people in Sarasota, Florida, but about 9% of people in Anchorage, Alaska.

Light exposure therapy, while controversial, may show benefits for people with SAD, particularly intense broad-spectrum light or blue wavelength light. People with SAD may have something unusual about their intrinsically photosensitive retinal ganglion cells.

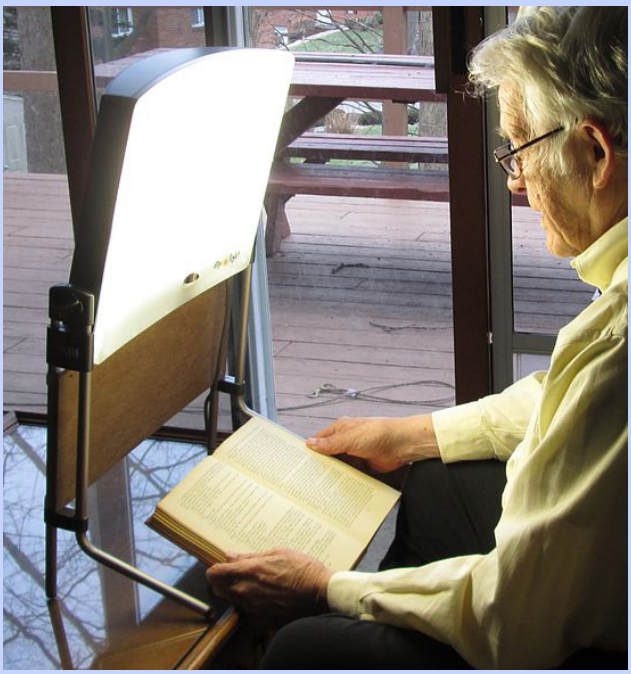


Figure 16.11 Light exposure therapy may be effective at treating some cases of seasonal affective disorder.

A wide variety of drugs are used for the treatment of depression. The **first-generation antidepressants** were developed in the 1950s and 1960s. These drugs acted to increase the action of the monoamine neurotransmitters: primarily dopamine, norepinephrine, and serotonin. Our body uses an enzyme called **monoamine oxidase (MAO)** which degrades these chemicals into inactive components that do not signal at receptors. These first generation antidepressants block the action of MAO; biochemically, we call them **monoamine oxidase inhibitors (MAOIs).** In the presence of an MAOI, the neurotransmitter signal remains in the synapse longer, similar to how an acetylcholinesterase inhibitor increases ACh signaling (Chapter 10).

Most of the MAOIs, while sometimes effective, have fallen out of fashion clinically because of the adverse side effects associated with their biochemical activity (however, they are still commonly used in treatment of Parkinson's disease). Some of them interact dangerously with foods rich in tyramine (particularly fermented foods, such as aged cheeses or beer, as well as beans and processed meats), an amino acid that is degraded by MAO. Excess tyramine can activate the sympathetic nervous system, and body levels of tyramine can rise to dangerous levels in the presence of an MAOI, leading to adverse cardiovascular events like stroke.

Many of the common MAOIs, like phenelzine and isocarboxazid, can be damaging to the liver. Some MAOIs also produce unwanted side effects, such as psychosis or nausea.

A different class of antidepressant drugs called the **tricyclic antidepressants (TCAs)**, named for the shape of their chemical structure, was also developed around this time. They generally act as monoamine reuptake inhibitors, resulting in elevated neurotransmitter signaling.

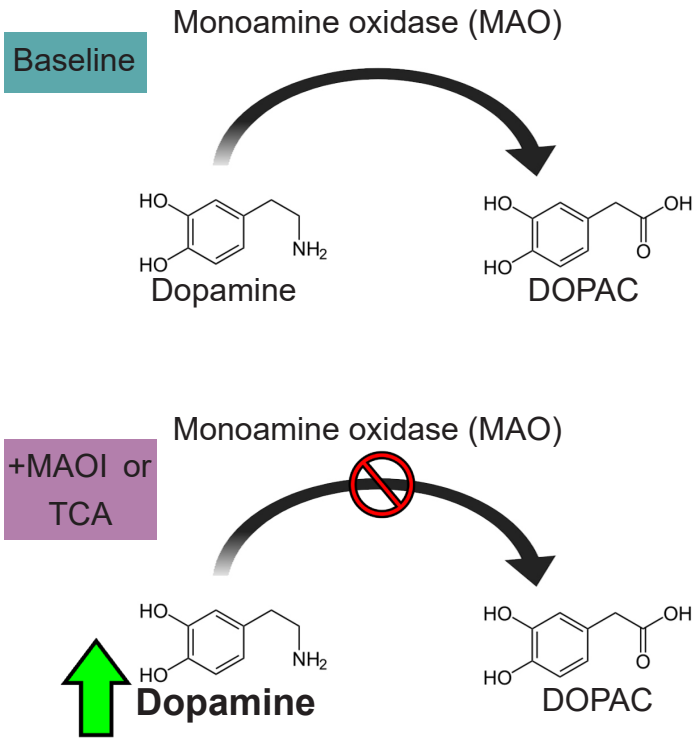


Figure 16.12 MAOIs and TCAs both increase neuronal signaling by decreasing the metabolic degradation of neurotransmitters, such as dopamine.

Unfortunately, these tricyclics may produce many severe side effects, such as seizures, tachycardia, and heart attacks, so prescriptions must be monitored closely. The tricyclics are still prescribed today for several other nervous system disorders, ranging from insomnia to neuropathic pain, but due to their potential cardiotoxicity, they are not often the first line of treatment in depression.

Our current, most often prescribed class of compounds for MDD, called the **third-generation antidepressants**, are focused on boosting the signaling activity of serotonin. Instead of preventing degradation, like the MAOIs, these compounds block the reuptake of serotonin out of the synapse. These chemicals are called **selective serotonin reuptake inhibitors (SSRIs),** of which fluoxetine (Prozac) is one of the most well-known examples.

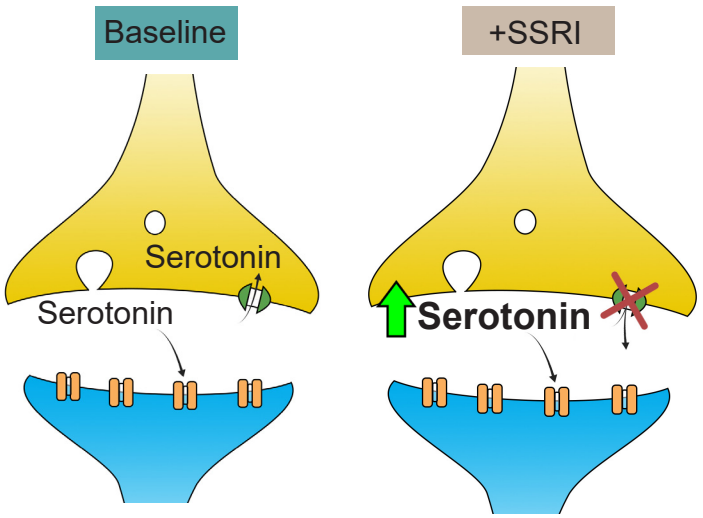


Figure 16.13 Third-generation antidepressants act to increase neurotransmission at the synapse by inhibiting reuptake.

While SSRIs can be effective at reversing the side effects of depression, they are not perfect drugs. One shortcoming is that a person needs to be on the drug for 2-4 weeks before they start to experience a clinically meaningful reversal of depressive symptoms. This finding is highly unusual since the pharmacological, molecular-level effects of SSRIs take place within hours after taking the medication. Similar to SSRIs, SNRIs (serotonin and norepinephrine reuptake inhibitors) can also be used to treat depression.

One of the unsavory side effects of SSRIs is **serotonin syndrome**, a set of somatic changes resulting from excessive serotonergic signaling. In mild cases, a person may have an elevated body temperature, excessive sweating, rapid heart rate, and elevated blood pressure. In more severe cases, a patient may have severe fevers or seizures. Serotonin syndrome can happen in the event of an SSRI overdose, or as a consequence of some interaction between an SSRI and other drugs like MAOIs, MDMA (ecstasy), amphetamines, or cocaine.

Our newest treatment options for depression, recently approved by the FDA in March of 2019, is **ketamine**, a dissociative anesthetic typically used as a veterinary tranquilizer and a recreational club drug. Branded as esketamine, it can be administered rapidly via nasal spray. The strength of esketamine is the speed of its action. After taking a dose, the antidepressant effects of the substance can be observed within hours.

For severe or treatment-resistant depression, **electroconvulsive therapy (ECT)** is a treatment option. Introduced in 1938, the procedure has since been refined over the years (it is currently performed under anesthesia) and is considered to be well-tolerated and highly effective. However, side effects include aches, nausea, and memory loss.

One future therapy that is currently making significant medical advancements in curing depression is the use of psychedelic drugs, particularly **psilocybin** – a substance found in some species of wild mushrooms. As a potent serotonergic agonist, a single dose of psilocybin has been shown to decrease depression scores in various self-report studies with little to no adverse side effects.

Animal behavioral tests of depression

Behavioral neuroscientists have developed a variety of tests to assess the effectiveness of antidepressant drugs in non-humans. They are roughly divided into two categories.

1. Despair-based tests. One symptom of depression is that a person "gives up," a behavior that can be modeled in a variety of rodent tests. One such example is in a **tail suspension test**. The mouse is held by the tail upside down. While it does not cause injury to the animal, they do experience discomfort, and will generally struggle

to either free themselves or to get upright again. The sooner they stop struggling (more time spent immobile), the more despair they experience, or the more depressive they are. Giving a mouse an antidepressant like fluoxetine causes them to fight for a longer duration.

A similar test to assess "giving up" is the **forced swim test**. Here, a rodent is put into a container with some water. While they are naturally buoyant and are not at risk of drowning, they do not like being wet and will try to swim so that they can climb out of the water-filled container. As in the tail suspension test, they are unable to escape their predicament, and will eventually become immobile. Giving them an antidepressant decreases time spent immobile.

Figure 16.14 The tail suspension test (left) and the forced swim test (right) are despair-based tests that assess depression-like behaviors in non-humans.

2. Reward-based tests. These tests seek to measure the severity of anhedonia, one of the main symptoms of depression. For example, a rodent may be presented with a **two-bottle choice task**, where they are put into a cage with two different bottles to drink from: one filled with standard water, and the other filled with a more desirable sugar-water solution. A depressed rat will not drink from the sugar-water bottle as frequently as a healthy rat but give that rat an antidepressant, and they will prefer the sweet water.

An intracranial self-stimulation paradigm can also be tested here in the context of depression (see chapter 11). When an electrical stimulator is placed in a reward area of the brain and the rodent is trained to perform some operant task to receive activation of these areas, we find that depressed mice do not activate these areas of their brain as much as one on antidepressants.

16.3 Bipolar disorder (BD)

A person with bipolar disorder (BD) experiences phases of clinical depression as described above, and at other times they experience **mania**, a state of exhilarating high energy. During this state, they may sleep very little, have difficulty concentrating, and experience pressure of speech: a perceived need to speak very rapidly to get their thoughts out. They might make poor financial or life decisions, such as deciding to abruptly leave their family behind. Historically, BD has been called "manic depression." Notably, actress Carrie Fisher and musician Demi Lovato both struggled with BD.

The estimated prevalence of BD is around 2.5%, but the disease is often misdiagnosed in the clinic as MDD. One reason this happens is that more people are aware of the symptoms of depression, and these symptoms are generally more easily observed. Mania is more difficult to identify, since in mild cases it may be hard to distinguish from a person just "being in a really good mood."

For a diagnosis of BD according to the DSM-V, a mood cycle has to last for a week or more. The word "bipolar" is often misused in pop culture. Frequent changes in mood from happy to sad does not characterize BD. In fact, for a person to be diagnosed with **rapid-cycling bipolar disorder**, they need to experience four mood transitions annually!

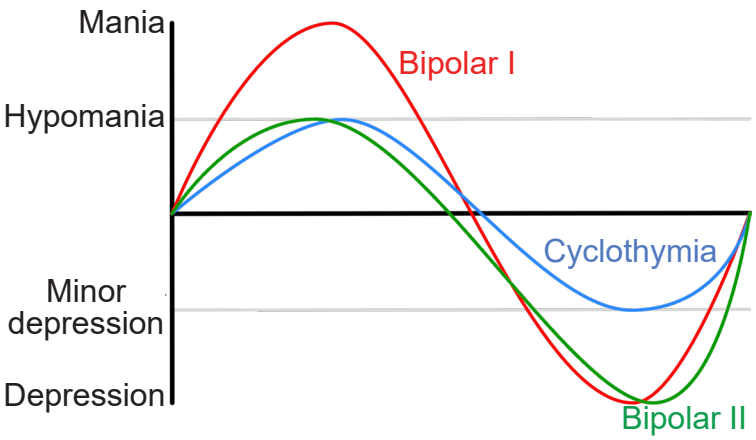


Figure 16.15 Severity of symptoms is used clinically to differentiate between Bipolar I, bipolar II, and cyclothymia.

BD is usually diagnosed in adolescence and early adulthood. As with depression, there are some genetic factors involved, since a family history of BD is a risk factor. Concordance rates for identical monozygotic twins are estimated to be between 35%-80%. But, environmental influences may be the precipitating factor in the onset of BD.

BD is diagnosed into two categories, based roughly on the severity of symptoms. **Bipolar 1 disorder** is the more severe of the two conditions, with a clinical diagnosis made when a patient experiences depressive or manic events that cause significant social or occupational impairment, or hospitalization to prevent serious self-harm. A diagnosis of **bipolar 2 disorder** is less severe, but the behavioral changes are still noticeable by friends and family. A related diagnosis is **cyclothymia**, where a person has alternating mood states that shift from depression to hypomania, a less severe state of mania. Like most other disorders, BD exists on a spectrum, and these labels only exist for simplicity.

Treatments for BD

The main issue with BD therapy is bringing the patient to some "middle" state: an antidepressant may treat the depression phase, but could also swing the patient into mania. Similarly, a mania-controlling drug could initiate depression.

Currently, our most reliable therapy for BD is **lithium** drugs. These compounds are described as mood stabilizers since they act to move the patient's mood to the center, rather than being at either the high end of mood (mania) or the low end (depression). The way lithium acts to reverse the symptoms is still unknown, and it probably acts on multiple pharmacological targets.

The main downside of this therapy is that lithium is very toxic. It has a very narrow therapeutic window: blood levels of lithium lower than 0.6 mEq/L produce no effect, and anything above 1.5 mEq/L causes delirium, tremor, fatigue, and deadly side effects like seizures and coma. It is also harmful to the kidneys after long exposure. Therefore, a person taking lithium drugs regularly undergo **therapeutic drug monitoring (TDM),** a procedure by which the concentration of lithium is assayed. TDM requires frequent visits to a hospital. Usually, patients get multiple blood draws in the first month when they start lithium treatment, decreasing to one test every 2 months, before decreasing to about four times a year.

BD is very challenging to model in non-humans. A genetically modified disruption of the circadian rhythm can induce mania-like symptoms, as can extending the length of daytime light exposure; oppositely, decreasing daily light exposure can induce depressive behaviors. Exposure to amphetamine can increase manic behaviors, while withdrawal from the drug can induce depression. Clearly, neither of these models for BD exhibit strong validity.

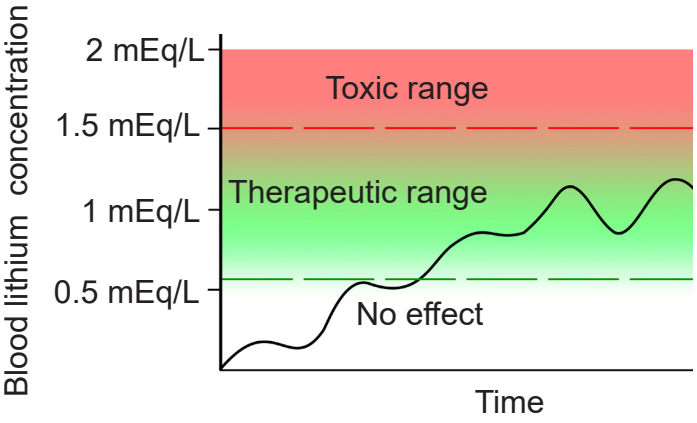


Figure 16.16 Therapeutic drug monitoring is important for people taking lithium for BPD since the medication is ineffective at low doses, but toxic at high doses.

16.4 Anxiety disorder

Anxiety is something that everyone has experienced at many points in their life. An anxious person may experience cardiovascular symptoms such as elevation of blood pressure and heart rate, shortness of breath, profuse sweating, and a state of panic. In many ways, the anxiety response is similar to the fight-or-flight response observed during sympathetic nervous system activity.

However, a clinical diagnosis of anxiety is different from the passing anxiety that we all experience. Anxiety disorders can be very common, and lifetime prevalence estimates suggest 29% of people could develop clinically significant anxiety over their life span.

According to the DSM-V, anxiety disorders have different presentations.

1. Generalized anxiety disorder (GAD). People with GAD experience a constant sensation of being overwhelmed, accompanied by fear and worry. Many times, this worry is not about a single concern, but rather a combination of issues all at once, such as financial issues, relationship issues, uncertainty of the future, and many others. GAD is much more severe and persists longer than the normal worries that affect everyone.

In GAD, worry persists for several months and is uncontrollable. There are also associated cognitive symptoms, such as fatigue, irritability, difficulty with concentration, and changes in sleep patterns.

2. Specific phobias. With specific phobias, a person develops the anxiety-related symptoms (cardiovascular and psychological changes) in response to highly specific stimuli, such as snakes, enclosed spaces, deep ocean, or public speaking. The person with the phobia perceives the stimulus to be a great threat, even though it does not actually pose a genuine threat. Most people with specific phobias will go to great lengths to avoid exposure to their particular phobia trigger. These phobias are often influenced by social and cultural conditions.

Developing a specific phobia has a lifetime prevalence of about 7%, but only a very small number of people with specific phobias ever seek treatment for their phobia. Like other forms of anxiety, there is a range of severity of these phobias.

Figure 16.17 A person with a specific phobia such as agoraphobia, the fear of unfamiliar environments where they have little control over their circumstance, may experience a panic attack in a crowd.

3. Panic disorder. A person with panic disorder experiences frequent **panic attacks**, characterized by sudden increases in heart rate, shortness of breath, dizziness, and sudden numbness or tingling (panic attacks can also be seen in specific phobias, but are not observed in GAD.) In panic disorder, these panic attacks may occur independently of external influences.

Pharmacologically, there are a wide variety of drugs that can be used to treat anxiety,

broadly called anxiolytics. The first-line therapies are usually SSRIs, the same class of compounds that are used in depression treatment. Other anxiolytics, such as the benzodiazepines alprazolam or clonazepam, act as positive allosteric modulators which increases the effect of the GABA system. Benzodiazepines are not always preferred since they may have abuse potential and can be addictive. Opioids and norepinephrine inhibitors can also decrease anxiety.

The exact cause of anxiety is still unknown. One theory suggests that anxiety is a maladaptive evolutionary response to our modern living conditions. The argument is based on the observation that an anxiety response looks a lot like a mild version of the fight-or-flight, sympathetic nervous system response: both elicit cardiovascular and respiratory changes. For 99% of the evolutionary history of *Homo sapiens*, we benefited from the sympathetic nervous system as a reflex to improve the odds of survival in dangerous situations. However, our modern civilized living conditions over the past few centuries have been very tame in comparison to the risks that our earlier ancestors experienced. The relative ease of living has let the main function of the sympathetic nervous system fall into disuse. The theory argues that people experience GAD because a part of them encourages sustained activity in the sympathetic nervous system. Although thought-provoking, this theory can't be tested experimentally and offers no explanation about a biological mechanism that can help to develop a therapy.

Animal behavioral tests for anxiety

As with depression above, there are non-human behavioral tests used to assess anxiety in rodents, such as the **elevated plus maze.** The maze is a raised platform, with four arms in the shape of a plus sign. Two of the arms have walls surrounding the sides, while the other two are open, exposed on all sides. The rodent is free to move between any of the arms as they choose. Standing in one of the open arms, where they can see the floor far below them, is an anxiety-provoking condition. Under normal circumstances, rodents choose to spend more time in the arms that are surrounded by walls. But if you give these animals an anti-anxiety drug, they increase the time spent in the open arms, indicating a decrease in the behavioral expression of anxiety.

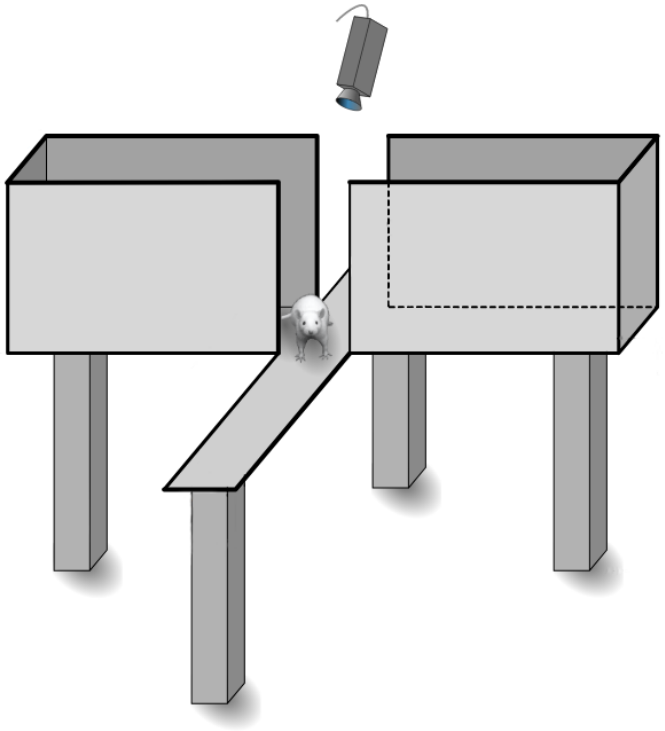


Figure 16.18 An elevated plus maze is one behavioral test for measuring anxiety behaviors in non-human animals.

A related behavioral test is the **open field test**. The test apparatus consists of a large, flat area where the rodent can move around freely, and some method to track the animal - either an aerial view camera or a series of parallel invisible

infrared beams that can locate the animal in the field. In the wild, rodents, as prey animals, prefer to spend more time close to the sides of the testing arena up against the wall, avoiding the wide-open space in the middle where their instinct warns them that they may be snatched up by some predator. However, if you give the rodent an anti-anxiety drug, they will spend more time venturing into the middle of the open field.

Another non-human model of anxiety is the **predator exposure paradigm**. In this paradigm, an ethologically-relevant stimulus is presented to the rodent, such as one of their naturally occurring predators. In this paradigm, rodent anxiety presents itself as a freezing response, an autonomic nervous system activity spike, and a reduction in non-survival behaviors. Although the predator exposure paradigm has good predictive validity, they may struggle with poor face validity, since the anxiety measures also may appear as many of several other conditions, such as PTSD or stress.

The Open Neuroscience Initiative is funded by a grant from the Vincentian Endowment Fund of DePaul University.

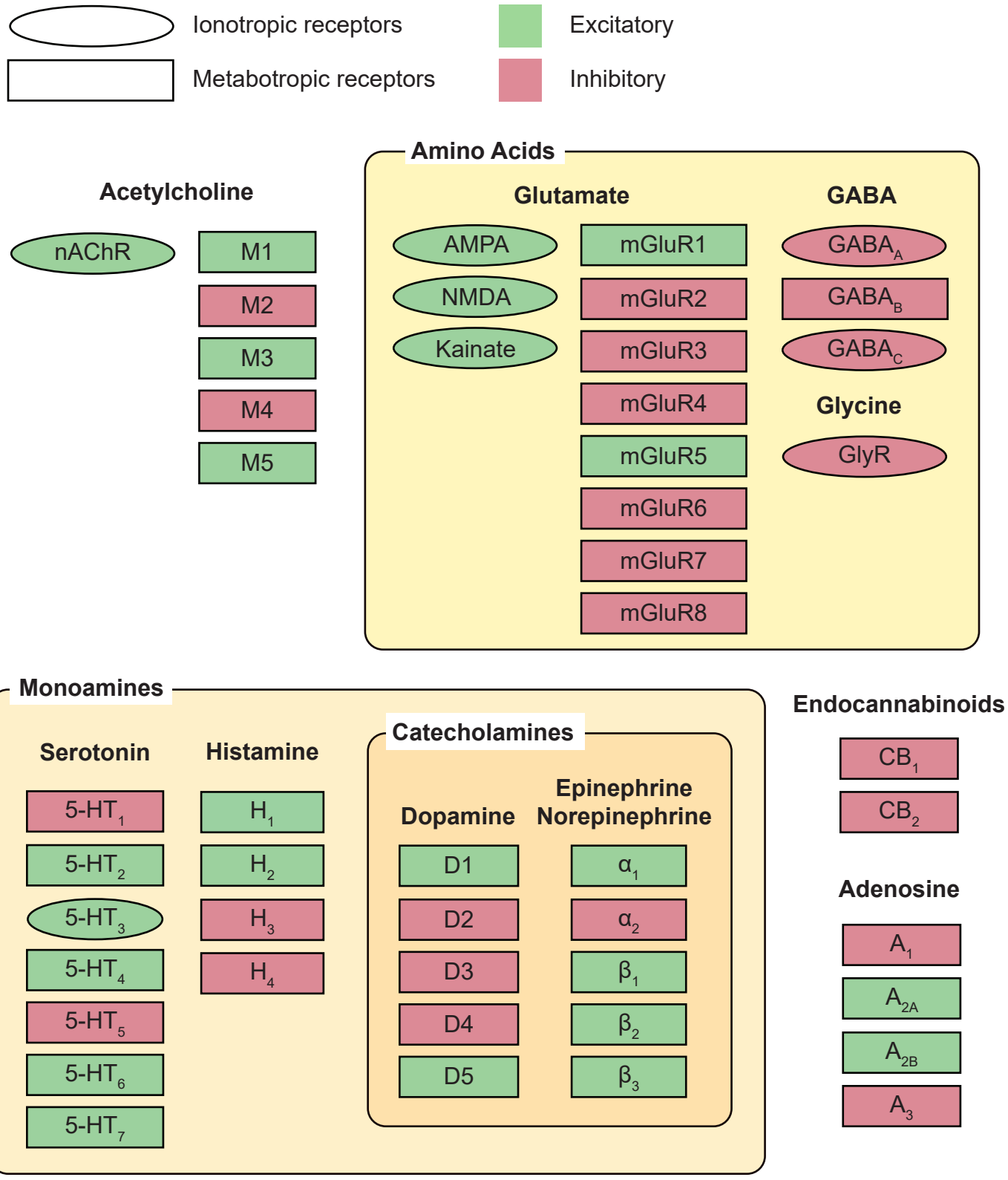
Cover https://pixabay.com/photos/lewy-body-dementia-dementia-2965713/
16.1 https://pixabay.com/photos/children-twins-girls-young-652270/
16.2 https://upload.wikimedia.org/wikipedia/commons/4/4b/Dsm-5-released-big-changes-dsm5.jpg
16.3 https://pixabay.com/photos/cat-animal-cute-pet-room-play-3162458/
16.4 https://upload.wikimedia.org/wikipedia/commons/a/a9/John_Forbes_Nash%2C_Jr._by_Peter_Badge.jpg
16.5 https://upload.wikimedia.org/wikipedia/commons/a/a6/Schizophrenia_image.jpg
16.6 https://pixabay.com/photos/hiding-boy-girl-child-young-box-1209131/
16.7 https://wellcomecollection.org/works/wv5n6jjx
16.8 https://upload.wikimedia.org/wikipedia/commons/7/7f/Schizophrenia_risk_factors-PLOS.png
16.9 Amato D, Kruyer A, Samaha A-N and Heinz A (2019) Hypofunctional Dopamine Uptake and Antipsychotic Treatment-Resistant Schizophrenia. Front. Psychiatry 10:314. doi: 10.3389/fpsyt.2019.00314
16.10 Pablo Picasso, The Old Guitarist
16.11 https://commons.wikimedia.org/wiki/File:Light_Therapy_for_SAD.jpg
16.14 https://upload.wikimedia.org/wikipedia/commons/8/89/201407_tail_suspension_test.png
https://upload.wikimedia.org/wikipedia/commons/e/e0/201407_porsolt_forced_swim_test.png
16.15 https://upload.wikimedia.org/wikipedia/commons/f/f0/Bipolar_disorder_subtypes_comparison_between_Bipolar_I%2C_II_disorder_and_Cyclothymia.svg modified by Austin Lim
16.17 https://pixabay.com/photos/crowd-of-people-crowd-football-fans-1488213/
16.18 https://upload.wikimedia.org/wikipedia/commons/a/ab/ElevatedPlusMaze.svg modified by Austin Lim

Orders of magnitude

Number	Power of 10	Name	Prefix (symbol)	Example in biology (approx)
0.000000000000000001	10^{-18}	Quintillionth	atto- (a)	
0.000000000000001	10^{-15}	Quadrillionth	femto- (f)	
0.000000000001	10^{-12}	Trillionth	pico- (p)	
0.000000001	10^{-9}	Billionth	nano- (n)	20 nanometers = synapse length
0.000001	10^{-6}	Millionth	micro- (μ)	10 micrometers = neuron diameter
0.001	10^{-3}	Thousandth	milli- (m)	2 milliseconds = duration of action potential
1	10^{0}			1.7 meters = height of a human
1,000	10^{3}	Thousand	kilo- (k)	20 kilohertz = highest pitch humans can hear
1,000,000	10^{6}	Million	mega- (M)	
1,000,000,000	10^{9}	Billion	giga- (G)	86 billion = neurons in the brain
1,000,000,000,000	10^{12}	Trillion	tera- (T)	150 trillion = synapses in the brain
1,000,000,000,000,000	10^{15}	Quadrillion	peta- (P)	
1,000,000,000,000,000,000	10^{18}	Quintillion	exa- (E)	

Measurement	Unit (short)	Examples
Length	Meters (m)	Length of wavelength of visible light: 500 nanometers Length of a typical neuron: 20 micrometers Height of a typical human: 1.7 meters
Volume	Liters (L)	Volume of cytoplasm in a single spin: 0.1 femtoliters Volume of CSF in ventricles: 150 milliliters
Weight	Grams (g)	Weight of typical brain: 1.3 kilograms Weight of typical grown adult human: 70 kilograms
Time	Second (s)	Duration of action potential: 2 millisecond
Velocity	Meters per second (m/s)	Speed of slow action potential propagation: 0.1 m/s Speed of fast action potential propagation: 100 m/s
Concentration	Molar (M)	Calcium ion concentraion in cell: 100 nanomolar Sodium ion concentration in ACSF: 140 millimolar
Temperature	Celsuis (C) Kelvin (K)	Freezing point of water: 0 Celsius or 273 Kelvin Typical body temperature: 37 Celsius or 310 Kelvin
Electrical potential	Volts (V)	Charge of a typical neuron: -70 millivolts
Current	Ampere (A)	Current passing through single ion channel: 5 picoamps
Resistance	Ohm (Ω)	Typical input resistance of neuron: 200 megaohms
Conductance	Siemens (S)	Conductance through single ion channel: 10 picosiemens
Capacitance	Farad (F)	Capacitance in neuron: 100 picofarads
Magnetic field	Tesla (T)	Strength of fMRI machine: 5 Tesla
Frequency	Hertz (Hz)	Range of human hearing: 20 Hertz to 20 kilohertz
Loudness	Decibels (dB)	Pain threshold: 130 dB Conversation: 65 dB

Neurotransmitters and their receptors

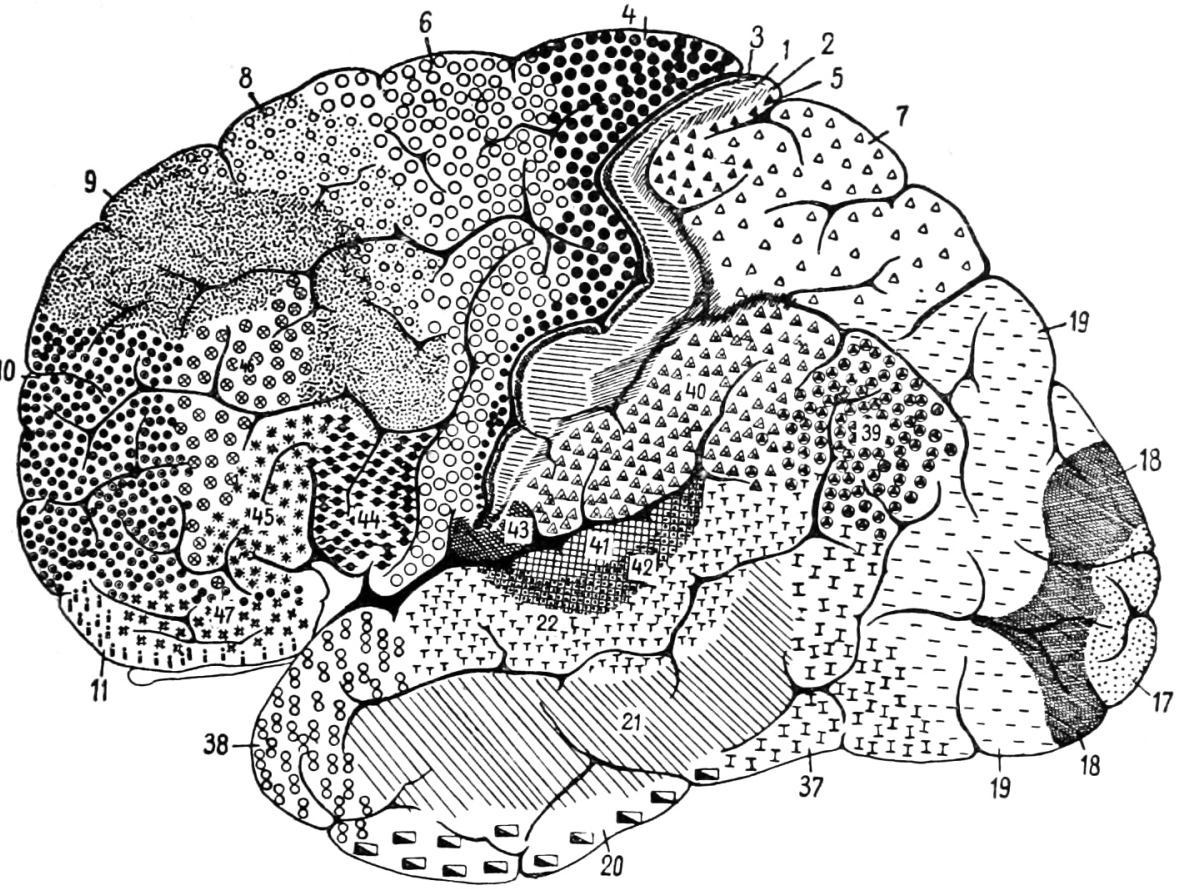


Brodmann areas and functions

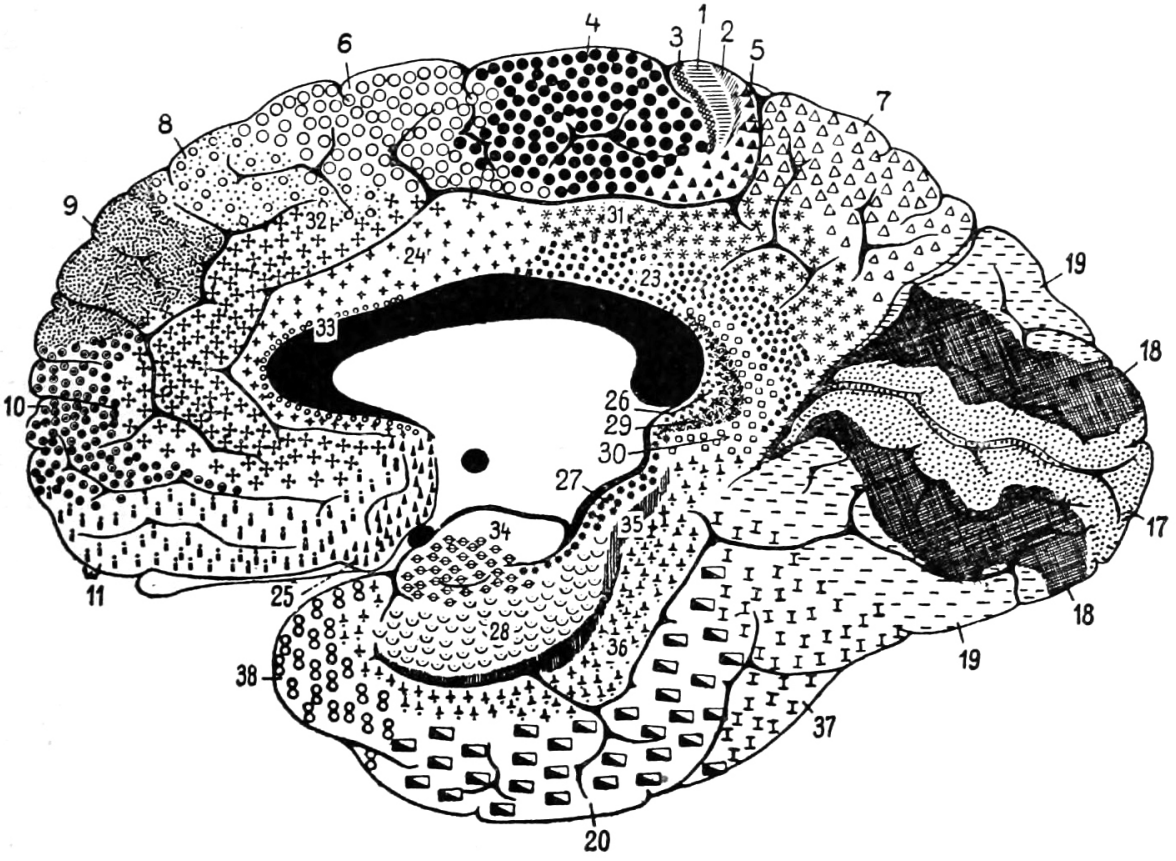
Brodmann area	Name	Function
1, 2, 3	Primary somatosensory cortex (S1)	Touch sensory and perception
4	Primary motor cortex (M1)	Voluntary motor control
5	Superior parietal lobule	Identification of objects based on somato-sensory cues (stereognosis)
6	Premotor; supplemental motor	Limb movement planning
7	Posterior parietal association area	Integration of visual and motor
8	Frontal eye fields	Visual perception and motor, saccades
9, 46	Dorsolateral prefrontal cortex	High order executive functions
10	Anterior prefrontal cortex	High order executive functions
11, 12	Orbitofrontal area	Emotion, decision making
13, 14, 16	Insular cortex	Emotion, empathy, taste, homeostasis
15	Anterior temporal lobe	Social knowledge and memories
17	Primary visual cortex (V1)	Vision, pattern recognition
18	Secondary visual cortex (V2)	Vision, illusory contours
19	Associative visual cortices	Vision, color, motion, depth
20	Inferior temporal gyrus	Visual memory, face perception
21	Middle temporal gyrus	Visual memory, emotional recognition
22	Superior temporal gyrus	Language comprehension, attention, hear-ing
23, 31	Posterior cingulate cortex	Emotions
24, 32, 33	Anterior cingulate cortex	Emotions, attention, decision making
25	Subgenual area	Inhibition of emotion, decision making
26	Ectosplenial area	Emotions
27	Presubiculum	Emotions, head direction
28, 34	Entorhinal cortex	Memory, navigation, smell, emotions
29, 30	Retrosplenial cortex	Memory, navigation
35, 36	Perirhinal cortex	Perception, memory
37	Fusiform gyrus	Facial processing, perception
38	Temporopolar area	Socio-emotional processing, smell
39	Angular gyrus	Reading, speech, perception
40	Supramarginal gyrus	Language perception and processing
41, 42	Auditory cortex (A1)	Hearing
43	Gustatory cortex	Taste
44, 45, 47	Broca's area	Language, movement planning, cognition
48	Retrosubicular area	Memory
49	Parasubicular area	Navigation
52	Parainsular area	

Location of Brodmann areas

Lateral view



Midsagittal view



https://commons.wikimedia.org/wiki/File:Cholinergic_synapse.svg

https://commons.wikimedia.org/wiki/File:Human_Brodmann_areas_(K._Brodmann,_1909,_p._131,_Fig._85-86).jpg

The Open Neuroscience Initiative is funded by a grant from the Vincentian Endowment Fund of DePaul University.