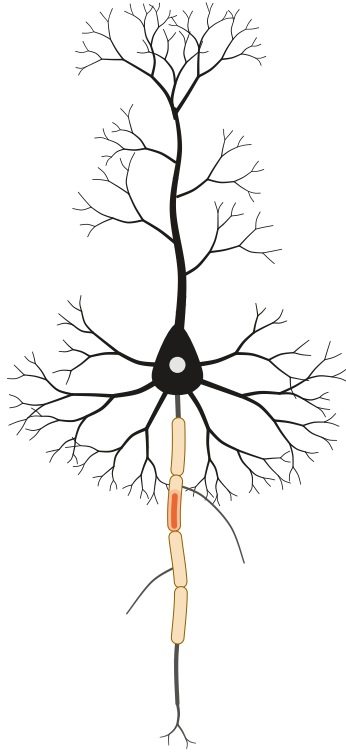




FOR AI ENGINEERS

Neuroscience Book

Javier Luraschi

Neuroscience Book — For AI Engineers

By Javier Luraschi, founder of Hal9 (hal9.com).

Notes on *Principles of Neural Science*, 6th edition, edited by Eric R. Kandel, John D. Koester, Sarah H. Mack, and Steven A. Siegelbaum (McGraw Hill, 2021). All figures are original illustrations.

© 2026 Javier Luraschi. Licensed under CC BY-NC-ND 3.0 US (<https://creativecommons.org/licenses/by-nc-nd/3.0/us/>).

Draft edition, 2026-10-06. The latest version is at neurosciencebook.com.

Contents

- 1 Brain5
 - 1.1 Overview 6
 - 1.2 History.....7
 - 1.3 Components11
 - 1.4 Cortex 14
 - 1.5 Neurons 21
 - 1.6 Signaling 26
 - 1.7 Circuits 30
 - 1.8 Systems 34
 - 1.9 Computation 41
 - 1.10 Genes..... 50
 - 1.11 Behavior 54
 - 1.12 Imaging 59
 - 1.13 Recap..... 65
- 2 Neurons — in progress
- 3 Synapses — in progress
- 4 Cognition — in progress
- 5 Perception — in progress
- 6 Movement — in progress
- 7 Consciousness — in progress
- 8 Behavior — in progress
- 9 Learning — in progress
- 10 Civilization — in progress
- 11 Models — in progress

CHAPTER 1

Brain

The parts, the cells, the wiring, and the principles that organize them

BASED ON

Principles of Neural Science, 6th edition, Part I, "Overall Perspective" (chapters 1–6)

Everything you do — reading this sentence, recognizing a face, reaching for a cup, remembering yesterday — is produced by about 86 billion nerve cells wired into circuits. Those circuits pick out the events that matter from a constant stream of sensory signals, build your perception of the world, make your decisions, and move your body. The working hypothesis of modern neuroscience is that this is the whole story: the mind is a set of operations carried out by the brain, and every mental process, from a reflex to a sense of self, can in principle be explained by how nerve cells signal and how they are connected.

This chapter is the system overview. It walks through the parts of the brain and what each one does, the cells it is made of, the signals those cells exchange, the circuits they form, the computations those circuits perform, how genes build and tune the whole system, and how we can watch it work in a living person. Later chapters zoom into each layer; here the goal is the map.

Along the way we collect the organizing principles that the rest of the book builds on. They are numbered and set apart, starting with Principle 1. Asides labeled *Engineering lens* translate ideas into terms familiar from computing and machine learning, and say where the analogy stops working. Each section notes the chapters of Kandel's *Principles of Neural Science* it summarizes, so you can go deeper.

IN THIS CHAPTER

- **Parts.** The central nervous system has seven major parts. Functions are localized in specific regions, yet every complex behavior is distributed across many of them.
- **Cells.** Neurons are discrete cells that signal in one direction — from dendrites, through the cell body and axon, to synapses — and connect with high specificity. Glia, about as numerous, support them.
- **Signals.** A neuron turns graded analog inputs into all-or-none spikes, and spikes back into graded chemical output. A spike means whatever its pathway means.
- **Circuits.** A few wiring motifs — divergence, convergence, feed-forward and feedback inhibition, recurrence — repeat from reflexes to the cortex, and experience reshapes them.
- **Computation.** Populations of neurons encode information that can be decoded. Hierarchies, expansions, and recurrent loops transform it, and plasticity rules implement unsupervised, supervised, and reinforcement learning.
- **Genes.** Genes build and regulate circuits but do not dictate behavior; environment and experience act through them.
- **Measurement.** Functional MRI shows the human brain at work — indirectly, at the scale of millimeters and seconds.

1.1

Overview

Before looking at any structure, it helps to have a sense of scale. Table 1.1 collects some numbers worth remembering. Treat them as orders of magnitude: most are estimates with real uncertainty, and all vary between individuals.

Table 1.1 The human brain at a glance.

QUANTITY	VALUE	NOTES
Mass	≈ 1.3–1.5 kg	About 2% of body mass
Power	≈ 20 W	About 20% of the body's energy use at rest ¹
Neurons	≈ 86 billion	16 billion in the cerebral cortex, 69 billion in the cerebellum ²
Glia and other cells	≈ 85 billion	About one per neuron in direct counts; Kandel et al. give 2–10 ³
Synapses	≈ 10 ¹⁴ –10 ¹⁵	Typically thousands per neuron
Cell types	Thousands	More than 3,000 molecular subtypes in a recent atlas ⁴
Action potential	≈ 100 mV, ≈ 1 ms	All-or-none
Conduction speed	≈ 1–100 m/s	Fastest in thick, insulated axons

QUANTITY	VALUE	NOTES
Axon length	< 1 mm to > 1 m	The longest run from the spinal cord to the foot
Cerebral cortex	≈ 0.2 m ² , 2–4 mm thick	A folded sheet; about two-thirds is hidden in the folds

1 Raichle ME, Gusnard DA (2002). Appraising the brain's energy budget. *PNAS* 99:10237–10239.

2 Azevedo FAC et al. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology* 513:532–541.

3 von Bartheld CS, Bahney J, Herculano-Houzel S (2016). The search for true numbers of neurons and glial cells in the human brain: a review of 150 years of cell counting. *Journal of Comparative Neurology* 524:3865–3895.

4 Siletti K et al. (2023). Transcriptomic diversity of cell types across the adult human brain. *Science* 382:eadd7046.

Three things stand out to an engineer. The components are slow: a spike lasts about a millisecond, and even the fastest axons carry it at about the speed of a racing car — a million times slower than a signal in a copper wire. The system is frugal: the whole brain runs on the power of a dim light bulb. And the network is huge but sparse: each neuron connects to thousands of others out of tens of billions. Whatever the brain’s advantages are, they are not raw component speed. They are organization, which is what this chapter is about.

1.2

History

For most of history, two views of the brain competed. One held that the brain works as a whole, every part able to do everything. The other held that it is a collection of specialized parts, made of individual cells. The evidence that settled the question also produced the field’s founding principles.

Electricity and chemistry

The idea that nerves work electrically came first. In the 1790s Luigi Galvani showed that frog nerves and muscles not only respond to electricity but produce it themselves. Emil du Bois-Reymond later recorded the electrical activity of active nerves, and in 1850 Hermann von Helmholtz measured how fast a nerve impulse travels: tens of meters per second — fast, but finite. The nerve signal was a physical process that could be timed, not the instantaneous flow of “animal spirits” imagined since antiquity. Pharmacology added chemistry. Claude Bernard, Paul Ehrlich, and John Langley showed that drugs act on cells by binding to specific *receptors*, the first step toward our understanding of chemical communication between neurons.

Localization

In the early 1800s Franz Joseph Gall argued that the brain is not one organ but many — dozens, each responsible for a mental faculty such as generosity or secretiveness — and that the strength of a faculty could be read from bumps on the skull. *Phrenology* was wrong in its method and in its list of faculties, but its central claim, that mental functions are localized, was radical. Pierre Flourens tested it in the 1820s by removing parts of animals' brains. Behavior degraded with how much tissue he removed, not with which part, and he concluded that the cerebral hemispheres work as a whole. This became the *aggregate-field* view.

Clinical observation slowly overturned it. In 1861 Pierre Paul Broca described a patient who understood language but could barely speak — though he could still whistle and sing a melody, so the muscles of speech worked. After the patient died, Broca found damage to the left frontal lobe, and after seeing similar cases he concluded that “we speak with the left hemisphere.” In the same decade John Hughlings Jackson noticed that some epileptic seizures start in one part of the body and spread to others in a fixed order, and inferred that different parts of the cortex control different parts of the body. In 1870 Gustav Fritsch and Eduard Hitzig showed that electrically stimulating specific spots of a dog's cortex moves specific limbs. And in 1874 Carl Wernicke described the complementary language disorder — fluent speech that makes little sense, with poor comprehension — caused by damage farther back, in the temporal lobe (Section 1.4).

The neuron

The argument about cells ran in parallel. By 1839 Matthias Schleiden and Theodor Schwann had established that all tissue is made of cells, but nervous tissue seemed to be an exception: under the microscope it looked like a continuous tangle. In 1873 Camillo Golgi discovered a silver stain that fills only a small, random fraction of neurons — but fills each of them completely, so that a single cell stands out with all its branches. Golgi concluded that nerve cells are fused into one continuous network. Santiago Ramón y Cajal, using Golgi's own stain, concluded the opposite: every neuron is a discrete cell, and neurons contact one another only at specific points. Golgi and Cajal shared the 1906 Nobel Prize and used their lectures to disagree. Cajal was right. In 1907 Ross Harrison watched axons grow out of individual nerve cells kept alive in a dish, and in the 1950s the electron microscope finally showed the narrow gap that separates neurons at their contacts.

PRINCIPLE 1**Neuron doctrine**

The nervous system is made of discrete cells — neurons — which are its elementary signaling units. Neurons are not fused into a continuous net: they communicate at specialized contacts called synapses.

Cajal went further. He and Charles Sherrington, who named the synapse in 1897, carried the clinicians' regional findings down to the level of cells, championing what Kandel calls *cellular connectionism*: behavior is produced by specific groups of neurons connected in specific ways.

PRINCIPLE 2**Cellular connectionism**

Individual neurons are the signaling units of the brain. They are arranged in functional groups and connect to one another in a precise way. The computations behind behavior live in those groups and their connections.

Even so, the aggregate-field view dominated experiments and clinical practice until about 1950. Its champions included the neurologist Henry Head and the physiologist Ivan Pavlov, but the most influential was Karl Lashley. In the 1920s Lashley trained rats to run mazes, removed varying amounts of their cortex, and found — as Flourens had — that performance depended on how much cortex was removed, not where. His *theory of mass action* held that what matters for learning and memory is the total mass of cortex, not any particular region. The explanation turned out to be the task: a rat solves a maze with vision, touch, smell, and movement all at once, so no single lesion abolishes the skill, but every lesion degrades it. Meanwhile, electrical recording and stimulation in the 1930s and 1940s revealed orderly maps of the body in the cortex of cats, monkeys, and people (Section 1.8). Localization won — with a twist that we return to in Section 1.4.

ENGINEERING LENS**Ablation studies and their trap**

Lashley's experiments are the ancestors of the ablation studies used to interpret deep networks: remove a unit, a layer, or an attention head, and measure what breaks. They share a trap. When a capability is distributed and redundant, no single ablation abolishes it and every ablation degrades it a little — exactly the pattern Lashley saw. That pattern says nothing about whether the parts are specialized. To find specialization you need a task narrow enough to depend on one component, or a double dissociation (Section 1.4).

The milestones behind this chapter, in order:

1791 Galvani: nerves and muscles produce electricity.

1810 Gall: mental functions are localized in the brain (phrenology).

1824 Flourens: lesion experiments suggest the hemispheres act as a whole.

1850 Helmholtz measures the speed of the nerve impulse.

1861 Broca identifies a frontal region for speaking.

1873 Golgi's silver stain reveals whole neurons.

1874 Wernicke identifies a temporal region for understanding speech.

1888 Cajal: neurons are discrete cells.

1897 Sherrington names the synapse.

1909 Brodmann maps the cortex into about fifty areas.

1928 Adrian: nerve impulses are all-or-none and encode intensity in their rate.

1929 Lashley proposes the theory of mass action.

1937 Penfield maps the body onto the cortex.

1949 Hebb proposes a rule for learning at synapses.

1952 Hodgkin and Huxley explain how the action potential is generated.

1962 Hubel and Wiesel describe feature-selective neurons in the visual cortex; split-brain studies begin.

1971 Konopka and Benzer find a gene that sets the circadian clock; O'Keefe and Dostrovsky discover place cells.

1973 Bliss and Lømo discover long-term potentiation.

1990 Ogawa describes the BOLD signal, the basis of functional MRI.

2005 Optogenetics makes it possible to switch on genetically chosen neurons with light.

1.3

Components

The nervous system has two divisions. The *central nervous system* (CNS) is the brain and spinal cord. The *peripheral nervous system* (PNS) is everything outside them: the nerves and clusters of nerve cells (*ganglia*) that connect the CNS to the skin, muscles, and internal organs (Section 1.8). Kandel divides the CNS into seven major parts (Figure 1.1). From the bottom up:

- **Spinal cord.** The lowest part of the CNS, running inside the vertebral column. It receives sensory information from the skin, joints, and muscles of the trunk and limbs, and it contains the motor neurons that command their muscles. It is not just a cable: reflexes, and even the basic rhythm of walking, are generated by circuits inside it. Top to bottom, it has four regions — cervical, thoracic, lumbar, and sacral — made of 8, 12, 5, and 5 segments, each with its own pair of nerve roots on either side.
- **Medulla oblongata.** The lowest part of the brain, continuous with the spinal cord. It contains centers for vital autonomic (automatic) functions such as breathing, heart rate, blood pressure, and digestion.
- **Pons.** Above the medulla. It relays information about movement from the cerebral hemispheres to the cerebellum, and it takes part in controlling breathing and sleep.
- **Midbrain.** Above the pons. It controls many sensory and motor functions, including eye movements and the coordination of visual and auditory reflexes.
- **Cerebellum.** Behind the pons and medulla, attached to the brain stem by three pairs of thick fiber bundles, the *cerebellar peduncles*. It modulates the force and range of movements, refines their timing, and is central to learning motor skills. Although it is small, it holds about 80% of the brain's neurons (Section 1.9).
- **Diencephalon.** Two structures above the midbrain. The *thalamus* processes and relays most of the information that reaches the cerebral cortex from the rest of the CNS. The *hypothalamus*, below it, regulates autonomic, endocrine, and visceral functions — body temperature, hunger, thirst, sleep, and reproduction — partly through the hormones of the pituitary gland and partly through its projections to autonomic centers in the brain stem and spinal cord.
- **Cerebrum.** The two cerebral hemispheres: the *cerebral cortex*, the white matter beneath it, and three deep structures. The *basal ganglia* regulate motor performance, select actions, and learn habits; the *hippocampus* is essential for forming memories of people, places, things, and events; the *amygdala* co-

ordinates the autonomic and hormonal responses of emotional states, including memories of threats. A thick bundle of axons, the *corpus callosum*, connects the two hemispheres.

The medulla, pons, and midbrain together form the *brain stem*. Besides its own functions, the brain stem carries all the traffic between the brain and the spinal cord, receives sensory information from the head and controls the muscles of the face and neck through the cranial nerves, and contains networks of neurons that regulate arousal and awareness.

The parts differ in how their neurons are arranged. In the brain stem and diencephalon, neurons are grouped in clusters called *nuclei*, each with its own connections. In the cerebrum and cerebellum, the surface is a folded sheet of neurons arranged in layers — a *cortex* — with stereotyped patterns of connection.

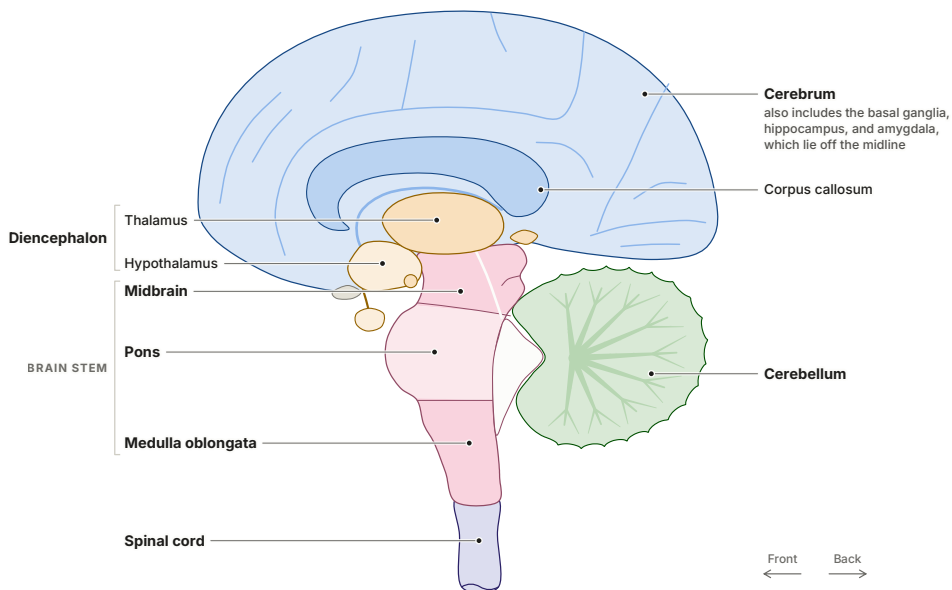


Figure 1.1 The seven major parts of the central nervous system, in a section along the midline of the brain. The medulla, pons, and midbrain form the brain stem. Three structures of the cerebrum — the basal ganglia, hippocampus, and amygdala — lie deep in each hemisphere, away from the midline, so they do not appear in this section.

BOX 1.1

Neuroanatomy's coordinate system

Anatomy locates structures along three axes and cuts sections in three planes (Figure 1.2).

- **Rostral–caudal:** toward the nose (Latin *rostrum*, beak) or toward the tail (*cauda*).
- **Dorsal–ventral:** toward the back (*dorsum*) or toward the belly (*venter*).
- **Medial–lateral:** toward the midline or toward the side.

A *sagittal* section is vertical and runs front to back; the *midsagittal* section splits the brain into left and right halves, as in Figure 1.1. A *coronal* (or frontal) section is vertical and runs side to side. A *horizontal* (or axial) section is parallel to the ground.

One complication: because we stand upright, the long axis of the CNS bends where the midbrain meets the diencephalon. In the spinal cord and brain stem, *dorsal* points toward the back and *rostral* points up. In the forebrain, *dorsal* points up and *rostral* points toward the face.

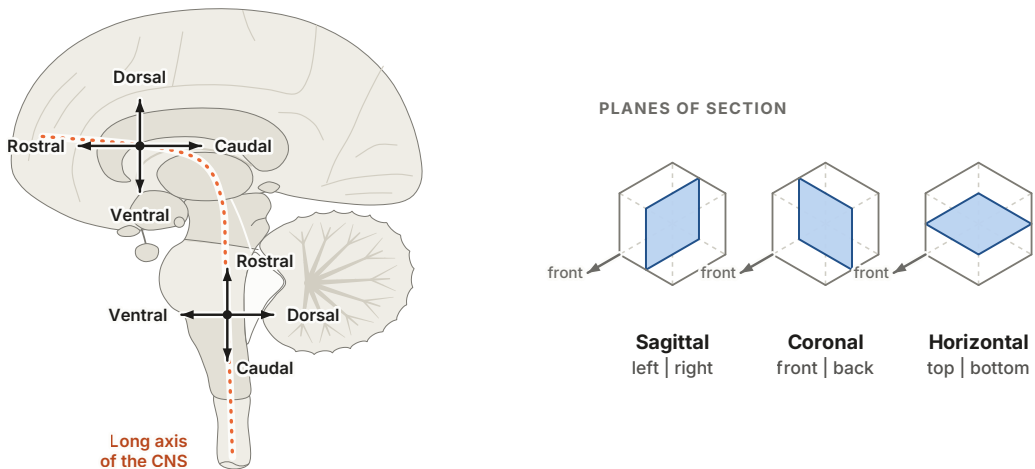


Figure 1.2 Axes and planes of the central nervous system. In humans the rostral–caudal axis is flexed by about 110° where the brain stem meets the forebrain, and the meanings of *dorsal* and *ventral* bend with it.

ENGINEERING LENS

Layered control

The CNS looks like a system that grew by adding layers rather than rewriting old ones. The spinal cord runs reflexes and the rhythm of stepping on its own; the brain stem keeps you breathing and awake; the cortex adds flexible, learned control on top. Hughlings Jackson noticed in the nineteenth century that damage to a higher level can *release* older behaviors that it normally suppresses. In adults, for example, stroking the sole of the foot curls the toes down; after damage to the pathway from the motor cortex, the big toe extends upward instead, as it does in infants (the Babinski sign). Roboticists will recognize the idea in Rodney Brooks's *subsumption architecture*, which builds robots from layers of simple behaviors, each able to override the one below. The analogy breaks at the wiring: the brain's layers are densely and reciprocally connected, not a clean stack.

1.4

Cortex

The cerebral cortex, the part of the brain most highly developed in humans, computes our most elaborate abilities: perceiving objects, using language, planning, reasoning. It is a sheet of neurons only 2–4 mm thick — the *gray matter* — folded so that about 0.2 m² of it fits inside the skull (Figure 1.3). Beneath the sheet lies the *white matter*: bundles of axons that connect each part of the sheet with other parts and with the structures below. To an engineer the layout is familiar: the processing elements sit in a thin layer, and much of the volume is taken up by the cabling between them.

Folding is how evolution packs a large sheet into a limited space. The ridges of the folds are called *gyri* and the grooves *sulci*; about two-thirds of the cortex lies hidden in the sulci, and the deepest grooves are called *fissures*. Two of them serve as landmarks. The *central sulcus* runs from the top of each hemisphere down its side, and the *lateral* (or Sylvian) *fissure* separates the temporal lobe, below it, from the frontal and parietal lobes above.

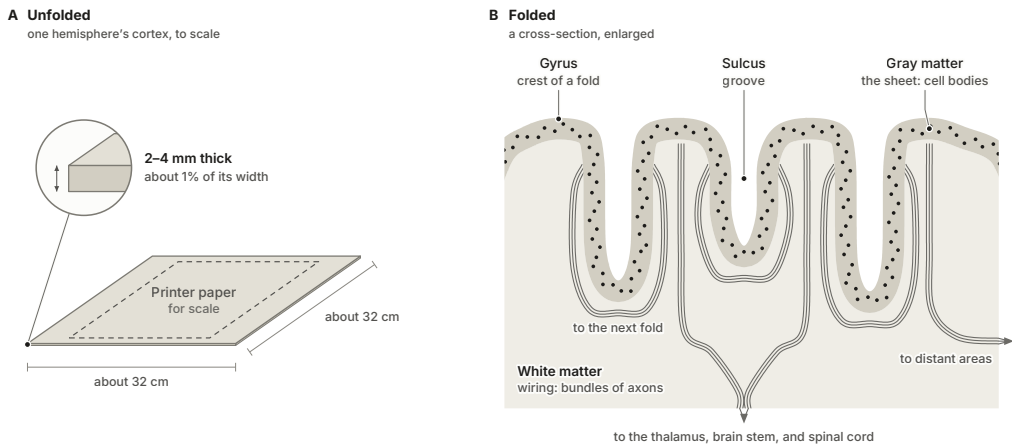


Figure 1.3 The cortex is a folded sheet. A. Unfolded, the cortex of one hemisphere would be a sheet about 32 cm on a side but only 2–4 mm thick, drawn here to scale around a sheet of printer paper. B. Folded into gyri and sulci (enlarged), it fits inside the skull. The sheet itself, the gray matter, holds the cell bodies of neurons; beneath it, the white matter is wiring: bundles of axons that loop under a sulcus to the next fold, run to distant areas, or descend to structures below the cortex.

These landmarks divide each hemisphere into four lobes, named after the bones of the skull that cover them (Figure 1.4):

- The **frontal lobe** is concerned with planning future actions, working memory, and the control of movement. The *primary motor cortex* occupies the *precentral gyrus*, just in front of the central sulcus.
- The **parietal lobe** handles somatic sensation (touch, pressure, the position of the body), the body's image, and its relation to the space around it. The *primary somatosensory cortex* occupies the *postcentral gyrus*, just behind the central sulcus.
- The **occipital lobe**, at the back, is devoted to vision. From there, visual information flows along two main streams: a *dorsal stream* into the parietal lobe, concerned with where objects are and how to act on them, and a *ventral stream* into the temporal lobe, concerned with what they are — including whose face you are looking at.
- The **temporal lobe** is concerned with hearing and — through the hippocampus and amygdala, deep inside it — with learning, memory, and emotion. Its underside also performs high-level vision, such as recognizing objects and faces.

Two more cortical regions are hidden from the outside: the *insula*, folded inside the lateral fissure, which takes part in emotion, the sense of the body's internal state, and taste; and the *cingulate cortex*, on the inner surface of each hemisphere above the corpus callosum, which takes part in emotion, pain, and the control of behavior.

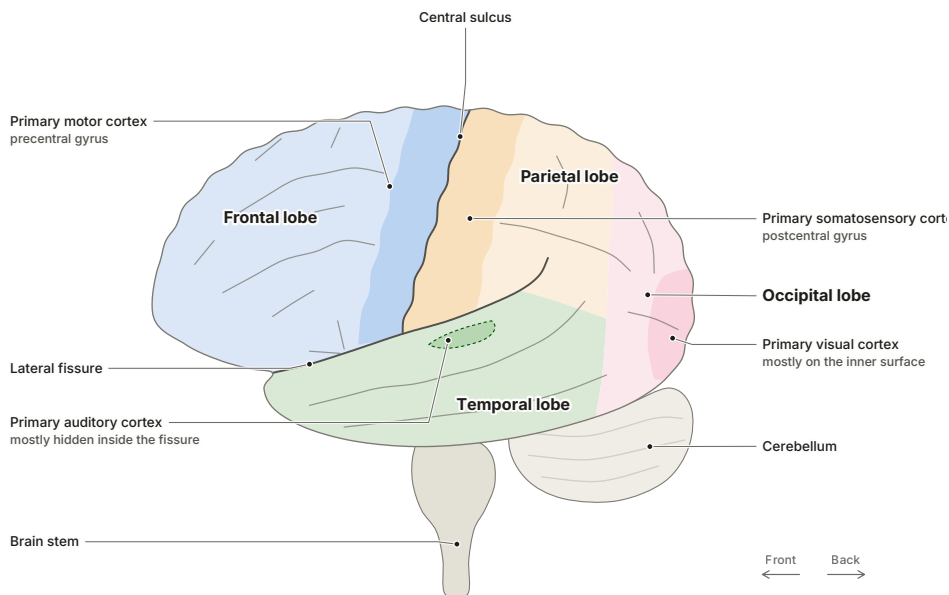


Figure 1.4 The four lobes of the left cerebral hemisphere and some landmark areas, seen from the side. The primary motor and somatosensory areas flank the central sulcus; the primary auditory cortex lies on the upper surface of the temporal lobe, mostly hidden inside the lateral fissure.

Three facts of organization apply throughout. Each hemisphere is concerned mostly with the opposite side of the body: the left hemisphere feels and moves the right hand, because the pathways cross the midline on the way (Section 1.8). The two hemispheres are neither perfectly symmetrical in structure nor equivalent in function. And each lobe contains several specialized areas: *primary* areas are the first cortical stop of a sense or the last cortical stage of motor output, and the *association* areas around them combine information into more abstract representations.

Language: the first evidence for localization

Language provided the first strong evidence that specific mental abilities depend on specific regions, through the study of *aphasias* — disorders of language caused by brain damage.

Broca's patients had what is now called *Broca's aphasia* (expressive, or nonfluent, aphasia): they understood language but spoke slowly and with great effort, in short, ungrammatical phrases. The damage was in the lower rear part of the left frontal lobe, in *Broca's area*, next to the motor cortex that controls the face and mouth. Wernicke's patients showed the opposite pattern, *Wernicke's aphasia* (receptive, or fluent, aphasia): they spoke fluently, but their speech made little sense and they could not understand what they heard. The damage was at the back of the left temporal lobe, in what is now *Wernicke's area*. Together, the two syndromes form a *double dissociation* — some patients can understand but not speak, others can speak but not understand — strong evidence that the two abilities depend on separate components.

Wernicke went beyond localization: he proposed how the components work together. Simple perceptual and motor functions, he argued, are localized in single areas, but complex abilities arise from the interconnections between several functional sites. In his model, as later extended by Norman Geschwind, a heard word is processed by the auditory cortex, recognized in Wernicke's area, and sent through a bundle of axons, the *arcuate fasciculus*, to Broca's area, which turns it into a program of articulation that the motor cortex executes. A written word enters by a different route, proposed by Jules Dejerine in the 1890s: from the visual cortex through the *angular gyrus*, where the visual and auditory forms of a word are associated, so that written and spoken words converge on a common code (Figure 1.5).

The model made a prediction: damage to the connection, sparing both areas, should spare comprehension and articulation but break the link between them. That disorder exists. Patients with *conduction aphasia* understand what they hear and read, and have no trouble moving the muscles of speech. Yet their speech is full of errors — dropped syllables, wrong sounds — they struggle to repeat words and phrases verbatim, and their verbal working memory is severely limited.

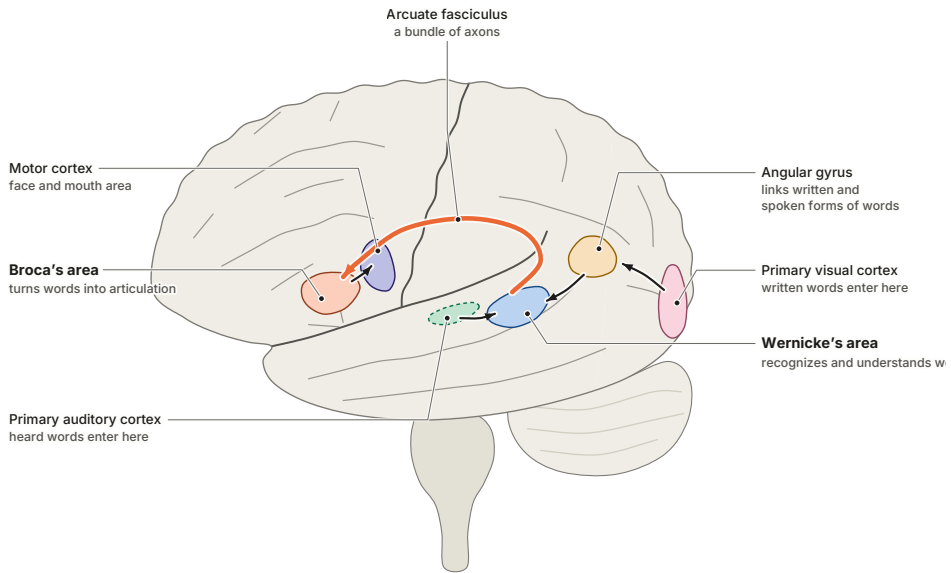


Figure 1.5 The classic model of language processing. To repeat a heard word, information travels from the auditory cortex to Wernicke's area, through the arcuate fasciculus to Broca's area, and on to the motor cortex. To read a word aloud, it enters from the visual cortex through the angular gyrus. Modern imaging shows a broader network, but the model's central idea — specialized nodes linked by pathways — has held.

The Wernicke–Geschwind model is now seen as a simplification. Language engages a much broader network that includes more of the frontal, temporal, and parietal cortex and structures beneath it, and imaging has revised the nodes themselves: Broca's area takes part in comprehension as well as speech, and reading also relies on a specialized region on the underside of the occipital and temporal lobes. But its central idea — that a complex ability is a network of specialized components, and that a broken connection is a failure of its own kind — became a pillar of neuroscience, now called *distributed processing*.

What is localized

If the language areas were specialized for sound, sign languages, which are visual and spatial, should depend on other regions — perhaps on the right hemisphere, which dominates spatial processing. They do not. Deaf people who use sign language develop sign-language aphasia after damage to the left hemisphere, and these resemble spoken-language aphasia: damage near Broca's area impairs producing signs, and damage near Wernicke's area impairs understand-

ing them. Damage to the right hemisphere impairs their spatial abilities but not their signing.⁵¹ The left-hemisphere language areas are specialized for language itself, whatever its sensory form.

Imaging refines the picture. When people who learned a second language as adults speak each of their languages, the two activate adjacent but separate regions within Broca's area; in people who learned both languages in infancy, the regions overlap. Wernicke's area shows overlap in both groups.⁶²

Nor is all of language left-sided. The emotional tone of speech, its *prosody*, depends mostly on the right hemisphere, in regions that mirror the classic language areas. Damage to the right-hemisphere counterpart of Wernicke's area impairs recognizing the emotion in other people's voices; damage to the counterpart of Broca's area produces flat, emotionless speech.⁷³

So is function localized or distributed? Both, at different scales. The specialized units are not whole faculties such as "language" or "memory," as phrenology imagined; they are elementary operations. A faculty is assembled from many such operations, carried out in many regions, working together.

The brain also carves up tasks differently than introspection suggests. What you know about an apple — how it looks, how you hold it, what it is called — is stored in separate regions and bound together when you recall it, and the color and the motion of a single object are analyzed in different pathways. Experience feels unified; the machinery behind it is not.

PRINCIPLE 3

Localization of function

Specific regions of the brain carry out specific elementary operations, not whole faculties. Because many operations are supported by more than one region or pathway, damaging one region often degrades a behavior rather than abolishing it.

⁵ Hickok G, Bellugi U, Klima ES (1996). The neurobiology of sign language and its implications for the neural basis of language. *Nature* 381:699–702.

⁶ Kim KHS, Relkin NR, Lee K-M, Hirsch J (1997). Distinct cortical areas associated with native and second languages. *Nature* 388:171–174.

⁷ Ross ED (1981). The aprosodias: functional-anatomic organization of the affective components of language in the right hemisphere. *Archives of Neurology* 38:561–569.

PRINCIPLE 4**Distributed processing**

Every complex behavior is carried out by many specialized regions working in series and in parallel, linked by bundles of axons. A function lives in a network rather than in a spot: parallel pathways make it robust to the loss of any one of them, and function can partly return as the network reorganizes — yet a broken connection can be as damaging as a broken region.

ENGINEERING LENS**Reverse-engineering from failures**

Neuropsychology infers architecture the way you might debug a system you cannot open: by cataloging how it fails. A *dissociation* — a patient loses ability A but keeps B — suggests that A depends on a component that B does not. A *double dissociation* — another patient loses B but keeps A — makes the case much stronger, because it rules out the possibility that A was simply the harder task. Broca's and Wernicke's aphasia are the textbook example. The method has pitfalls familiar from fault analysis: lesions are rarely clean, the system reconfigures itself after damage, and a lesion can disturb distant regions that depend on the damaged one (*diaschisis*), so the observed failure is not always where the fault is.

Two hemispheres, one mind

Distributed processing raises a puzzle: if a behavior is computed in many regions working in parallel, how do the pieces come together into a single experience? A striking clue comes from patients whose corpus callosum — some 200 million axons linking the hemispheres — was cut to stop the spread of severe epileptic seizures. In the 1960s Roger Sperry, Michael Gazzaniga, and Joseph Bogen showed that in these *split-brain* patients each hemisphere can perceive, learn, and remember on its own, largely unaware of the other.⁸⁴ A word shown only to the right hemisphere (by flashing it in the left half of the visual field) cannot be read aloud, because speech depends on the left hemisphere; yet the left hand, which the right hemisphere controls, can pick out the named object by touch. Asked why the hand chose that object, the speaking left hemisphere, which never saw the word, confidently invents a reason. Occasionally the two

⁸ Gazzaniga MS (2005). Forty-five years of split-brain research and still going strong. *Nature Reviews Neuroscience* 6:653–659.

hemispheres act at cross purposes, as when one hand buttons a shirt while the other unbuttons it. Our sense of a single, unified mind depends on continuous communication between the parts.

Split-brain patients also show how much of the brain's work never reaches awareness. Most of its processing is unconscious, and Kandel warns that describing perception and reasoning in terms of conscious experience can mislead: we perceive an object less as a picture than as an opportunity — something we might eat, sit on, or pick up. One consequence is encouraging for science. If much of cognition does not depend on awareness, then studying simpler behaviors in animals can illuminate it.

1.5

Neurons

The brain has two classes of cells: *neurons* (nerve cells), which carry signals, and *glial cells* (glia), which support them in many ways. In most respects a neuron is an ordinary cell: it has a nucleus that holds its genes and machinery that makes its proteins and energy. What sets it apart is its shape and its membrane. A neuron is *polarized* — it has a distinct input end and output end — and its membrane can generate and conduct electrical signals.

A typical neuron has four regions, each with its own job (Figure 1.6):

- **Cell body (soma).** The metabolic center. It contains the *nucleus* and the endoplasmic reticulum, where the cell's proteins are made. Two kinds of processes grow out of it: several short *dendrites* and a single long *axon*.
- **Dendrites.** Branching like a tree (Greek *déndron*), they are the neuron's main receiving surface: most of the synapses from other neurons land on them.
- **Axon.** The output cable. It begins at a tapered region of the cell body, the *axon hillock*, followed by the *initial segment*, where outgoing signals are triggered. It carries *action potentials* — brief electrical impulses — over distances from a fraction of a millimeter to more than a meter. Many axons are wrapped in *myelin*, an insulating sheath made by glial cells, which is interrupted at regular intervals by short gaps called *nodes of Ranvier*. The action potential is regenerated at each node as it travels.
- **Presynaptic terminals.** Near its end, the axon divides into fine branches that contact other cells at *synapses*. The transmitting cell is called *presynaptic* and the receiving cell *postsynaptic*; between them is a gap of 20–40 nanometers, the *synaptic cleft*. Most terminals contact the dendrites or the cell body of an-

other neuron, and a few its axon; others contact muscle or gland cells. Some neurons, called *neuroendocrine cells*, release their transmitter into the bloodstream, where it acts as a hormone.

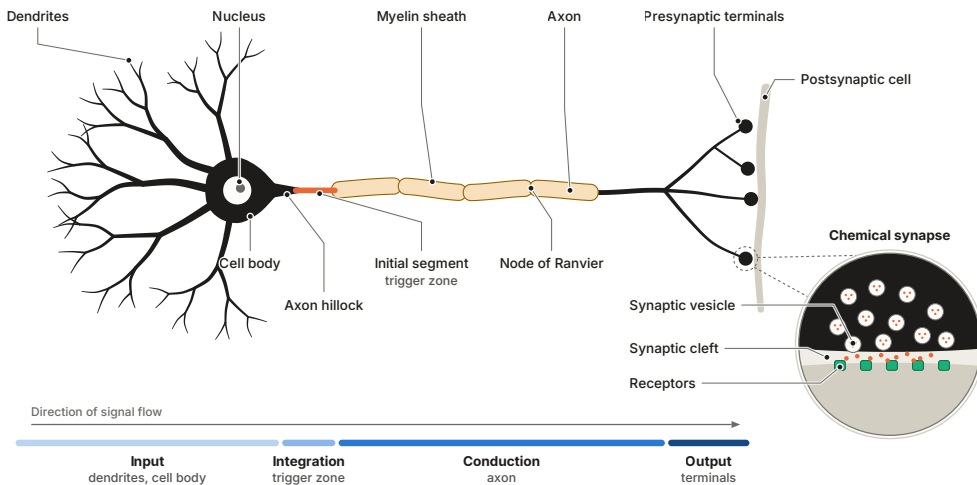


Figure 1.6 Anatomy of a neuron. Signals arrive on the dendrites and cell body, are integrated at the initial segment of the axon, travel along the axon as action potentials — regenerated at each node of Ranvier between segments of myelin — and are passed on at the presynaptic terminals. The inset shows a chemical synapse: the presynaptic terminal releases neurotransmitter from vesicles into the synaptic cleft, where it binds receptors on the postsynaptic cell.

From the shapes of neurons, Cajal deduced two more principles. Studying circuits in the retina, the olfactory bulb, the spinal cord, and elsewhere, he noticed that dendrites always face the direction from which signals arrive and axons the direction in which they leave.

PRINCIPLE 5

Dynamic polarization

Signals flow through a neuron in one predictable direction: from its receiving sites, the dendrites and cell body, to the trigger zone at the start of the axon, and from there along the axon to the presynaptic terminals. Because of this, a circuit can be read like a directed graph.

PRINCIPLE 6**Connectional specificity**

Neurons do not form random networks. Each neuron contacts particular postsynaptic cells, and not others, at particular sites. The wiring diagram is specific: set up during development and refined by experience.

Kinds of neurons

The classic way to classify neurons is by shape — specifically, by the number of processes that leave the cell body (Figure 1.7):

- **Unipolar** neurons have a single process that branches into receiving segments (dendrites) and a transmitting segment (axon). They predominate in invertebrate nervous systems; in vertebrates they occur in the autonomic nervous system.
- **Bipolar** neurons have two processes: a dendrite that receives signals from other cells and an axon that carries them toward the CNS. Many sensory cells are bipolar, including those of the retina and the sensory neurons of the nose.
- **Pseudo-unipolar** neurons start out as bipolar cells during development, but their two processes fuse into a single stem that splits into two long branches, both of which work as axons: one runs to the skin, a joint, or a muscle; the other runs into the spinal cord. These sensory neurons, whose cell bodies lie in the *dorsal root ganglia* beside the spinal cord, carry touch, pressure, body position, and pain. Their signals can travel from one branch to the other without passing through the cell body.
- **Multipolar** neurons have one axon and many dendrites, and they make up most of the vertebrate nervous system. Their shapes vary with the number of inputs they integrate. A *spinal motor neuron*, with a relatively modest dendritic tree, receives about 10,000 synapses — some 1,000 on its cell body and 9,000 on its dendrites. A *pyramidal cell* of the cortex or hippocampus, named for the shape of its cell body, has two dendritic trees, one at its apex and one at its base; a single pyramidal cell in the hippocampus receives about 30,000 synapses.⁹⁵ A *Purkinje cell* of the cerebellum spreads a flat, fan-shaped dendritic tree that receives as many as a million synapses, by Kandel's estimate; careful counts in the rat find about 175,000 from a single class of input alone.¹⁰⁶

⁹ Megías M, Emri Z, Freund TF, Gulyás AI (2001). Total number and distribution of inhibitory and excitatory synapses on hippocampal CA1 pyramidal cells. *Neuroscience* 102:527–540.

¹⁰ Napper RMA, Harvey RJ (1988). Number of parallel fiber synapses on an individual Purkinje cell in the cerebellum of the rat. *Journal of Comparative Neurology* 274:168–177.

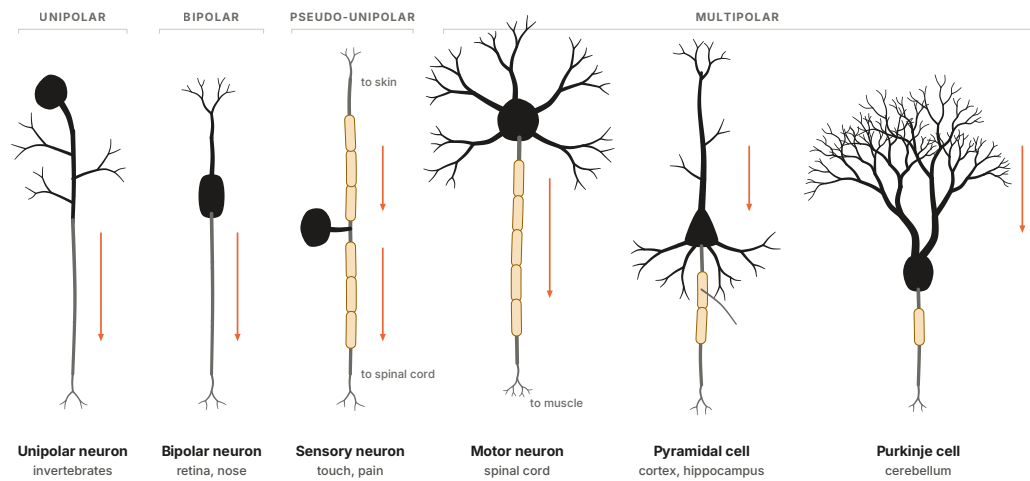


Figure 1.7 Neurons classified by the number of processes that leave the cell body. Cell bodies and dendrites are drawn in black and axons in gray; cream segments are myelin, and arrows show the direction of signal flow. Sizes are not to scale: a Purkinje cell's dendritic tree is a few hundred micrometers across, while a motor neuron's axon can be a meter long.

Neurons can also be classified by their role. *Sensory* (afferent) neurons carry information from the body's periphery into the CNS. *Motor* (efferent) neurons carry commands out to muscles and glands. *Interneurons*, by far the most numerous, do everything in between: *projection* (or relay) interneurons have long axons that connect distant regions, and *local* interneurons have short axons that work within a small circuit. Today neurons are increasingly classified by the genes they express, which has revealed thousands of types.

Glia

Glia are often said to vastly outnumber neurons — Kandel gives 2 to 10 glial cells per neuron in the vertebrate CNS — but careful counts of the human brain find roughly one per neuron.³⁷ Glia are not electrically excitable and do not carry fast signals, yet the brain cannot work without them. They also guide wiring during development and help stabilize connections changed by learning; once seen as mere support, they are now regarded as partners of neurons. They fall

3 von Bartheld CS, Bahney J, Herculano-Houzel S (2016). The search for true numbers of neurons and glial cells in the human brain: a review of 150 years of cell counting. *Journal of Comparative Neurology* 524:3865–3895.

into two classes. *Microglia* are the brain's resident immune cells: they respond to injury and infection, and during development they help remove unneeded synapses. *Macroglia* comprise three main types (Figure 1.8):

- **Oligodendrocytes**, in the CNS, and **Schwann cells**, in the PNS, make myelin. An oligodendrocyte extends many arms, each wrapping a segment of a different axon; a Schwann cell wraps a single segment of a single axon. Diseases that destroy myelin – multiple sclerosis in the CNS, Guillain-Barré syndrome in the PNS – slow or block conduction.
- **Astrocytes**, star-shaped cells, contact both blood vessels and synapses. They wrap synapses and clear away the neurotransmitter released there, which keeps neighboring synapses isolated from one another; absorb the potassium that active neurons release; supply neurons with fuel and release growth factors; help match blood flow to neural activity; and help regulate how synapses form and work.

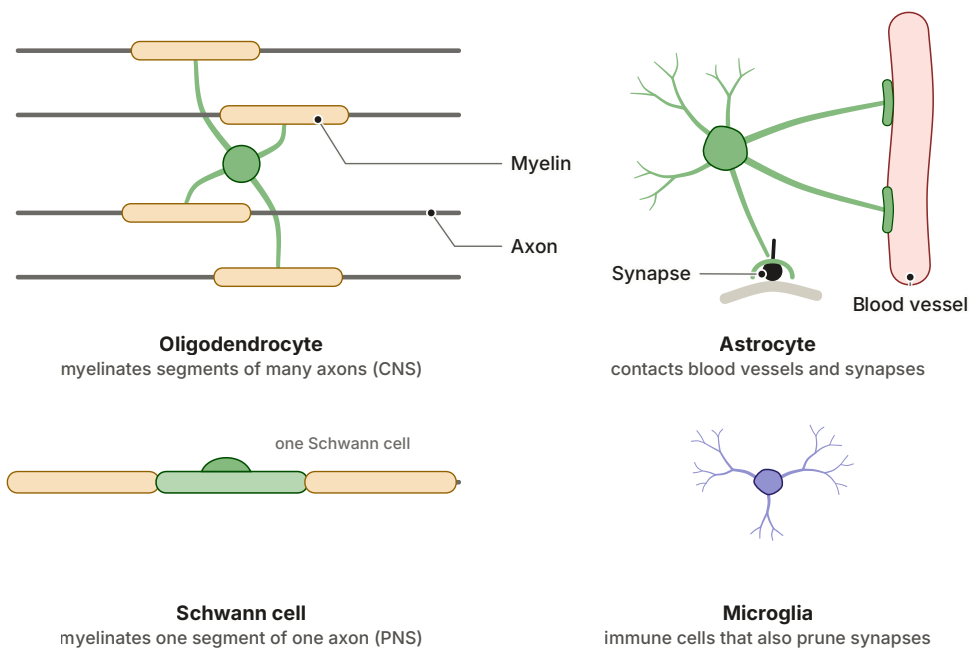


Figure 1.8 The main types of glia. Oligodendrocytes myelinate segments of many axons in the CNS, whereas each Schwann cell myelinates a single segment of one axon in the PNS. Astrocytes link blood vessels and synapses. Microglia survey the tissue for damage.

ENGINEERING LENS

The artificial neuron is a caricature

Artificial neural networks descend from a 1943 model by Warren McCulloch and Walter Pitts: a unit sums its weighted inputs and fires if the sum crosses a threshold. The abstraction keeps the essence of dynamic polarization — many inputs, one decision, one output — and Frank Rosenblatt's perceptron (1958) made its weights learnable. But it leaves out a lot. Dendrites are not passive wires: they produce their own nonlinear responses, so a single pyramidal neuron can behave more like a small two-layer network than like one unit.¹¹⁸ Real neurons are dynamical systems that integrate their inputs over time and fire in rhythms and bursts (Section 1.6). Most neurons are either excitatory or inhibitory at all of their outputs, a constraint that artificial weights, free to take either sign, do not obey. And where a typical network uses one kind of unit, a brain uses thousands.

1.6

Signaling

Almost every neuron, whatever its shape or chemistry, signals with the same sequence of events. Understanding it is the key to reading any circuit.

At rest, the inside of a neuron is about 65 millivolts more negative than the outside. This *resting membrane potential* is maintained by pumps that keep sodium and potassium ions at different concentrations inside and outside the cell, and by channels that let potassium leak out (the next chapter explains the mechanism). A signal is a change in this potential. Making the inside less negative, *depolarization*, brings a neuron closer to firing; making it more negative, *hyperpolarization*, moves it away.

Four components, four signals

Kandel's model of the neuron has four functional components, each producing a different kind of signal (Figure 1.9):

1. **Input.** A stimulus, for a sensory neuron, or a neurotransmitter, for any other neuron, opens ion channels and produces a small, local change in the membrane potential: a *receptor potential* or a *synaptic potential*. These signals are *graded* — their amplitude and duration track the strength and duration of their cause — and they spread passively, fading within a few millimeters at most.

¹¹ Poirazi P, Brannon T, Mel BW (2003). Pyramidal neuron as two-layer neural network. *Neuron* 37:989–999.

2. **Integration.** The input signals spread to the *trigger zone* at the axon's initial segment, where they add up. The trigger zone has the highest density of the voltage-gated sodium channels that generate action potentials, and it makes the neuron's one decision: if the summed potential crosses a *threshold*, typically about 10 mV above rest, the neuron fires.
3. **Conduction.** The *action potential* is a stereotyped pulse of about 100 mV that lasts about a millisecond. It is produced by positive feedback: depolarization opens sodium channels, sodium ions rush in and depolarize the membrane further, and the potential swings explosively from about -65 to about +40 mV before other channels restore it. It is *all-or-none*: a stimulus below threshold produces none, and every stimulus above threshold produces the same full-size pulse. It regenerates itself as it travels, so it reaches the end of the axon undiminished, at 1 to 100 meters per second. Because every pulse is the same, information is carried by how many pulses there are and when they occur: a stronger stimulus produces a higher firing rate, and a longer stimulus a longer train.
4. **Output.** At the presynaptic terminals, each action potential triggers the release of neurotransmitter, packaged in tiny *synaptic vesicles*. The amount released depends on the number and frequency of the arriving action potentials, so the signal becomes graded again — and starts the sequence over in the next cell.

Not every neuron uses all four components. Some local interneurons, with no axon or only a very short one, never fire action potentials: their graded potentials spread passively to their release sites.

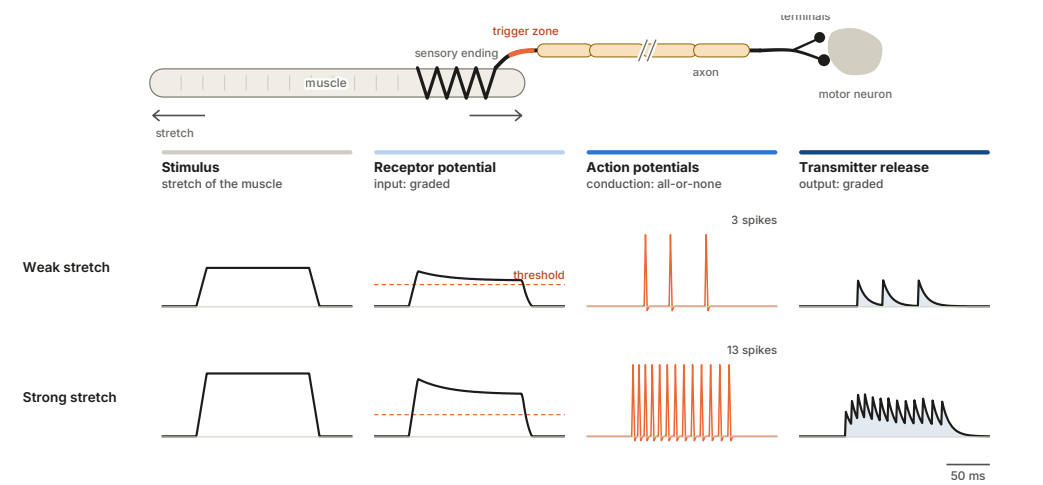


Figure 1.9 The four signals of a sensory neuron. A stretch of the muscle produces a graded receptor potential at the sensory ending. Where it reaches the trigger zone above threshold, it is converted into a train of all-or-none action potentials whose rate reflects the size of the stretch. The action potentials travel to the presynaptic terminals, where they release an amount of transmitter that again depends on their rate.

Table 1.2 Local signals and the action potential. Ranges from Kandel et al., chapter 3.

SIGNAL	AMPLITUDE	DURATION	SUMMATION	EFFECT	PROPAGATION
Receptor potential	Small (0.1–10 mV), graded	5–100 ms	Graded	Depolarizing (most) or hyperpolarizing	Passive
Synaptic potential	Small (0.1–10 mV), graded	5 ms to 20 min	Graded	Depolarizing or hyperpolarizing	Passive
Action potential	Large (70–110 mV), all-or-none	1–10 ms	All-or-none	Depolarizing	Active, regenerating

ENGINEERING LENS

Analog in, digital across, analog out

In engineering terms the neuron is a mixed-signal device. Its inputs are analog, summed in continuous time. The trigger zone is a comparator. The axon is a digital transmission line: all-or-none pulses are regenerated at every node of Ranvier, like a chain of repeaters, so noise does not accumulate along a meter of cable — something an analog signal could never survive. Intensity is carried by pulse rate, a scheme engineers call *pulse-frequency modulation*, and the synapse demodulates it back into an analog quantity. Rate is not the whole code, though: the precise timing of spikes, and their pattern across many neurons, carry information too (Section 1.9).

The meaning is in the pathway

Recording from single sensory fibers in the 1920s, Edgar Adrian noticed something puzzling: action potentials look the same in every nerve. A spike from the eye, a spike from the ear, and a spike commanding a muscle to contract are vir-

tually indistinguishable. How, then, does the brain know whether a spike means light or sound? By its pathway. The meaning of a signal is set by where the fiber that carries it comes from and where it goes.

You can check this yourself. Press gently on the corner of your closed eye and you will see a spot of light: the pressure excites the retina, and because the signals arrive along the visual pathway, they are experienced as light. Cochlear implants exploit the same principle: electrically stimulating fibers of the auditory nerve produces sound, and stimulating different positions along the cochlea produces different pitches.

PRINCIPLE 7

Labeled lines

The information carried by an action potential is determined not by the form of the signal, which is nearly the same everywhere, but by the pathway it travels — the neurons it comes from and the neurons it reaches. The brain reads meaning from which lines are active, and when.

Same input, different output

Neurons differ in how they respond to the same input. Many are silent until they are driven. Others fire spontaneously, with no input at all: regularly — *beating*, or pacemaker, neurons — or in brief, repeated bursts — *bursting* neurons. The same excitatory input may make a silent neuron fire a spike or two but only speed up one that is already active. Inhibition differs even more: it tells a silent cell little, but in an active cell it can carve periods of silence into ongoing firing, creating a pattern where none existed. These differences come from the particular mix of ion channels each neuron expresses.

Synapses

Most synapses are *chemical*: the presynaptic terminal releases a neurotransmitter that diffuses across the cleft and binds *receptors* on the postsynaptic cell. A minority are *electrical*: protein channels called *gap junctions* connect the two cells directly, so current flows between them almost instantly, which helps groups of neurons fire in synchrony. Chemical synapses are slower but far more versatile. They can amplify or invert a signal, they can be modulated, and their strength can change.

Whether a chemical synapse excites or inhibits its target is decided by the postsynaptic receptor, not by the transmitter. The same molecule can do both: acetylcholine excites skeletal muscle through one kind of receptor and slows the

heart through another. In practice, the brain's main excitatory transmitter is *glutamate* and its main inhibitory transmitter is *GABA*. A neuron receiving both kinds of input sums them at its trigger zone, so inhibition can veto excitation.

Neurons differ most in their molecules

Neurons that look alike can be very different molecularly: they use different transmitters, receptors, and ion channels. This molecular diversity is why drugs and diseases can be so selective. In Parkinson's disease, for example, a small population of dopamine-releasing neurons in the midbrain degenerates and movement suffers; supplying the missing dopamine's chemical precursor, L-DOPA, can relieve the symptoms for years. Other diseases strike a single class of cells: amyotrophic lateral sclerosis (ALS) and polio destroy motor neurons, while in myasthenia gravis the immune system attacks the acetylcholine receptors through which motor neurons excite muscle.

1.7

Circuits

Neurons never act alone. Each is part of a circuit, and what a neuron does for behavior is determined by its connections. Kandel counts this among the key organizational principles of the brain: nerve cells with similar properties can produce very different actions because of the way they are interconnected. The simplest circuits show how the pieces fit together.

The knee-jerk reflex

When a doctor taps the tendon below your kneecap, your leg kicks. The tap stretches the quadriceps, the muscle at the front of the thigh, and activates stretch receptors inside it, the *muscle spindles*. Sensory neurons carry the signal into the spinal cord (Figure 1.10). In cross-section the spinal cord is the cortex turned inside out: a butterfly-shaped core of gray matter, holding cell bodies, surrounded by white matter, made of axons running up and down. Sensory axons enter through the *dorsal roots*; motor neurons sit in the *ventral horns* of the gray matter and send their axons out through the *ventral roots*. Inside the cord, the reflex circuit does three things at once:

- Each sensory neuron excites the motor neurons of the quadriceps directly, through a single synapse, and the muscle contracts. This *monosynaptic* reflex is the simplest circuit in the body.

- The same sensory neurons excite *inhibitory interneurons* that suppress the motor neurons of the opposing muscle, the hamstring at the back of the thigh, so that it relaxes instead of resisting. The reflex would work without them, but less stably.
- Other branches carry information about the reflex up to higher centers, so that the brain can coordinate it with other behaviors, at the same time or in sequence.

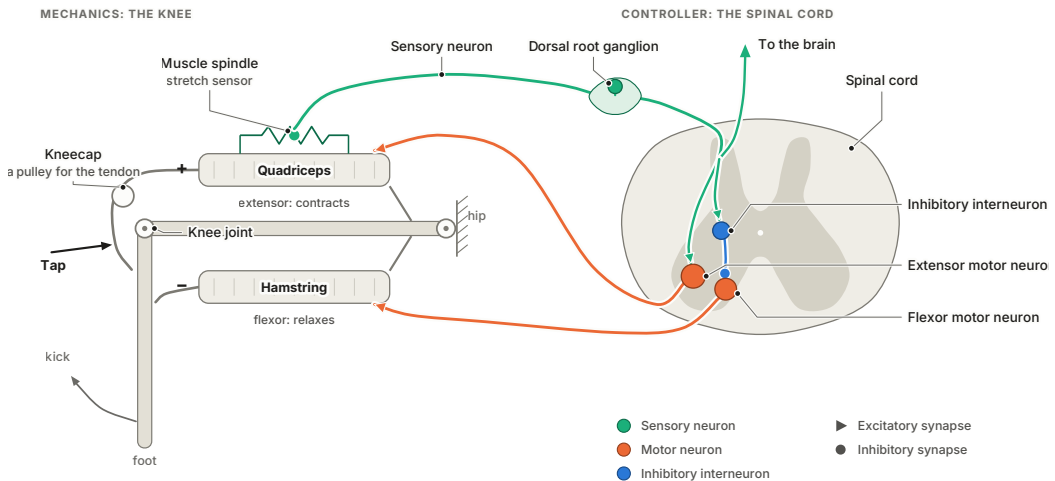


Figure 1.10 The knee-jerk reflex as a control loop. *Left:* the mechanics. The knee is a hinge driven by two opposing actuators, the quadriceps (extensor) and the hamstring (flexor); the muscle spindle is a stretch sensor in parallel with the quadriceps. *Right:* the controller. The sensory neuron, whose cell body lies in a dorsal root ganglion, excites the quadriceps motor neurons directly and, through an inhibitory interneuron, inhibits the hamstring motor neurons. Excitatory synapses are drawn as triangles and inhibitory synapses as circles.

The reflex shows wiring patterns that recur everywhere. Stretching the quadriceps activates several hundred sensory neurons, and each contacts 45 to 50 motor neurons — *divergence*, typical of the input stages of the nervous system. Each motor neuron, in turn, collects 200 to 450 contacts from about 130 sensory neurons — *convergence*, typical of its output stages. No single input is strong enough to fire a motor neuron; it fires only when many sensory neurons are active together. The inhibitory interneuron implements *feed-forward inhibition*: the signal that excites one pathway also inhibits its competitor, a pattern known in the spinal cord as *reciprocal innervation*. The opposite arrangement, *feedback inhibition*, has a neuron excite an interneuron that inhibits the neuron itself, limiting its own activity; spinal motor neurons do this through interneurons called Renshaw cells.

ENGINEERING LENS

A reflex is a servo loop

The stretch reflex is a negative-feedback controller: when a load stretches a muscle, the reflex contracts it back toward its set length, without waiting for the brain. It keeps tone in the quadriceps so that your knees do not buckle when you stand. Its gain is adjustable, and the brain sets it according to context: when you run, the knee bends with every stride and stretches the quadriceps, so the brain and spinal cord turn the reflex down. After some strokes, which cut the descending control, the reflex becomes exaggerated and the joint stiff. A clinician can even raise the gain on demand by asking you to hook your fingers together and pull while the knee is tapped (the Jendrassik maneuver). Fast local loops whose gains are set centrally are a design pattern the nervous system uses at every level.

Circuit motifs

Larger circuits reuse a small set of wiring patterns, or *motifs* (Figure 1.11). *Divergence* broadcasts one signal to many targets, and *convergence* combines many signals into one decision. *Feed-forward excitation* relays information from stage to stage, building hierarchies (Section 1.9). *Feed-forward inhibition* sharpens timing and arbitrates between competing actions. *Feedback inhibition* stabilizes activity and controls gain. *Recurrent excitation* — neurons exciting each other — amplifies signals and can sustain activity after its input has ended, a basis for short-term memory.

Because a neuron can receive excitation from hundreds or thousands of synapses, circuits must keep excitation in check with inhibition. Many operate near a balance between the two, which may increase their computing power but leaves them prone to runaway activity: an epileptic seizure is excitation escaping control.

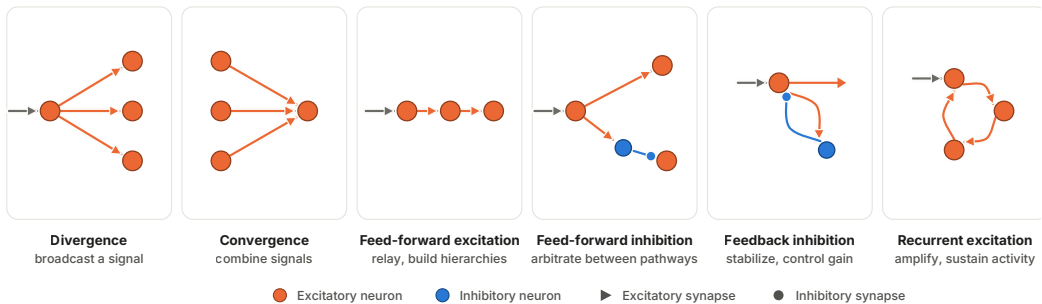


Figure 1.11 Six circuit motifs. Excitatory neurons are drawn in orange and inhibitory neurons in blue; triangles mark excitatory synapses and circles inhibitory ones. The same motifs appear in reflex circuits, in the cortex, and in artificial networks.

Complex behaviors are built from the same components — sensory neurons, interneurons, motor neurons — and the same motifs, but with many more stages between input and output, many parallel paths, and loops that run back from later stages to earlier ones. Kandel suggests that the reflex's basic plan often survives. Just as the firing rate of a stretch receptor encodes how far a muscle is stretched, the firing rates of neurons in sensory cortex encode the intensity of a feature, those in motor cortex shape the movement that follows, and those in association cortex encode quantities such as the evidence for a choice. What thought adds are two elaborations. One is flexible routing: depending on context and goals, the same input can lead to many actions, and the same action can follow from many inputs. The other is freedom from immediacy. A reflex acts as soon as its signals arrive — the ankle jerk lags the knee jerk only because its nerves are longer — while neurons in association cortex can sustain graded firing for many seconds, so that action can wait for more evidence or the right moment (Section 1.9).

Circuits are modified by experience

Circuits are not fixed, which raises a puzzle: if the connections between neurons are set so precisely during development, how can learning change behavior? The answer that has proved most farsighted is the *plasticity hypothesis*, proposed by Cajal around 1900 and given its modern form by the Polish psychologist Jerzy Konorski in 1948: synapses change with use. The strength of a synapse can change for seconds to hours, through modifications of proteins already present at the synapse, and with repeated experience the change can last days or longer, as neurons switch on new genes and grow new synaptic connections or prune old ones. Kandel's own work on the marine snail *Aplysia* showed how a simple reflex, the withdrawal of its gill, weakens with repetition (habituation) and strengthens after a noxious stimulus (sensitization) through exactly such changes at identified synapses. In 1949 Donald Hebb proposed a rule for which synapses should change (Section 1.9).

PRINCIPLE 8

Synaptic plasticity

The strength and the number of synaptic connections change with experience. Short-term changes alter the efficacy of existing synapses; long-term changes require new gene expression and the growth or removal of synapses. These changes are thought to be the cellular basis of learning and memory.

1.8

Systems

Circuits are assembled into systems that carry information from the senses to the cortex and from the cortex to the muscles. Following two pathways — touch in and movement out — introduces most of the anatomy needed in the rest of the book (Figure 1.12).

Touch: from skin to cortex

Touch the tip of your finger. Receptors in the skin excite the peripheral branch of a pseudo-unipolar sensory neuron whose cell body lies in a dorsal root ganglion next to the spinal cord. The neuron's central branch enters the spinal cord through a *dorsal root* and divides. Local branches end in the spinal gray matter, where they feed reflex circuits; the main branch, without making a synapse, ascends on the same side, in the *dorsal columns* of the white matter, all the way to the medulla. There it excites neurons of the *dorsal column nuclei*, whose axons cross the midline and ascend through the brain stem as the *medial lemniscus* to the thalamus. Thalamic neurons, in turn, project to the primary somatosensory cortex in the postcentral gyrus.

Three features of this pathway recur throughout the nervous system:

- **It is a chain of relays**, and the information is transformed at each synapse, not merely passed along. In Kandel's phrase, sensation is an abstraction, not a replication, of reality.
- **It is mapped**. Neighboring points on the skin are represented by neighboring neurons at every stage, so each relay holds a map of the body surface (it is *somatotopic*). In the cortex, the map is distorted in proportion to sensory acuity: the fingers and lips occupy far more cortex than the back, a distortion depicted by Wilder Penfield's *homunculus*. The map is not fixed: in violinists, the area devoted to the fingers that press the strings is enlarged.
- **It crosses the midline** — in the medulla, for this pathway — which is why the left hemisphere feels the right hand. Not every pathway crosses completely: only about half the fibers of each optic nerve cross, hearing reaches both hemispheres, and smell stays mostly on the same side.

Different kinds of bodily sensation travel in separate, parallel pathways. Pain and temperature, for example, ascend in the *anterolateral system*, which crosses the midline within the spinal cord rather than in the medulla.

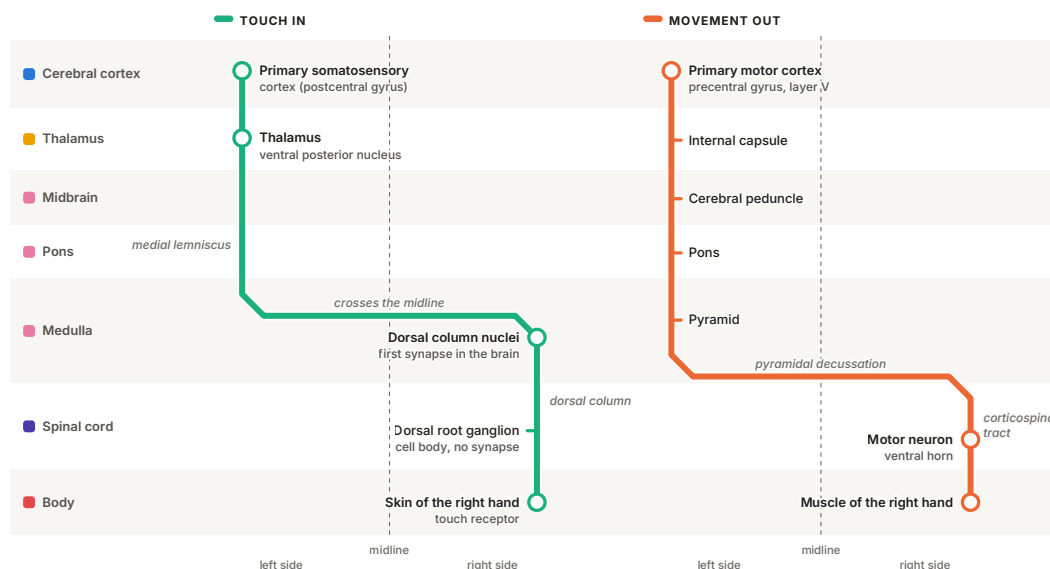


Figure 1.12 Touch in, movement out. The touch pathway (dorsal column–medial lemniscal system) relays through the dorsal column nuclei and the thalamus and crosses the midline in the medulla. The motor command (corticospinal tract) descends from the primary motor cortex, and most of its fibers cross at the pyramidal decussation, at the junction of the medulla and the spinal cord. Both crossings explain why each hemisphere serves the opposite side of the body.

The thalamus, gateway to the cortex

Except for smell, all sensory information reaches the cortex through the thalamus, a cluster of as many as fifty nuclei. Each sense has its own *relay nuclei* — the *lateral geniculate nucleus* for vision, the *medial geniculate nucleus* for hearing, and the *ventral posterior nuclei* for touch — while others relay the output of the cerebellum and basal ganglia to the motor cortex, form loops with emotional and association areas, or regulate arousal. The thalamus is not a passive relay. In the lateral geniculate nucleus, for example, axons returning from the cortex make more synapses on the relay neurons than the retina does, although the retina still drives them. And a thin shell of inhibitory neurons surrounding the thalamus, the *reticular nucleus*, samples the traffic between thalamus and cortex and inhibits the relay cells, gating what gets through — a mechanism implicated in attention and sleep.

The cortical sheet

Most of the cortex is *neocortex*, which has six layers, numbered I at the surface to VI at the bottom. The layers differ in their cell types and connections, in a pattern repeated across areas (Figure 1.13): input from the thalamus arrives mainly in layer IV; layers II and III connect to other cortical areas; layer V sends the cortex's output to structures below it — the basal ganglia, brain stem, and

spinal cord; and layer VI projects back to the thalamus. Layer I contains mostly dendrites and axons. Perpendicular to the layers, neurons are organized in *columns*, each about a third of a millimeter wide, that share response properties: in the somatosensory cortex, the neurons of a column respond to the same kind of stimulus at the same place on the body.

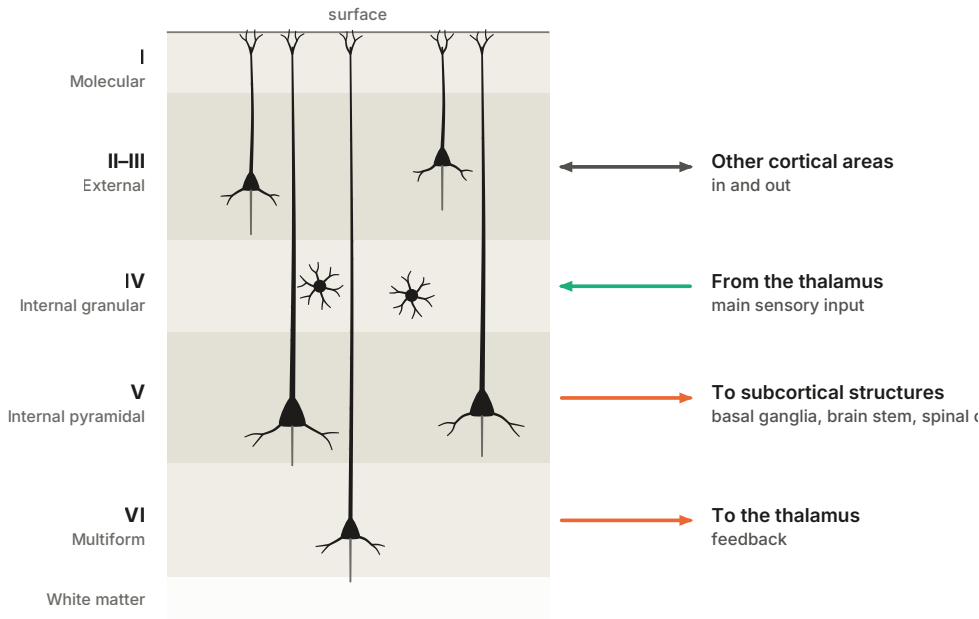


Figure 1.13 The six layers of the neocortex and their main connections. Thalamic input arrives mostly in layer IV, layers II–III exchange information with other cortical areas, layer V sends output to subcortical structures, and layer VI sends feedback to the thalamus.

The relative thickness of the layers varies from area to area, and the variation follows function. Layer IV, the input layer, is so thick in the primary visual cortex that it is divided into sublayers, and it is essentially absent from the primary motor cortex, whose output layer V is prominent instead. In 1909 Korbinian Brodmann used this variation — the *cytoarchitecture* of the cortex — to divide it into about fifty areas, whose numbers are still in use: the primary visual cortex is area 17 and the primary motor cortex area 4. Functionally, the areas form a hierarchy. Primary sensory areas feed *unimodal association* areas, which elaborate a single sense, and these feed *multimodal association* areas, in the pre-frontal, parietal-temporal, and limbic cortex, which combine the senses and support planning, language, and memory. Connections usually run in both directions, and the layers tell them apart: forward (*feed-forward*) projections

leave mainly from layer III and arrive in layer IV of the next area, while backward (*feedback*) projections leave from layers V and VI and end outside layer IV.

Movement: from cortex to muscle

The command for a voluntary movement leaves the primary motor cortex, in the precentral gyrus, along the axons of large pyramidal neurons in layer V. They descend through the *internal capsule*, the base of the midbrain, and the pons; gather into the *pyramids* on the front of the medulla; and most of them cross the midline at the *pyramidal decussation*, where the medulla meets the spinal cord. They continue down the *corticospinal tract*, about a million axons strong, to interneurons and to the motor neurons of the ventral horn. Because the pathway crosses, the left motor cortex moves the right side of the body. Two great loops refine the commands: the cerebellum compares intended and actual movements, and the basal ganglia help select which actions to perform. Both act largely through the thalamus, back onto the cortex.

Modulatory systems

Alongside these point-to-point pathways, the brain has broadcast systems (Figure 1.14). Small clusters of neurons in the brain stem and the base of the forebrain send highly branched axons over vast territories of the brain and release *neuromodulators*, which change how their targets respond rather than carrying specific messages:

- **Dopamine** neurons, in the *substantia nigra* and the *ventral tegmental area* of the midbrain, project to the striatum (the input stage of the basal ganglia) and to the frontal cortex. They are involved in movement, motivation, and learning from reward. Rats allowed to stimulate these reward circuits by pressing a lever prefer it to food, water, or sex.
- **Norepinephrine** neurons, in the *locus coeruleus* of the pons, project to almost the entire CNS and regulate arousal, vigilance, and the response to novelty and stress.
- **Serotonin** neurons, in the *raphe nuclei* along the midline of the brain stem, also project almost everywhere and influence mood, sleep, and impulse control.
- **Acetylcholine** neurons, in the *basal forebrain* (and in the brain stem), project to the cortex and hippocampus and regulate attention, arousal, and learning.

These systems contain few neurons — a human locus coeruleus has only tens of thousands — but they influence motivation, emotion, attention, and memory throughout the brain, and many psychiatric drugs act on them.

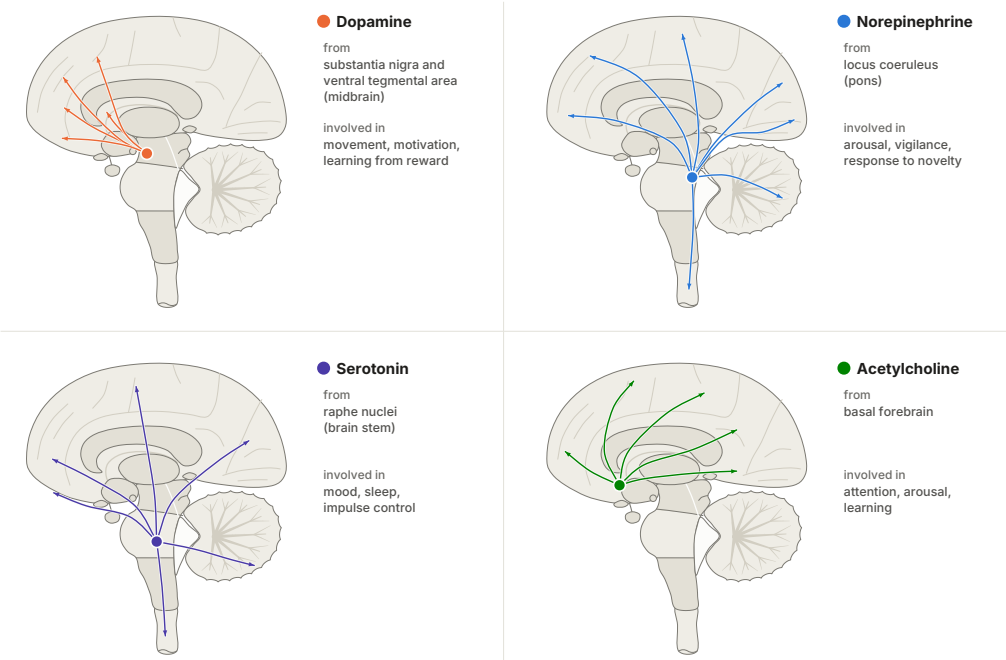


Figure 1.14 Four modulatory systems. Small groups of neurons in the brain stem and basal forebrain send widely branching axons across the brain. Each system is drawn in its own color; the target regions shown are the main ones, not a complete list.

ENGINEERING LENS

Global hyperparameters

Neuromodulators look less like messages and more like global configuration: a few broadcast signals that change the gain, the plasticity, and the mode of operation of many circuits at once. Kenji Doya proposed a specific mapping onto reinforcement learning: dopamine signals the reward prediction error, serotonin the time horizon over which future rewards count (the discount factor), norepinephrine the randomness of choices (the balance between exploring and exploiting), and acetylcholine the learning rate.¹²⁹ Only the dopamine part is firmly established (Section 1.12); treat the rest as a productive hypothesis. The lesson that holds is architectural: alongside its point-to-point wiring, the brain has a low-bandwidth broadcast channel that sets the mode of computation.

¹² Doya K (2002). Metalearning and neuromodulation. *Neural Networks* 15:495–506.

The peripheral nervous system

Outside the CNS, the *somatic* division of the PNS comprises the sensory neurons that serve the skin, muscles, and joints and the axons of the motor neurons that command the skeletal muscles. The *autonomic* division carries sensation from the internal organs and controls them, along with the blood vessels and glands. Its *sympathetic* system prepares the body for action (the fight-or-flight response), its *parasympathetic* system for rest and digestion, and its *enteric* system — a network of hundreds of millions of neurons in the wall of the gut — runs digestion largely on its own.

Memory: the hippocampal loop

Memory shows that the brain's systems are not only sensory and motor. In 1953 a young man known for decades only as H.M. — Henry Molaison — had much of the hippocampus and the surrounding medial temporal lobe removed from both sides of his brain to treat severe epilepsy. His seizures improved, but he lost the ability to form new memories of people, places, and events: he could hold a conversation but forgot it minutes later. Yet his memories from long before the operation were largely spared, and so were his perception, his intelligence, and his ability to learn new motor skills. The hippocampal system is needed to form lasting *explicit* memories — the kind we can recall consciously — but not to store them indefinitely; they are ultimately stored elsewhere, in the cortex, by mechanisms still unknown. *Implicit* memories, such as skills and habits, depend on other structures, including the basal ganglia and the cerebellum.

The *hippocampal formation*, in the medial temporal lobe, sits at the top of the cortical hierarchy. It comprises the *entorhinal cortex*, the *dentate gyrus*, the hippocampus proper, and the *subiculum*. The entorhinal cortex is its entry point, collecting information from the highest-level, multimodal association areas. Inside, connections run mostly in one direction, around a loop (Figure 1.15): from the entorhinal cortex to the dentate gyrus, then to the CA3 and CA1 fields of the hippocampus proper, then to the subiculum, and back to the entorhinal cortex, which returns the result to the neocortex. One stage breaks the pattern: CA3 neurons are richly connected to each other. A network that excites itself this way can store patterns of activity and recall an entire pattern from a fragment, which makes CA3 a candidate for the brain's associative memory.

The loop also shows that circuits are not all alike: their structure suits their function. Sensory and motor systems are hierarchical and reciprocal, with each stage sending connections back to the one before it, but the hippocampal system is mostly serial and one-way. That makes it fragile in a particular way: dam-

age to a single stage can break the whole chain. A patient known as R.B., who lost only the neurons of one field, CA1, after an interruption of blood flow to the brain, suffered a profound loss of memory.

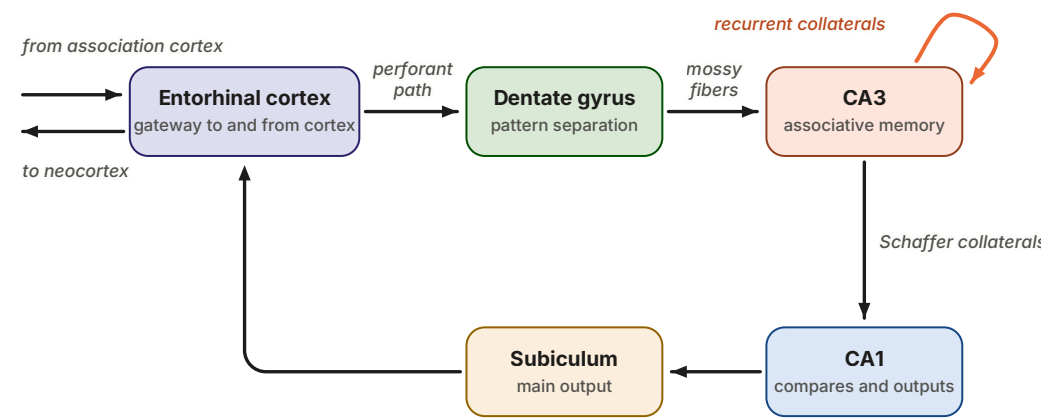


Figure 1.15 The hippocampal loop. Information flows mostly one way: from the entorhinal cortex to the dentate gyrus (along the perforant path), to CA3 (along the mossy fibers), to CA1 (along the Schaffer collaterals), and through the subiculum back to the entorhinal cortex. Recurrent collaterals connect CA3 neurons with each other. The roles in each box are the leading theoretical proposals, not settled facts.

ENGINEERING LENS

Content-addressable memory

Starting with David Marr in 1971, theorists have modeled CA3 as an *autoassociative* network: a recurrent network whose connections, set by a Hebbian rule (Section 1.9), store activity patterns as stable states, so that a partial cue settles into the complete stored pattern.¹³¹⁰ John Hopfield’s 1982 network is the canonical formulation.¹⁴¹¹ It is a content-addressable memory: you retrieve an item by part of its content rather than by an address. In this account the dentate gyrus first makes similar inputs more different (*pattern separation*) so that they can be stored without interfering with each other, and the hippocampus as a whole works like an index, binding together the cortical patterns that make up an event so that they can later be reactivated as a whole.

¹³ Marr D (1971). Simple memory: a theory for archicortex. *Philosophical Transactions of the Royal Society of London B* 262:23–81.

¹⁴ Hopfield JJ (1982). Neural networks and physical systems with emergent collective computational abilities. *PNAS* 79:2554–2558.

1.9

Computation

Classical artificial intelligence tried to reproduce intelligence with serial, symbolic algorithms. It worked well for some problems, such as chess, and poorly for problems that brains find easy, such as recognizing faces and understanding speech. Neural network models took the opposite approach: many simple units, connected in feed-forward and recurrent architectures, computing in parallel. Such networks share two properties with real circuits. Their power comes less from the complexity of their units than from the number of units and the pattern of their connections. And, like the brain, which stays busy during sleep, recurrent networks can generate activity of their own, without any input. A chapter new to the sixth edition of Kandel's book asks the question from the biological side: what computations do real circuits perform, and how?

Recording and controlling circuits

For decades, neurophysiologists recorded from one neuron at a time. Today, silicon probes with hundreds of recording sites capture the spikes of hundreds of neurons at once,^{15,12} and microscopes image the activity of thousands of neurons through genetically encoded indicators that glow when calcium enters an active cell — more slowly than electrodes, but following the same identified cells for weeks. Just as important, neurons can now be controlled. *Optogenetics* introduces genes for light-sensitive proteins into chosen types of neurons — channels such as channelrhodopsin to excite them, pumps such as halorhodopsin to silence them — so that light switches them on or off within milliseconds; *chemogenetics* does the same, more slowly, with engineered receptors that respond only to a designer drug. Recording reveals what a circuit represents; control reveals whether that representation causes behavior. Making sense of the data is a computational problem in its own right: statistical methods extract spikes from noisy signals, dimensionality reduction finds the few patterns that dominate the activity of many neurons, and models range from detailed simulations of single cells to abstract descriptions of whole populations.

Encoding and decoding

A neuron *encodes* a variable if its firing depends on it. Much of what we know comes from measuring *tuning curves*: firing rate as a function of some stimulus or movement variable. In the 1960s David Hubel and Torsten Wiesel found that neurons in the primary visual cortex respond best to edges of a particular ori-

¹⁵ Jun JJ et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551:232–236.

entation, each neuron preferring its own; neurons in the motor cortex similarly prefer particular directions of arm movement. Codes can also use time: the precise timing of spikes, and their order across a population, can carry information that a firing rate averages away.

Decoding runs the other way: from the activity of a population, infer the variable. It is not only something experimenters do — every brain area must, in effect, decode the activity of its inputs. The hippocampus provides the clearest example. In 1971 John O’Keefe discovered that some hippocampal neurons fire only when an animal is in a particular part of its environment — the cell’s *place field* — as if each marked a location on a map. Together, *place cells* cover the whole environment, and from the activity of about a hundred of them it is possible to track where a rat is, moment by moment, to within a few centimeters (Figure 1.16).^{16,13} Single cells are noisy, but the population is redundant, and a good decoder weights each cell by how reliable it is. Place cells have since been found in bats, monkeys, and people. The entorhinal cortex adds *grid cells*, each of which fires at many locations arranged in a triangular lattice, giving the map a metric.^{17,14} O’Keefe and May-Britt and Edvard Moser shared the 2014 Nobel Prize for these discoveries, which made concrete Edward Tolman’s 1948 idea of a *cognitive map*.

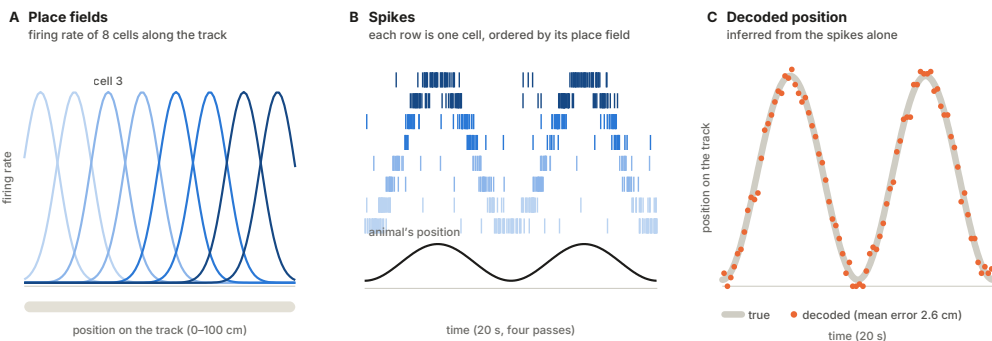


Figure 1.16 Encoding and decoding position in the hippocampus. A. As a rat runs along a track, each place cell fires near one location, its place field. B. Spike rasters of the same cells over several laps. C. A decoder that knows each cell’s tuning can infer the rat’s position from the population’s spikes alone; the decoded position closely tracks the true one.

¹⁶ Wilson MA, McNaughton BL (1993). Dynamics of the hippocampal ensemble code for space. *Science* 261:1055–1058.

¹⁷ Hafting T, Fyhn M, Molden S, Moser M-B, Moser EI (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature* 436:801–806.

Decoding also reveals what the brain does offline. During pauses and sleep, hippocampal populations emit brief, high-frequency bursts called *sharp-wave ripples*, lasting 50 to 500 milliseconds, and within them place cells fire in sequences that replay paths the animal has taken — compressed 10- to 20-fold in time, so that a ripple of a fifth of a second can replay several seconds of running. Replay sometimes runs in reverse, and sometimes traces paths the animal has never taken.^{18,15} It is thought to help consolidate memories and plan routes.

ENGINEERING LENS

Decoding is supervised learning

A decoder is a model fit to data: neural activity in, variable out. Linear readouts, Bayesian decoders, and neural networks are all used, and the best decoders drive brain-machine interfaces that let people with paralysis move cursors and robotic arms (Kandel et al., chapter 39). Two cautions transfer directly from machine learning. Evaluate on held-out data, because decoders overfit like any other model. And decodable is not the same as used: if a variable can be read out of a brain area, the information is there, but that does not prove that the brain reads it from there.

Feed-forward hierarchies

In the visual system, information flows from the retina through the thalamus to the primary visual cortex (V1), and then along the ventral stream (Section 1.4), from the occipital lobe to the underside of the temporal lobe — through V2, V4, and the inferotemporal cortex (IT) (Figure 1.17). Along the way, *receptive fields* — the part of the visual field to which a neuron responds — grow larger, and preferences grow more complex: edges in V1, contours and simple shapes in V2 and V4, and in IT whole objects and categories such as faces, recognized across changes in position, size, and viewing angle. Each stage builds on the features computed by the one before.

Deep convolutional networks share the design ideas Kandel identifies in this hierarchy: receptive fields that tile the image like a map (the biological counterpart of convolution), that grow from stage to stage, and that become more selective for features and more invariant to position and size. The resemblance turned out to run deeper than design. Networks trained only to classify images develop stages that resemble the visual hierarchy, their later layers predict the responses of IT neurons better than models built by hand,^{19,16} and on hard images their errors partly match those of people and monkeys. The differences

¹⁸ Gupta AS, van der Meer MAA, Touretzky DS, Redish AD (2010). Hippocampal replay is not a simple function of experience. *Neuron* 65:695–705.

¹⁹ Yamins DLK et al. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *PNAS* 111:8619–8624.

matter just as much. The visual cortex has massive feedback and lateral connections, learns from very few labeled examples, runs on a fraction of the brain’s 20 watts, and is not fooled by the tiny, carefully chosen changes to an image that fool many networks.

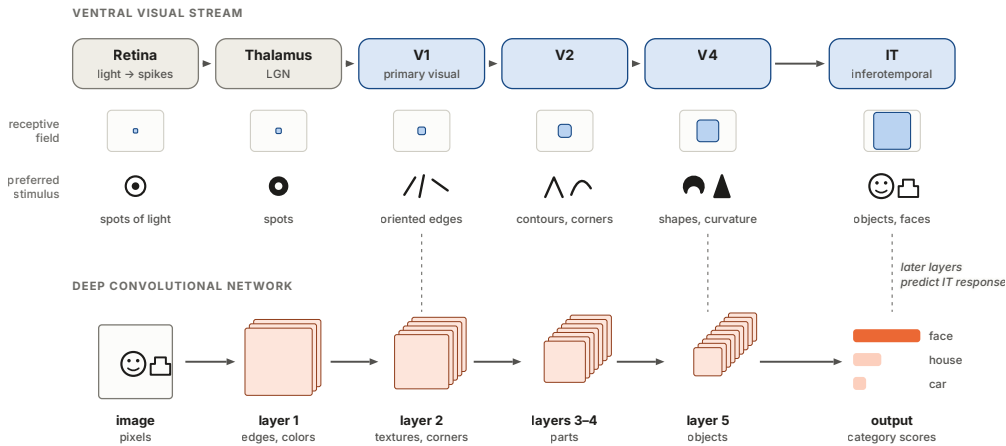


Figure 1.17 A feed-forward hierarchy in the visual cortex and in a deep network. In the ventral visual stream (top), receptive fields grow and preferred features become more complex from V1 to the inferotemporal cortex. A convolutional network trained to recognize objects (bottom) develops a similar progression across its layers. Arrows show the feed-forward sweep; the cortex also has extensive feedback connections, not drawn.

Expansion: the cerebellum

The cerebellum uses a different motif (Figure 1.18). Its input arrives through some 200 million *mossy fibers*, which excite about 50 billion *granule cells*, the most numerous neurons in the brain. Each granule cell receives only about four mossy fibers, a different combination for each cell and often from different sources — one signaling, say, the sight of a moving ball and another the flexing of the wrist — so each responds to a particular conjunction of inputs, a *mixed* representation. The combinations multiply quickly: 100 input channels can be paired in 4,950 ways and grouped in threes in more than 150,000. The granule layer thus recodes the input into a much larger and sparser representation: an *expansion*. Its wiring appears to be largely random — convenient, since a specific pattern would be hard to specify genetically.

The granule cells’ axons, the *parallel fibers*, run along the cerebellar cortex and cross the flat dendritic trees of the Purkinje cells, each of which collects input from more than a hundred thousand of them. Each Purkinje cell also receives one very powerful input from a single *climbing fiber*, which comes from the in-

ferior olive in the brain stem and signals errors or unexpected events. When a climbing fiber fires, the parallel-fiber synapses that were active just before it are weakened, a change called *cerebellar long-term depression*.

David Marr (1969) and James Albus (1971) recognized this as a learning machine.²⁰¹⁷ The expansion makes many different input patterns separable by a simple, linear readout, and the climbing fiber acts as a teacher that adjusts the readout's weights. This is how you recalibrate a movement after errors — for example, how your eye movements adapt when you start wearing new glasses.

Eyeblink conditioning shows the mechanism at work. A tone is repeatedly followed by a puff of air to the eye, and the animal learns to close the eye just before the puff arrives. The tone reaches the Purkinje cells through mossy fibers and granule cells, the puff through the climbing fiber. Pairing them weakens the parallel-fiber synapses that were active just before the puff, which carves a well-timed pause into the Purkinje cells' steady, inhibitory firing; the pause releases neurons in the deep cerebellar nuclei, which drive the eyelid closed. The timing is thought to come from granule cells that respond at different delays after the tone, giving the readout a set of time-shifted signals to choose from. In machine-learning terms the cerebellum learns from errors — supervised learning, though with a crude teacher: in eyeblink conditioning the climbing fiber reports only that something aversive happened, which is why Kandel's chapter calls it a form of reinforcement learning.

²⁰ Marr D (1969). A theory of cerebellar cortex. *Journal of Physiology* 202:437–470. Albus JS (1971). A theory of cerebellar function. *Mathematical Biosciences* 10:25–61.

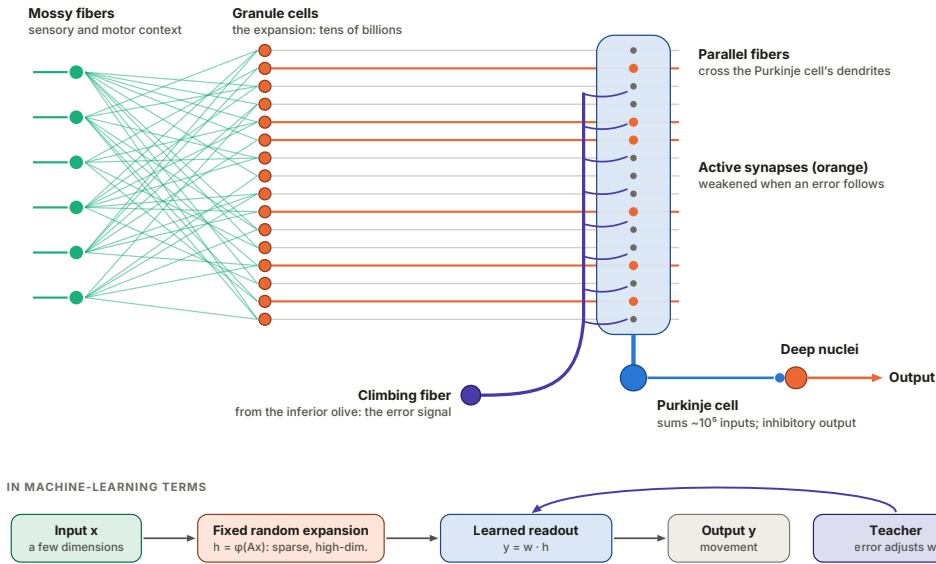


Figure 1.18 The cerebellum as a learning machine. A few mossy fibers fan out to many granule cells, expanding the input into a high-dimensional, sparse code. Parallel fibers converge on Purkinje cells, whose output leaves through the deep cerebellar nuclei. A single climbing fiber delivers an error signal to each Purkinje cell and weakens the parallel-fiber synapses that were active just before the error. Below, the same architecture in machine-learning terms: a fixed expansion followed by a trained linear readout.

ENGINEERING LENS

Random features with a trained readout

Machine learning has rediscovered the cerebellar design several times: project the inputs into a high-dimensional space through fixed, random-looking weights, then train only a linear readout. Random-feature methods, reservoir computing, and extreme learning machines all work this way. The expansion does the heavy lifting of making a problem linearly separable, and the readout is cheap to train with a local error signal. One analysis found that about four inputs per granule cell — the number found in real cerebella — is close to the optimum for maximizing the dimension of the expanded representation.²¹¹⁸

Recurrence: holding and integrating

Feed-forward circuits respond to their input and fall silent when it stops. Many functions need the opposite: activity that outlasts its input. When you move your eyes to a new target, premotor neurons in the brain stem and the superior colliculus send the eye's motor neurons a brief burst of spikes that encodes the velocity of the movement. But to hold the eyes still at their new position, the motor neurons must keep firing at a new, steady rate for as long as you look

²¹ Litwin-Kumar A, Harris KD, Axel R, Sompolinsky H, Abbott LF (2017). Optimal degrees of synaptic connectivity. *Neuron* 93:1153–1164.

there. A circuit in the brain stem, the *neural integrator*, converts one signal into the other: it integrates velocity into position, in the mathematical sense (Figure 1.19). The leading explanation is *recurrent* connections: the integrator's neurons excite one another, so their activity sustains itself after the burst ends. Persistence survives blocking the ion channels that might let single neurons sustain their own firing, which points to the network, though its exact wiring is not settled. Integration is a general-purpose operation. Similar circuits could accumulate evidence over time — as when you pick out a familiar voice in a noisy room — and are thought to produce the *persistent activity* that holds information in prefrontal cortex for seconds, during working memory.

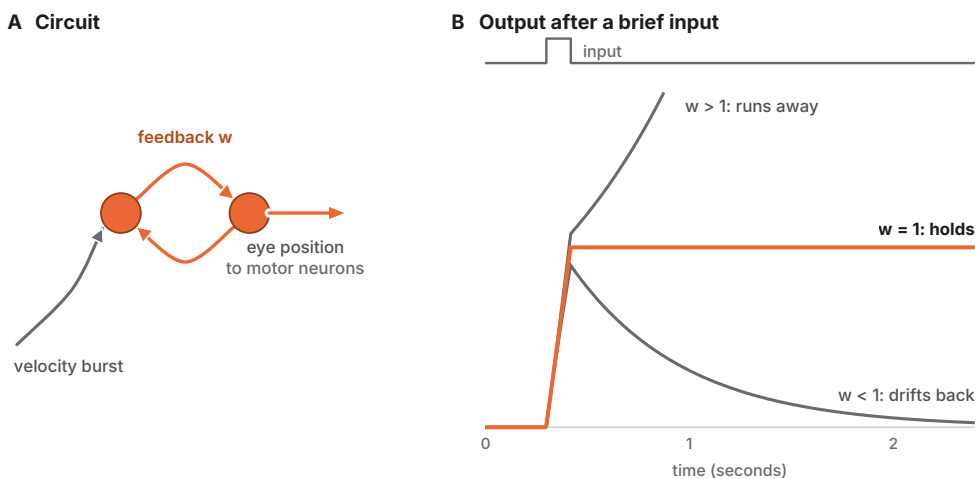


Figure 1.19 A recurrent circuit integrates. *Top:* a brief burst of input encodes the velocity of an eye movement; the integrator's output steps to a new level and holds it, encoding the eye's position. *Bottom:* with feedback weaker than the leak, the output decays back to rest and gaze drifts; with feedback stronger than the leak, it runs away. Only finely tuned feedback holds steady.

ENGINEERING LENS

An integrator is a tuned recurrent network

A single neuron forgets its input within about 20 milliseconds, yet in the dark the integrator holds the eyes off center for about 20 seconds before they drift back: feedback stretches the time constant a thousandfold. For a linear recurrent network, holding a value requires feedback that exactly cancels the leak — an eigenvalue of one along the direction that encodes the value, which makes the network a *line attractor*.²²¹⁹ With too little feedback, the remembered position decays and the eyes drift back toward center; with too much, it runs away. Both failures are seen in patients as forms of nystagmus. Anyone who has trained recurrent networks knows the same knife-edge as vanishing and exploding gradients. The brain appears to keep its integrator tuned continuously, using error signals, in part from the cerebellum.

Learning rules

How do circuits come to be wired correctly? Two kinds of learning rule are well established.

Hebbian plasticity. In 1949 Donald Hebb proposed that when one neuron repeatedly helps to fire another, the connection between them strengthens — often summarized as “cells that fire together wire together.” In its simplest form, the change in a synaptic weight w is proportional to the product of presynaptic activity x and postsynaptic activity y : $\Delta w = \eta xy$, where η is a small learning rate. The physiological counterpart was found in 1973: *long-term potentiation*, a lasting strengthening of synapses after strong, coincident activity on both sides.²³²⁰

Pure Hebbian growth is unstable — a strengthened synapse makes the postsynaptic cell fire more, which strengthens the synapse further — so neurons also regulate their total input, a stabilizing process called *homeostatic plasticity* that has been observed experimentally. With normalization, the rule does something remarkable: the neuron’s weights converge to the dominant pattern in its input, the direction along which the input varies most (Figure 1.20). A Hebbian neuron with normalization extracts the first principal component of its inputs.

²⁴²¹ In a few lines of code:

²² Seung HS (1996). How the brain keeps the eyes still. *PNAS* 93:13339–13344.

²³ Bliss TVP, Lømo T (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology* 232:331–356.

²⁴ Oja E (1982). Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology* 15:267–273.

```
# Oja's rule: Hebbian growth plus a decay term that keeps |w| near 1.
w = rng.normal(size=d)
for x in inputs:
    y = w @ x
    w += eta * y * (x - y * w) # strengthen co-active inputs; normalize
# w converges to the first principal component of the inputs
```

Hebbian learning is *unsupervised*: it finds structure in the input without any teacher.

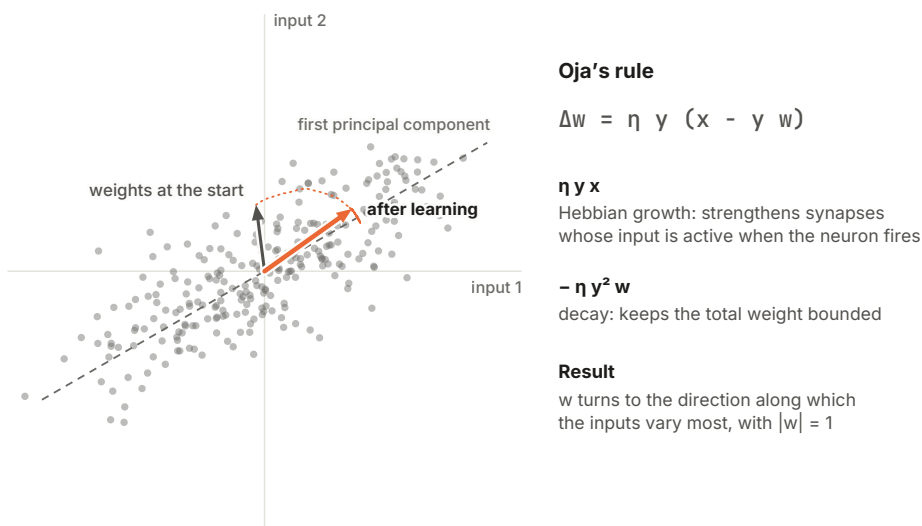


Figure 1.20 Hebbian learning finds the dominant pattern in its input. Each dot is one input pattern presented to a neuron with two synapses. Starting from a random direction, the weight vector rotates during learning until it aligns with the axis along which the inputs vary most — the first principal component.

Error-driven plasticity. In the cerebellum, as we saw, a teaching signal adjusts the synapses that were active just before an error: *supervised* learning. A third kind, learning from reward, uses the neuromodulator dopamine. Dopamine neurons fire when an outcome is better than expected and pause when it is worse: they signal a *reward prediction error*, the same quantity that drives temporal-difference learning in *reinforcement* learning (Section 1.12).

ENGINEERING LENS

Three learning machines

Kenji Doya proposed a division of labor that maps directly onto machine learning: the cerebellum specializes in supervised learning, the basal ganglia in reinforcement learning, and the cerebral cortex in unsupervised learning.²⁵²² It is a simplification — every structure uses more than one kind of plasticity — but it is a useful first map, and it suggests that a brain-like system needs all three, working together.

1.10

Genes

The nervous system's design starts with genes. Genes do not encode behaviors; they encode RNAs and proteins, and these build neurons, wire them into circuits, and run their signaling. This section covers how genes work and how we know they matter; the next shows what they do for behavior.

Heritability

The oldest evidence that genes influence behavior comes from families. In 1875 Francis Galton, who popularized the phrase “nature and nurture,” became the first to use twins to weigh one against the other. *Identical* (monozygotic) twins develop from a single fertilized egg and share all their genes; *fraternal* (dizygotic) twins share about half, like any siblings. If identical twins resemble each other more than fraternal twins do, genes are the likely reason, since both kinds of twins usually share their environment. Identical twins separated early in life and raised in different families make the point vividly: as adults they still show remarkable similarities in personality. A meta-analysis of fifty years of twin studies, covering 17,804 traits, found that on average about half of the variation in human traits is attributable to genetic differences.²⁶²³ The other half shows that the environment matters about as much.

Psychiatric disorders follow the same logic (Figure 1.21). The lifetime risk of schizophrenia, about 1% in the general population, rises with genetic relatedness to an affected person, to about 48% for an identical twin.²⁷²⁴ That number cuts both ways: genes matter enormously, yet about half of the identical twins of people with schizophrenia never develop it. The pattern also rules out a single

²⁵ Doya K (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Networks* 12:961–974.

²⁶ Polderman TJC et al. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics* 47:702–709.

²⁷ Gottesman II (1991). *Schizophrenia Genesis: The Origins of Madness*. W. H. Freeman.

cause. Parents, siblings, children, and fraternal twins all share half their genes with an affected person, yet their risks range from 6% to 17% — a spread that one gene could not produce, but many genes acting together with the environment could.

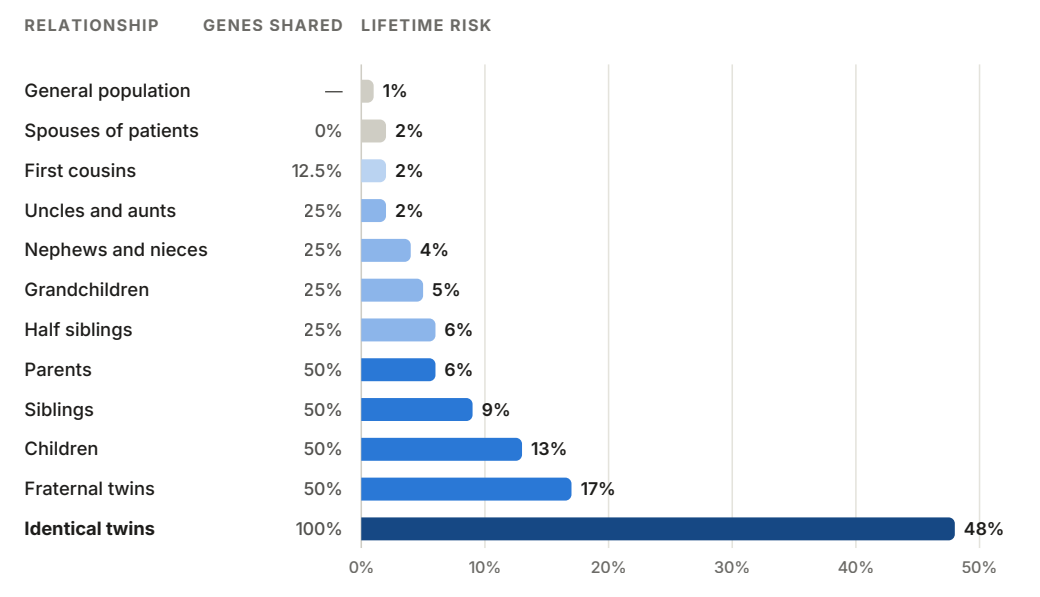


Figure 1.21 Lifetime risk of schizophrenia rises with genetic relatedness to an affected person. Pooled family and twin studies compiled by Gottesman (1991). Relatives also share environments, so these risks mix genetic and environmental influences — which is why the comparison between identical and fraternal twins is the most informative.

How genes work

A gene is a segment of DNA, a double-stranded molecule whose strands are chains of four nucleotides — adenine (A), thymine (T), guanine (G), and cytosine (C) — that pair across the strands, A with T and G with C. Because each strand is a template for the other, DNA can be copied faithfully when a cell divides. The human genome has about 3 billion base pairs and about 20,000 genes that encode proteins; the protein-coding sequences make up less than 2% of the genome. The rest includes the regulatory sequences that control when and where genes are used, tens of thousands of genes for RNAs that are never translated into protein, the introns described below, and long stretches of repeated DNA whose function, if any, is unclear.

A gene has two kinds of parts (Figure 1.22). Its *regulatory regions* — a *promoter* next to the gene and *enhancers* that can be far away — bind *transcription factors*, proteins that switch transcription on or off. Its transcribed region alternates *ex-*

ons, which are kept, with *introns*, which are removed. Expression takes three steps: the gene is *transcribed* into a single-stranded copy made of RNA; the introns are *spliced* out, leaving a *messenger RNA* (mRNA) that leaves the nucleus; and the mRNA is *translated* by ribosomes into a protein. Because splicing can join exons in different combinations, one gene can produce several proteins.

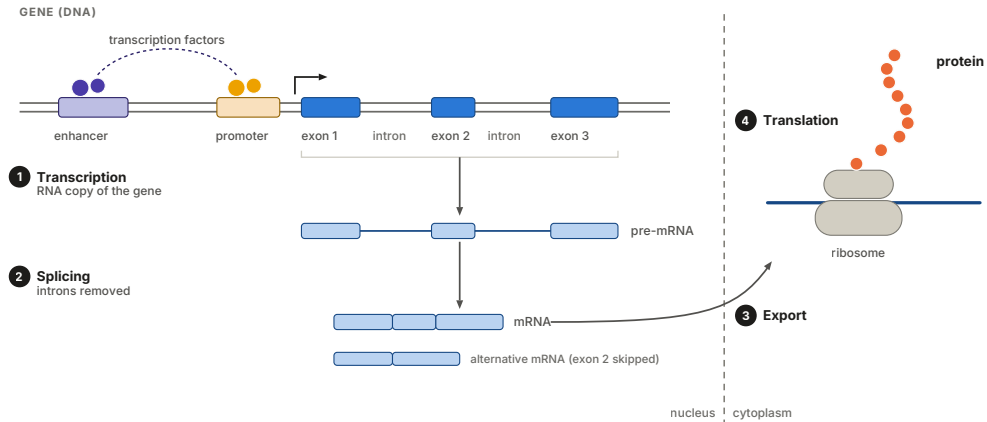


Figure 1.22 From gene to protein. Transcription factors bound to the promoter and enhancers control when the gene is transcribed. The RNA copy is spliced to remove the introns, and the mature messenger RNA is exported from the nucleus and translated into protein.

The brain uses more of the genome than any other organ, and different types of neurons express different combinations of genes. That is how a fixed set of about 20,000 genes can produce thousands of neuron types: the variety comes from combinatorial regulation, not from one gene per type. Gene expression is also dynamic. Experience and neural activity switch genes on and off, which is how long-term memories are stabilized (Section 1.7).

Chromosomes, genotype, and phenotype

Genes are arranged along *chromosomes*. Humans have 23 pairs: 22 pairs of *autosomes* and a pair of sex chromosomes (XX in females, XY in males). We are *diploid*: every cell except eggs and sperm carries two copies of each autosome, one from each parent. The position of a gene on its chromosome is its *locus*, and the different versions of a gene that occur in a population are its *alleles*. Mitochondria, the cell's power plants, carry a small genome of their own, inherited from the mother.

It is essential to distinguish an individual's *genotype*, its genetic makeup, from its *phenotype*, its observable traits. The mapping between them is rarely one to one. A variant may have a large effect, a small one, or none, depending on other genes and on the environment; most traits depend on many genes; and the same

genotype can produce different phenotypes. Even eye color, often taught as a simple trait, depends on several genes, most strongly on variants in and near *OCA2*, which control how much pigment the iris makes.

Genes vary between individuals. Most variants are *single-nucleotide polymorphisms* (SNPs) — two people differ at about one position in a thousand — but stretches of DNA can also be deleted or duplicated (*copy-number variants*), and everyone carries some 70 to 90 new, *de novo* point mutations not present in either parent — on average about one of them changes a protein — with more in the children of older fathers. A *recessive* mutation shows its effect only when both copies of a gene are affected; a *dominant* one needs only one. Many disorders of brain development arise from *haploinsufficiency*: one working copy of the gene is not enough. As a rule, common variants have small effects and rare ones large effects, though there are exceptions: a common variant of the *APOE* gene raises the risk of late-onset Alzheimer's disease about fourfold.

Conserved from worms to humans

Genes are conserved through evolution. About 80% of mouse genes have a single counterpart, an *ortholog*, in the human genome;²⁸²⁵ more than half of human genes resemble genes of worms or flies; and where the human and chimpanzee genomes can be aligned, they differ by only about 1%.²⁹²⁶ Yet humans and chimpanzees are profoundly different. Most of the difference lies not in new genes but in how shared genes are regulated: when, where, and how much they are expressed. The basic molecular machinery of neurons — for development, signaling, and gene regulation — was already present in the common ancestor of worms, flies, mice, and humans, and genes important for the human brain tend to be among the most conserved. That is why simpler animals can reveal how genes shape behavior.

Two strategies are used. *Forward genetics* starts with behavior: mutate many animals at random, find individuals that behave abnormally, and identify the responsible gene. *Reverse genetics* starts with a gene: delete or alter it — in specific cell types and at specific times, with tools such as Cre/loxP recombination and CRISPR gene editing — and observe the consequences.

²⁸ Mouse Genome Sequencing Consortium (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–562.

²⁹ Chimpanzee Sequencing and Analysis Consortium (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437:69–87.

ENGINEERING LENS

The genome is a compressed prior

The genome holds roughly 750 megabytes of information (3 billion base pairs at 2 bits each), and the brain has on the order of 10^{14} synapses. The genome cannot possibly specify the wiring synapse by synapse; it has to encode rules — which cell types to make, which regions to connect, which learning rules to use. Anthony Zador calls this the *genomic bottleneck*, and argues that it acts as a regularizer: evolution compresses what generations of animals learned into wiring rules, which become innate circuits and inductive biases — enough for a newborn foal to stand and walk within hours — while learning tunes the details.³⁰²⁷ In machine-learning terms, genes are closer to the architecture, the initialization, and the learning algorithm than to the trained weights.

PRINCIPLE 9

Genes and environment

Genes contribute to behavior, but behavior is not inherited directly. Genes encode RNAs and proteins that build and regulate the circuits underlying behavior; experience and the environment shape those circuits in turn, partly by changing which genes are expressed.

1.11

Behavior

Single genes rarely explain a behavior. In animals, though, some genes have strikingly specific effects, and they illuminate how genes act. Kandel's examples span timekeeping, foraging, social bonds, and human cognition.

A genetic clock

Many behaviors — sleep and waking, activity, body temperature, hormone release — follow a daily, or *circadian*, rhythm. The rhythm is not simply a response to daylight: animals kept in constant darkness keep cycling, with a period close to but not exactly 24 hours (about 24.2 hours in humans). An internal clock generates the rhythm, and light keeps the clock synchronized with the sun.

In 1971 Ronald Konopka and Seymour Benzer screened mutant fruit flies for abnormal daily rhythms and found three mutations in a single gene, which they named *period*: one shortened the cycle to about 19 hours, one lengthened it to

³⁰ Zador AM (2019). A critique of pure learning and what artificial neural networks can learn from animal brains. *Nature Communications* 10:3770.

about 28, and one abolished it.³¹²⁸ Until then, many doubted that there could be true “behavior genes” — genes not needed for an animal’s basic physiology — and the *period* mutants were otherwise healthy. A single gene could change the speed of a behavioral clock, so its product had to be part of the timekeeper itself, not merely a component the clock needed to run. In 1994 Joseph Takahashi’s group found a comparable mutation in mice, in a gene they named *Clock*: mice with one mutant copy run slow, and mice with two lose their rhythm altogether in constant darkness. A group of genes, not a single one, keeps time, and the same genes do it in flies and mice.

The clock is a *transcriptional oscillator*: a molecular negative-feedback loop (Figure 1.23). Two transcription factors, CLOCK and CYCLE (called BMAL1 in mammals), switch on the *period* and *timeless* genes. Their proteins, PER and TIM, accumulate in the cytoplasm over hours, pair up, and enter the nucleus, where they shut off CLOCK and CYCLE — and therefore their own production. As PER and TIM are degraded, the inhibition lifts and the cycle starts again. Most of the delay comes after the proteins are made: enzymes called kinases tag PER with phosphate groups that mark it for rapid destruction, so it accumulates only when enough TIM is present to bind and protect it. These delays stretch each turn of the loop to about a day. The clock’s transcription factors also switch on output genes that carry the time signal to the rest of the body — in flies, for example, the gene for a neuropeptide, PDF, that sets the daily rhythm of locomotor activity. Light resets the clock by triggering the destruction of TIM.

A similar loop, built from related genes, runs in mammals, where a master clock in the hypothalamus receives light information directly from the retina. In people with *familial advanced sleep phase syndrome*, a mutation in a *PER* gene shifts the whole cycle earlier: they fall asleep around 7:30 p.m. and wake around 4:30 a.m.³²²⁹ Jeffrey Hall, Michael Rosbash, and Michael Young received the 2017 Nobel Prize for working out the mechanism.

³¹ Konopka RJ, Benzer S (1971). Clock mutants of *Drosophila melanogaster*. *PNAS* 68:2112–2116.

³² Toh KL et al. (2001). An *hPer2* phosphorylation site mutation in familial advanced sleep phase syndrome. *Science* 291:1040–1043.

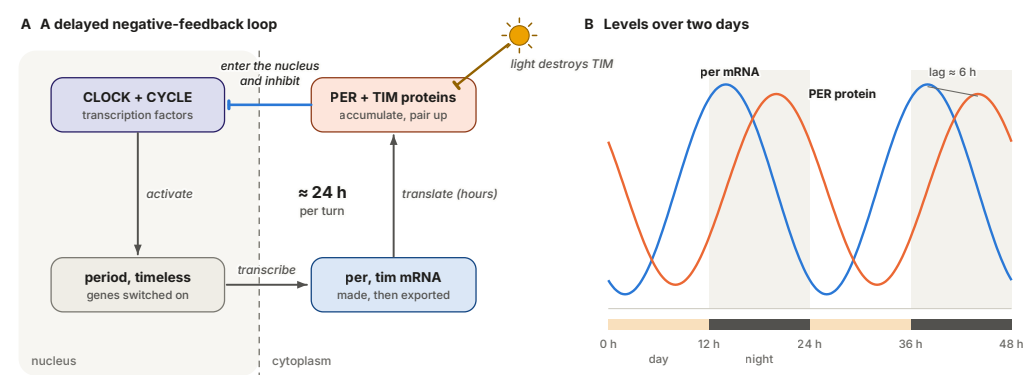


Figure 1.23 The circadian clock is a delayed negative-feedback loop. *Left:* CLOCK and CYCLE activate the *period* and *timeless* genes; their proteins accumulate, enter the nucleus, and inhibit CLOCK and CYCLE. Light triggers the degradation of TIM, which resets the phase of the loop. *Right:* over a day, *per* mRNA peaks in the early night and PER protein lags it by several hours.

ENGINEERING LENS

A phase-locked loop made of proteins

The circadian clock is a textbook delayed negative-feedback oscillator, and its relation to daylight is exactly that of a *phase-locked loop*: a free-running oscillator with its own natural period, plus a phase detector — light — that nudges it toward an external reference. The design explains familiar experience. In constant conditions the clock drifts at its natural period. After a flight across time zones, it relocks slowly, roughly one time zone per day: jet lag. It is also a reminder that innate is not the same as rigid. The oscillator is genetically built, but its phase is continuously corrected by the environment. A system that must behave sensibly across days may be better served by a built-in clock corrected by evidence than by having to discover the 24-hour cycle from scratch.

Exploring or exploiting

Fruit-fly larvae come in two natural varieties. *Rovers* move a lot while feeding and travel between patches of food; *sitters* move little and stay on one patch. Marla Sokolowski traced the difference to the *foraging* gene, the first gene found to underlie a natural variation in behavior rather than one created in the laboratory. It encodes a protein kinase (PKG) activated by the intracellular messenger cGMP — one of a family of enzymes that are especially important for turning short-term signals into long-term changes in neurons — and rovers have more PKG activity than sitters.³³³⁰ Both variants persist in wild populations because each wins under different conditions — in crowded environments rovers do bet-

³³ Osborne KA et al. (1997). Natural behavior polymorphism due to a cGMP-dependent protein kinase of *Drosophila*. *Science* 277:834–836.

ter, and in sparse ones sitters exploit their food more thoroughly. Reinforcement-learning practitioners will recognize the trade-off between exploration and exploitation, here with a single gene setting the default policy. The same gene shapes the division of labor in honeybees, where young bees work inside the hive as nurses and older bees forage outside. Foragers express more of it, and boosting PKG activity in young bees makes them start foraging early.^{34,31}

Social behavior

Social behaviors vary enormously between species, yet they have large innate components. In the roundworm *C. elegans*, some strains feed alone and others feed in groups. The difference maps to a single amino acid in *npr-1*, a receptor for a *neuropeptide* — one of a large class of signaling molecules that neurons release to coordinate the activity of whole networks; in mammals, neuropeptides have been implicated in feeding, sleep, pain, and much else.^{35,32} Two of them, *oxytocin* and *vasopressin*, regulate social bonds, oxytocin mainly in females and vasopressin mainly in males. Prairie voles form lifelong pair bonds, and the males help raise the young; closely related meadow and montane voles do neither. The species differ in where their brains express receptors for these neuropeptides, notably in reward circuits, and adding vasopressin receptors to a reward area in the brains of male meadow voles is enough to make them prefer a single partner.^{36,33} Whether these molecules play a similar role in human attachment is not known.

Genes and the human brain

In humans, rare genetic syndromes open windows onto the biology of cognition and social behavior:

- **Phenylketonuria (PKU)** is caused by recessive mutations in the gene for phenylalanine hydroxylase, the enzyme that converts the amino acid phenylalanine into tyrosine. Untreated, it causes severe intellectual disability; a diet low in phenylalanine, begun in infancy, largely prevents the damage. The phenotype comes from genotype and environment together: the same mutation is devastating on one diet and largely harmless on another.

³⁴ Ben-Shahar Y, Robichon A, Sokolowski MB, Robinson GE (2002). Influence of gene action across different time scales on behavior. *Science* 296:741–744.

³⁵ de Bono M, Bargmann CI (1998). Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in *C. elegans*. *Cell* 94:679–689.

³⁶ Lim MM et al. (2004). Enhanced partner preference in a promiscuous species by manipulating the expression of a single gene. *Nature* 429:754–757.

- **Fragile X syndrome**, the most common inherited cause of intellectual disability, and **Rett syndrome** are caused by mutations in genes (*FMR1* and *MECP2*) that regulate the expression of many other genes in neurons. Both genes lie on the X chromosome, which is why fragile X affects boys more severely, and why Rett syndrome is seen almost only in girls: the mutations are often lethal to male embryos, which have a single X.
- **Williams syndrome**, caused by the loss of one copy of about 27 genes on chromosome 7, produces unusually sociable, talkative people with poor visuospatial skills, intellectual disability, and high anxiety. Gene dosage cuts both ways: an extra copy of the same region raises the risk of autism.
- The **22q11.2 deletion syndrome** raises the risk of schizophrenia about twenty-five-fold, to roughly one in four.
- Mutations in **FOXP2**, a transcription factor, disrupt the fine control of mouth and face movements needed for speech, and language with it. The human form of the protein differs slightly from that of other primates, which made the gene a candidate for one of the changes that made speech possible — an idea still debated.

Common psychiatric disorders are different: they are *polygenic*. Autism spectrum disorder, which affects 2 to 3% of people, is highly heritable. Part of the risk comes from rare de novo mutations and copy-number variants of large effect — sequencing parents and their children has implicated more than a hundred genes — and part from many common variants of small effect. The genes fall into two broad groups: genes for synaptic proteins, such as the *neurexins* and *neuroligins*, which bind each other across the synaptic cleft and help hold synapses together and set their strength; and genes that regulate the expression of other genes. The same rare variants often raise the risk of several conditions — autism, schizophrenia, bipolar disorder, intellectual disability — a one-to-many pattern that cuts across psychiatric diagnoses.

Schizophrenia shows how such risk is being tracked down. Early studies tested candidate genes, chosen by hypothesis, in small samples, and their findings mostly failed to replicate.^{37,34} Genome-wide association studies scan the whole genome without a hypothesis, correct for the vast number of comparisons, and need enormous samples; one of nearly 40,000 patients and more than 110,000 controls found 108 risk loci.^{38,35} The strongest signal lies in the major

³⁷ Among the most studied candidates were *neuregulin 1* and *DISC1*, which affect the migration of neurons and the formation of synapses. Genome-wide studies have not confirmed them as risk genes, although the broader idea they supported — that schizophrenia begins with disturbed brain development — has held. Farrell MS et al. (2015). Evaluating historical candidate genes for schizophrenia. *Molecular Psychiatry* 20:555–562.

³⁸ Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511:421–427.

histocompatibility complex, a cluster of immune genes, and part of it maps to *C4*, a gene of the complement cascade, which tags synapses for elimination. Variants that raise the expression of its *C4A* form increase risk, suggesting that excessive pruning of synapses during adolescence contributes to the disease.

3936

Each common variant has only a tiny effect: even the highest-risk form of *C4* raises a person's lifetime risk from about 1% to about 1.3%, whereas having an affected parent or sibling raises it about tenfold. But the effects add up. Summed into a *polygenic risk score*, they separate people sharply: in the largest study, people in the top tenth of scores had 8 to 20 times the risk of those in the bottom tenth, depending on the sample. Even here genes act alongside the environment: risk is also raised by factors such as famine or infection during pregnancy and heavy cannabis use in adolescence. In the end, the behavioral differences between people emerge from the interplay of genes, environment, chance, and individual choice.

1.12

Imaging

Most of what we know about circuits comes from animals, where electrodes, microscopes, and genetic tools can reach individual neurons. To watch the healthy human brain at work, neuroscience relies mostly on *functional magnetic resonance imaging* (fMRI), which measures changes in blood oxygenation throughout the brain while people perceive, remember, and decide. Kandel calls it the dominant approach in humans for measuring biological processes and linking them to behavior.

What fMRI measures

MRI exploits the magnetism of hydrogen nuclei, which are abundant in the water of the body. In a strong magnetic field — 3 tesla, about 60,000 times Earth's — a tiny excess of the nuclei, about ten in a million, lines up with the field, and they precess around it at a frequency proportional to its strength, the *Larmor frequency*: about 128 MHz at 3 tesla. A radio-frequency pulse at that frequency tips them out of alignment, and as they precess in step they induce an oscillating current in the scanner's receiver coil. The signal fades as the nuclei fall out of step with one another — a decay measured by the time constant T_2 , or T_2^* when local distortions of the field are included — and as they relax back into

39 Sekar A et al. (2016). Schizophrenia risk from complex variation of complement component 4. *Nature* 530:177–183.

alignment (T_1). Small gradients added to the main field make the frequency depend on position, which is how the signal is mapped to locations in the brain. Fast pulse sequences, such as *echo-planar imaging*, acquire a slice in under a tenth of a second and the whole brain every second or two.

The functional part comes from blood. Hemoglobin that has released its oxygen is paramagnetic: it distorts the local magnetic field, so nearby nuclei fall out of step faster and the signal dims. When neurons in a region become active, blood flow to the region increases more than their oxygen consumption does, so the blood becomes more oxygenated and the signal brightens. This *blood-oxygen-level-dependent* (BOLD) signal, discovered by Seiji Ogawa in 1990,^{40,37} tracks neural activity indirectly, through a chain of events called *neurovascular coupling* whose mechanisms, including the role of astrocytes, are still being worked out. It is slow: after a brief burst of neural activity, the BOLD response begins to rise after a second or two, peaks at about five seconds, and returns to baseline after 12 to 15 seconds (Figure 1.24). And it is coarse: each measurement element, or *voxel*, is typically 2–3 mm across and contains hundreds of thousands to millions of neurons.^{41,38} What exactly the signal reflects is still debated. It seems to follow the synaptic input and local processing in a region more closely than its output spikes; it cannot tell excitation from inhibition; and blood flow can even rise without extra firing, as when an expected visual stimulus fails to appear.

⁴⁰ Ogawa S, Lee TM, Kay AR, Tank DW (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *PNAS* 87:9868–9872.

⁴¹ Logothetis NK (2008). What we can do and what we cannot do with fMRI. *Nature* 453:869–878.

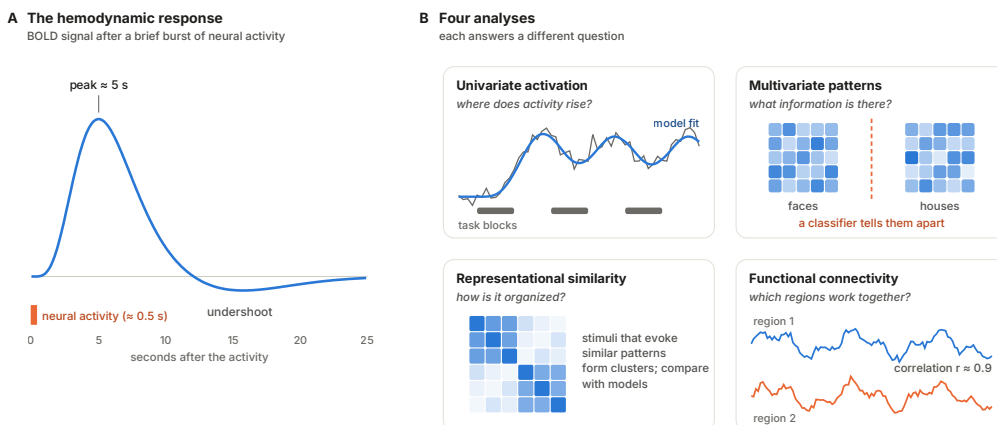


Figure 1.24 The BOLD signal and how it is analyzed. A. A brief burst of neural activity produces a slow hemodynamic response that peaks about five seconds later. B. Univariate analysis fits each voxel's time series with the task's timing convolved with that response, to find where activity rises. Multivariate pattern analysis reads what is represented in patterns across voxels, either with classifiers or, in representational similarity analysis, by comparing which stimuli evoke similar patterns. Functional connectivity correlates activity between regions.

Analyzing fMRI data

Raw fMRI data first go through *preprocessing*: correcting for head motion and for the different acquisition times of the slices — up to a second or two apart within one image of the brain — filtering out slow drifts, smoothing, and aligning each brain to an anatomical template so that people can be compared. Three families of analysis then answer different questions:

- **Univariate activation** asks *where*: which voxels respond more in one condition than in a matched control? Each voxel's time series is modeled as the task's timing convolved with the hemodynamic response, using a *general linear model*, and the difference between conditions is tested across people. This is how the *fusiform face area*, on the underside of the temporal lobe, was found: it responds much more strongly to faces than to other objects.⁴²³⁹ Variants replace the on-off timing of the task with a quantity from a model, such as memory load or a reward prediction error, or exploit *adaptation* — the weaker response to a repeated stimulus — to probe what a region is tuned to.
- **Multivariate pattern analysis** asks *what*: is there information about a stimulus or task in the pattern of activity across many voxels? It comes in two flavors. *Classification* trains a classifier on part of the data and tests it on the rest; categories such as faces, houses, and chairs can be decoded from the ventral

⁴² Kanwisher N, McDermott J, Chun MM (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience* 17:4302–4311.

temporal cortex, even from regions that do not respond most strongly to them.⁴³⁴⁰ *Representational similarity analysis* asks how the information is organized, by comparing which stimuli evoke similar activity patterns — across regions, people, species, and computational models such as the layers of a deep network.⁴⁴⁴¹

- **Functional connectivity** asks *which regions work together*, by correlating their activity over time. Two regions can correlate simply because both respond to the same stimulus, so responses evoked by the task are removed first. Even at rest, the brain is organized into large networks of regions whose activity fluctuates together; one of them, the *default mode network*, is most active when a person is not engaged in an external task.⁴⁵⁴²

What fMRI has taught us

fMRI has shaped neuroscience in three ways. It has guided research in animals: after face-selective regions were found in humans, fMRI in monkeys revealed similar “face patches,” and electrodes placed in them found that nearly all of their neurons respond selectively to faces.⁴⁶⁴³

It has challenged theories from psychology, often by accident, because it records the whole brain at once. Activity in prefrontal and parietal cortex while an experience is encoded predicts whether it will later be remembered, so memory formation involves more than the hippocampus.⁴⁷⁴⁴ And the hippocampus, thought to serve only memories that can be consciously recalled, takes part — interacting, and even competing, with the striatum — in trial-and-error learning tasks long considered unconscious.⁴⁸⁴⁵ Together these findings loosened the link between the hippocampus and conscious awareness, and brought human memory research closer to animal studies.

⁴³ Haxby JV et al. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425–2430.

⁴⁴ Kriegeskorte N, Mur M, Bandettini P (2008). Representational similarity analysis — connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience* 2:4.

⁴⁵ Raichle ME et al. (2001). A default mode of brain function. *PNAS* 98:676–682.

⁴⁶ Tsao DY, Freiwald WA, Tootell RBH, Livingstone MS (2006). A cortical region consisting entirely of face-selective cells. *Science* 311:670–674.

⁴⁷ Wagner AD et al. (1998). Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science* 281:1188–1191. Brewer JB et al. (1998). Making memories: brain activity that predicts how well visual experience will be remembered. *Science* 281:1185–1187.

⁴⁸ Poldrack RA et al. (2001). Interactive memory systems in the human brain. *Nature* 414:546–550.

And it has tested computational models: the human striatum responds to rewards like a reward prediction error, the quantity that dopamine neurons signal in monkeys and that drives temporal-difference learning.⁴⁹⁴⁶ Deep networks trained on vision likewise predict activity along the human ventral stream, layer by layer.

Reading fMRI carefully

fMRI results are easy to overinterpret. Three cautions:

- **Correlation, not causation.** fMRI shows that a region's activity varies with a task, not that the task needs the region. Lesions, brain stimulation, and causal manipulations in animals supply that evidence; and since every method has limits — recordings from neurons are correlational too — conclusions are strongest when several methods converge.
- **Reverse inference.** Most studies reason forward, from a process they manipulate to the regions it engages. Reasoning backward — “the amygdala was active, so the person felt fear” — is valid only to the extent that the region is selective for that process, and most regions are not: one region serves many processes, and one process engages many regions.⁵⁰⁴⁷ Databases that pool thousands of studies, such as Neurosynth, now estimate how selective each region is.⁵¹⁴⁸
- **Statistics at scale.** A colored brain map is a map of statistics, not of activity: the colored voxels are those that passed a threshold. A brain image has 100,000 or more voxels, so without correction for multiple comparisons some will pass by chance. In a famous demonstration, an uncorrected analysis found “activity” in a dead salmon that was shown photographs of people.⁵²⁴⁹ Selecting voxels with one analysis and then measuring the same data inflates effects in the same way.

⁴⁹ O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ (2003). Temporal difference models and reward-related learning in the human brain. *Neuron* 38:329–337.

⁵⁰ Poldrack RA (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences* 10:59–63.

⁵¹ Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods* 8:665–670.

⁵² Bennett CM, Baird AA, Miller MB, Wolford GL (2010). Neural correlates of interspecies perspective taking in the post-mortem Atlantic salmon: an argument for proper multiple comparisons correction. *Journal of Serendipitous and Unexpected Results* 1:1–5.

Used carefully, fMRI can do remarkable things. In a study published in 2006, a patient diagnosed as being in a vegetative state was asked, in the scanner, to imagine playing tennis and then to imagine walking through her home. Her brain produced the same distinct patterns of activity as healthy volunteers', suggesting that she understood the instructions and was following them.⁵³⁵⁰

ENGINEERING LENS

A thermal camera on a data center

fMRI is a little like inferring what a data center computes from a thermal camera pointed at its racks. You can see which racks run hot during which jobs, with a delay of seconds and the resolution of whole racks, and with enough data you can even tell which job is running from the pattern of heat. What you cannot see are the packets. And the analogy flatters fMRI in one respect: a thermal camera sees the heat a job produces, while fMRI sees the cooling system's response — a supply of fresh blood that overshoots demand and can even ramp up for a job that never arrives. The analysis toolkit will feel familiar — linear models, classifiers with cross-validation, similarity matrices, correlation networks, mostly in open-source packages such as AFNI, FSL, and SPM, with public datasets on OpenNeuro — and so will its pitfalls: selecting voxels and evaluating on the same data is data leakage, and testing 100,000 voxels without correction is the multiple-comparisons problem in its purest form.

No single method sees everything (Figure 1.25). Electrodes resolve individual spikes but sample a tiny fraction of the neurons; calcium imaging sees thousands of identified cells but mostly in animals and near the surface of the brain; EEG and MEG, which record electric and magnetic fields outside the head, resolve milliseconds but blur centimeters; fMRI covers the whole human brain at the scale of millimeters, but only second by second. Much of modern neuroscience consists of combining them.

⁵³ Owen AM et al. (2006). Detecting awareness in the vegetative state. *Science* 313:1402.

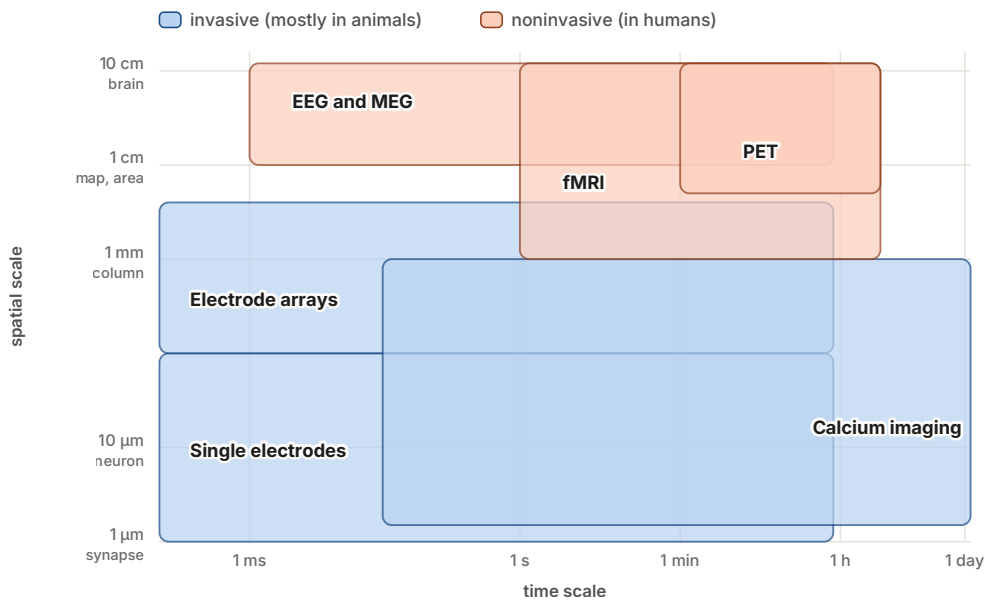


Figure 1.25 The resolution of methods for measuring brain activity. Each method covers a range of spatial scales (vertical) and temporal scales (horizontal). Invasive methods, used mostly in animals, reach single cells and milliseconds; noninvasive human methods trade spatial or temporal detail for coverage. Ranges are approximate.

1.13

Recap

This chapter assembled a system overview of the brain: seven major parts with specialized functions, built from neurons and glia, signaling with graded potentials and all-or-none spikes, wired by a handful of motifs into reflexes, sensory and motor pathways, modulatory broadcast systems, and memory loops; computing with population codes, hierarchies, expansions, and recurrent dynamics; learning through Hebbian, error-driven, and reward-driven plasticity; built and tuned by genes in constant interplay with the environment; and observable in humans, slowly and indirectly, with fMRI. Nine principles summarize it (Table 1.3).

Table 1.3 The principles introduced in this chapter.

#	PRINCIPLE	IN ONE SENTENCE
1	Neuron doctrine	The nervous system is built from discrete cells that signal to one another at synapses.
2	Cellular connectionism	Behavior arises from groups of neurons connected in precise ways.

#	PRINCIPLE	IN ONE SENTENCE
3	Localization of function	Specific regions carry out specific elementary operations.
4	Distributed processing	Complex behaviors use many specialized regions working in series and in parallel.
5	Dynamic polarization	Signals flow one way through a neuron: dendrites, cell body, axon, terminals.
6	Connectional specificity	Neurons connect to particular partners at particular sites, not at random.
7	Labeled lines	A spike's meaning is set by its pathway, not by its shape.
8	Synaptic plasticity	Experience changes the strength and number of synapses, which is thought to be the basis of learning.
9	Genes and environment	Genes build and tune circuits, experience reshapes them, and behavior is not inherited directly.

For engineers, four themes are worth carrying into the chapters ahead:

- **Organization beats component speed.** The brain's parts are slow, noisy, and frugal; its strength is the way they are wired — with specificity, in deep, parallel, recurrent architectures.
- **Learning is plural.** Unsupervised, supervised, and reinforcement learning coexist in different structures, and neuromodulators set the mode in which circuits operate and learn.
- **Much is innate.** Genes specify the architecture, the initial wiring, and the learning rules — strong priors that experience then tunes.
- **Every method sees part of the system.** Claims about the brain are only as good as the measurement behind them.

The next chapter, *Neurons*, opens the cell: how membranes, ion channels, and pumps produce the resting potential, and how the action potential is generated and propagated.

Further reading

- Kandel ER, Koester JD, Mack SH, Siegelbaum SA, eds. (2021). *Principles of Neural Science*, 6th ed. McGraw Hill. Part I covers everything in this chapter in depth.
- Luo L (2020). *Principles of Neurobiology*, 2nd ed. Garland Science. A modern, experiment-driven textbook with exceptionally clear figures.
- Sterling P, Laughlin S (2015). *Principles of Neural Design*. MIT Press. The brain as an engineered system, organized around the costs of energy and information.

- Dayan P, Abbott LF (2001). *Theoretical Neuroscience*. MIT Press. The standard mathematical introduction to neural coding, dynamics, and learning.
- Lindsay G (2021). *Models of the Mind*. Bloomsbury. How mathematics and computation have shaped neuroscience, for a general audience.
- Kandel ER (2006). *In Search of Memory*. W. W. Norton. A memoir of the science behind many of this chapter's principles.