

## CHAPTER 2

# COGNITIVE NEUROSCIENCE

**T**he field of cognitive neuroscience addresses how mental functions are supported by the brain. This close relative of cognitive psychology is exploding with new findings as a result of the discovery of methods for imaging the workings of the living brain. Neuroimaging technologies have revolutionized the study of the brain, but as will be seen in this chapter, their effective use requires the behavioral measures, research strategies, and theories of cognitive psychology. It is also important to understand that the core questions of cognitive psychology cannot be answered just by viewing the brain in action. One must first know which cognitive functions, such as short-term memory, to look for in a highly complex organ. In other words, cognitive psychology provides the theories that guide the search into the structures and activities of the brain.

The chapter begins with an introduction to the problem of how the mind and brain are related to each other. Next, a brief tour of functional neuroanatomy is provided, followed by a discussion of the methods used in cognitive neuroscience. Lastly, the fundamental properties of connectionist models are presented. As noted in Chapter 1, these are highly simplified

## 26 ● SCOPE AND METHODS

models of the brain using artificial neurons that mimic some of the basic properties of real neurons. Connectionist models are now a central tool in cognitive neuroscience and the broader field of cognitive psychology.

### ● MIND AND BRAIN

---

Cognitive neuroscience confronts us with one of the most challenging, if not *the* most challenging, philosophical and scientific questions. What exactly is the relation between the mind and the body? Put differently, how is consciousness produced by the brain? Is a mental state reducible to a physical state of the brain, or are they separate phenomena?

One view of the relation between the brain and the mind is that they are one and the same. **Materialism** regards the mind as the product of the brain and its physiological processes. The mind does not exist independently of the nervous system, according to materialism. One version of materialism contends that it is possible in theory to reduce all cognitive processes to descriptions of neural events (Crick, 1994). The reductionistic point of view was well-expressed by Dennett (1991) in these words:

The prevailing wisdom, variously expressed and argued for, is *materialism*: there is only one sort of stuff, namely *matter*—the physical stuff of physics, chemistry, and physiology—and the mind is somehow nothing but a physical phenomenon. In short, the mind is the brain. According to the materialists, we can (in principle!) account for every mental phenomenon using the same physical principles, laws, and raw materials that suffice to explain radioactivity, continental drift, photosynthesis, reproduction, nutrition, and growth. (p. 33)

Not all versions of materialism contend that the mind can be reduced to a description of brain states. An alternative version regards mental states as emergent properties of neural functioning (Scott, 1995). An **emergent property** implies that the whole is greater than the sum of its parts. It is not possible to predict the behavior of the whole just from knowing the behavior of the parts. In addition, it is necessary to understand how all of the parts interact with one another to produce the whole. A mental state can be viewed, then, as a whole that is more than the sum of the individual neurons firing. Regarding the mind as an emergent property is mentalistic but stays within the confines of materialism. Mental experience depends on, and is a functional property of, an active living brain. Sperry (1980) explained the mentalistic approach to materialism in the following passage:

Once generated from neural events, the higher order mental patterns and programs have their own subjective qualities and progress, operate, and interact by their own causal laws and principles which are different from and cannot be reduced to those of neurophysiology. (p. 201)

An alternative to materialism contends that attempts to connect mental states with brain states are mistaken. **Dualism** holds that the mind is an immaterial entity that exists independently of the brain and other bodily organs. This idea can be traced at least as far back as the French philosopher René Descartes. For a dualist, the attempt to reduce mental states to brain states is mistaken because it misinterprets correlation as causation. The dualist account recognizes that a subjective experience is correlated with activities in the brain. But as all students of psychology are aware, correlation does not prove causation. Perhaps mind and brain are correlated and have no influence on each other, or perhaps the mind actually causes brain activity rather than vice versa. Descartes assumed, as do contemporary dualists, that the immaterial mind interacts with the brain through a flow of information in ways not yet understood (Eccles, 1966, 1994; Popper & Eccles, 1977).

Clearly, these deep fundamental questions will not soon be resolved. But progress in cognitive psychology and cognitive neuroscience does not depend on resolving them, and measurements at different levels of analysis are appropriate and necessary. Measurements of brain activity can be useful, but they are not sufficient by themselves. Behavioral measurements such as verbally reporting a memory, describing thoughts leading to the solution of a problem, and making a decision and rapidly pressing a button reveal the mind in a way that brain activity cannot. Cognitive psychologists, then, often adopt dualism as a methodological approach to research, as Hilgard (1980) observed:

My reaction is that psychologists and physiologists have to be modest in the face of this problem (consciousness) that has baffled the best philosophical minds for centuries. I do not see that our methods give us any advantage at the ultimate level of metaphysical analysis. A heuristic solution seems to me to be quite appropriate. . . . That is, there are conscious facts and events that can be shared through communication with others like ourselves, and there are physical events that can be observed or recorded on instruments, and the records then observed and reflected upon. Neither of these sets of facts produces infallible data. . . . It is the task of the scientist to use the most available techniques for verification of the database and for validation of the inferences from these data. (p. 15)

---

For materialists, mental experiences can be reduced to states of the brain, or they may be an emergent property, meaning that the mind is different from the sum of the activity of neurons. For dualists, mental states are correlated with brain states and may even interact with neural processes, but the mind is not seen as rooted in matter.

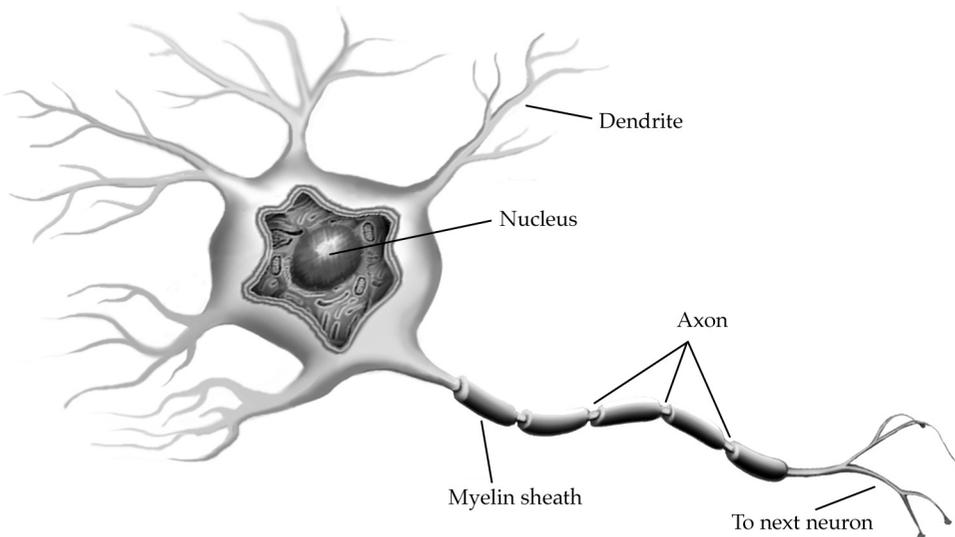
---

---

The cognitive sciences today recognize that behavioral techniques are needed to measure mental states at the same time as neural techniques are needed to measure brain states. Neither replaces the other.

---

## 28 ● SCOPE AND METHODS

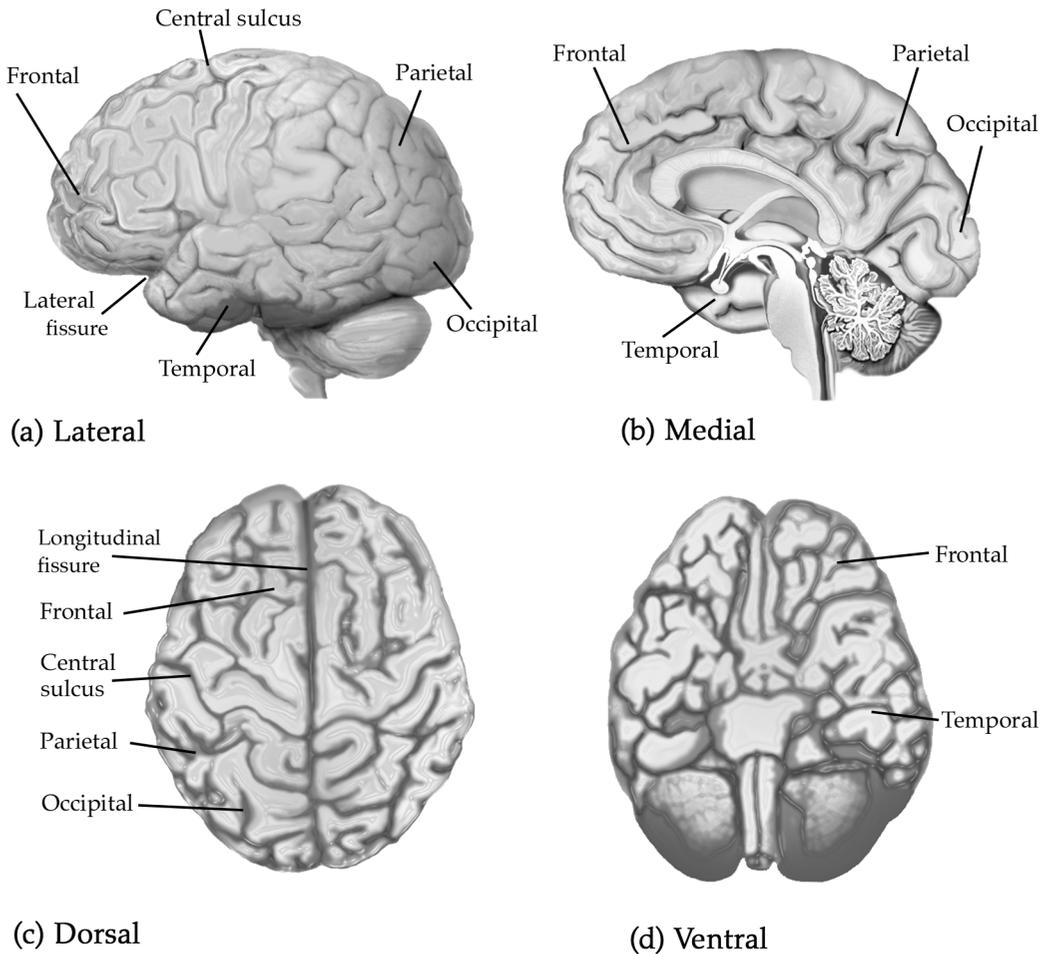


**Figure 2.1.** The basic components of a neuron.

## ● FUNCTIONAL NEUROANATOMY

The human brain may well be the most complex structure in the known universe. Consider just a few of the brain's properties to understand this point (Sejnowski & Churchland, 1989). A neuron is 1 of about 200 different types of cells that make up the 100 trillion ( $10^{14}$ ) cells of the human body. As shown in Figure 2.1, a neuron includes dendrites for receiving signals from other neurons, a cell body, and an axon for transmitting a signal to other neurons via a synaptic connection. This is an idealized illustration of one of several classes of neurons that vary in the size, shape, number, and arrangements of their dendrites and axons. The dendrites of a single neuron may receive as many as 10,000 synaptic connections from other neurons. The central nervous system is comprised of 1 trillion ( $10^{12}$ ) neurons of all kinds and about 1,000 trillion ( $10^{15}$ ) synaptic connections among these neurons (see Figure 2.1).

At a larger scale, the brain is organized into major structures such as the lobes of the cerebral cortex. Shown in Figure 2.2 are the four lobes from a lateral view (a), a medial view (b), a dorsal view (c), and a ventral view (d). These regions are separated in part by anatomical markers called the central sulcus, lateral fissure, and longitudinal fissure. The lobes of the neocortex are divided into a left and right hemisphere by the longitudinal fissure. Large folds in the cortex identify the boundaries among four lobes of



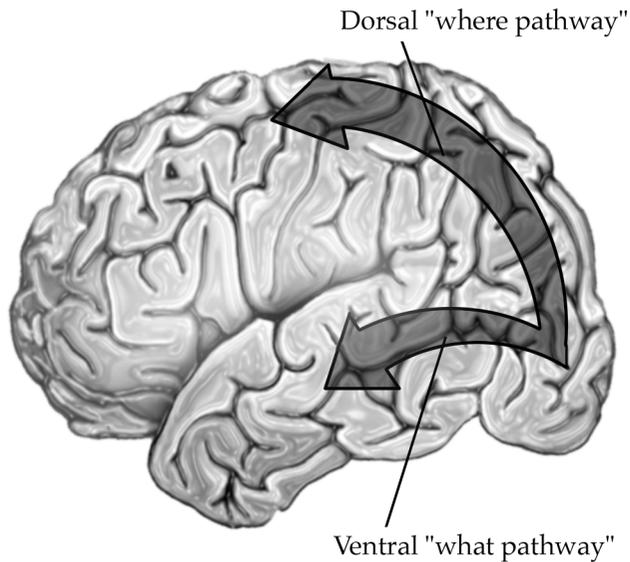
**Figure 2.2.** Four views of the lobes of the cerebral cortex.

the brain. The **frontal lobe** extends from the anterior of the brain back to the central sulcus. The **temporal lobe** lies on the side of the brain, beginning below the lateral fissure. The **parietal lobe** extends toward the rear of the brain, beginning at the central sulcus. The **occipital lobe** lies at the rear base of the brain.

## Parallel Processing

Another complexity of the brain is its dependence on parallel processing. Many separate streams of data are processed to support a single cognitive

## 30 ● SCOPE AND METHODS



**Figure 2.3.** The ventral “what” pathway versus the dorsal “where” pathway.

function. Each parallel stream involves a series of stages of processing. Consequently, it is misleading to think of a cognitive function, such as recognizing your friend across a crowded room, as dependent on just one cortical region. Although it is known that certain regions in the temporal cortex of the brain are necessary for face and other object recognition, in a parallel data stream in the parietal lobe, the location of your friend in the room is computed simultaneously (Gazzaniga, Ivry, & Mangun, 1998). As shown in Figure 2.3, a ventral or side pathway projects from the occipital lobe to the temporal lobe—the so-called “what pathway.” The dorsal or top pathway projects from the occipital lobe to the parietal lobe—the “where pathway.”

Shown in Color Plate 2 in the section of color plates are the results of a functional magnetic resonance imaging study in which the participants attended to the identity of a face (by matching it to another face) or attended to its location in a different matching condition. The red arrow marks the ventral pathway, and the green arrow marks the dorsal pathway. As may be seen, there was greater activation in the ventral pathway in the face matching condition and greater dorsal activation in the location matching condition (Haxby, Clark, & Courtney, 1997).

Although the brain uses parallel processing extensively, serial processing is also involved. For example, the streams of data corresponding to facial recognition and to identifying location both depend on an earlier serial stage of processing in the visual cortex of the occipital lobe. The occipital, parietal,

and temporal lobes all are necessary for seeing your friend. No one region is sufficient by itself, and both parallel and serial processing are necessary.

If the brain is so complex, then why bother trying to understand its structure and function when the goal is to understand cognition? One answer is that neuroscience provides converging evidence for the theories of cognitive psychology. A cognitive theory is best supported if both behavioral data and neurobiological data lead one to exactly the same conclusion. Going still further, it is possible that the results of neuroscience can point theorists in the right direction so as to avoid blind alleys. As Sejnowski and Churchland (1989) phrased this point, “Neurobiological data . . . provide essential constraints on computational theories. . . . Equally important, the data are also richly suggestive of hints concerning what might really be going on and what computational strategies evolution might have chanced upon” (p. 343). As may be seen throughout this book, there are already a number of examples in which the theories of cognitive psychology can be supported by both behavioral and neurobiological data.

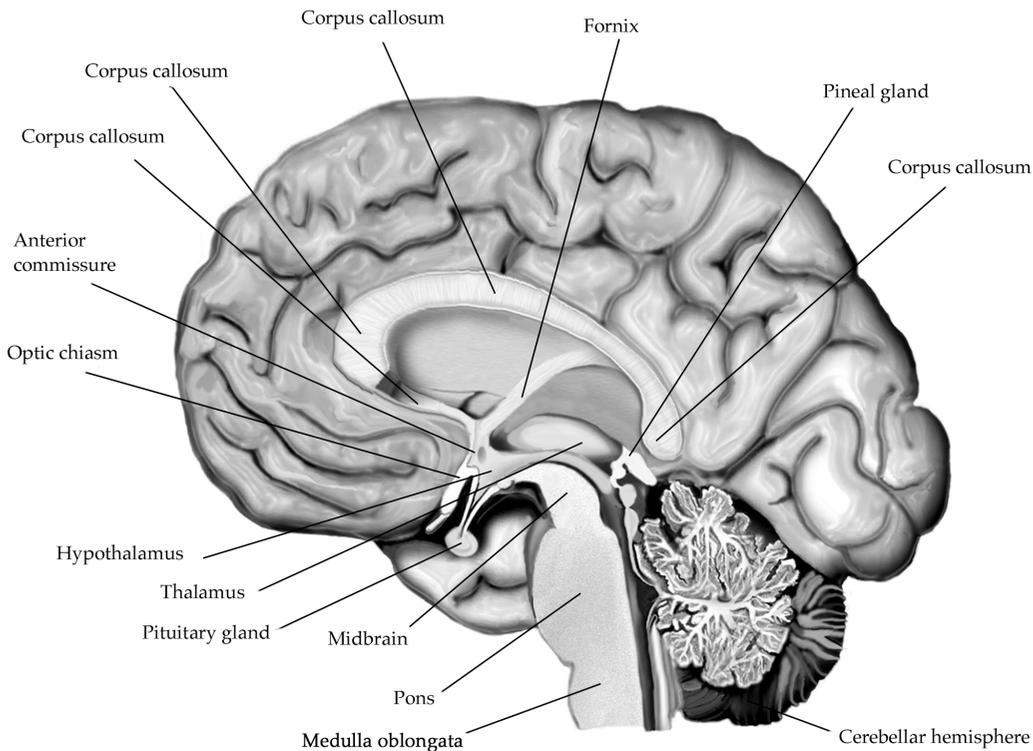
## Brain Structures and Functions

As shown in Figure 2.4, the cerebellum and brainstem lie at the base of the brain. These are very old parts of the brain that are found in species that evolved long before mammals and primates. The **cerebellum** is a large structure that lies over the brainstem at the rear of the head. The best-known function of the cerebellum is its role in coordinating complex motor skills. Signals are sent to the cerebellum regarding the position of the body and the output of the motor system. It uses this information to maintain posture and coordinate movements, enabling complex motor skills such as walking, swimming, and skiing.

*Brainstem and Forebrain.* The **brainstem** consists of the hindbrain—the medulla oblongata and pons—and the midbrain. These are identified as separate structures because they represent anatomically distinct collections of neural cell bodies or nuclei. Lying above and around the midbrain are structures of the forebrain called the diencephalon, which links the cerebral cortex with the brainstem. This includes two major structures: the thalamus and the hypothalamus. The **thalamus** is extensively interconnected with numerous regions of the cerebral cortex including, but not limited to, specific sensory areas such as vision and hearing.

The **hypothalamus** controls internal organs, the autonomic nervous system, and the endocrine system to regulate functions such as emotion, sex, hunger, and thirst (Beatty, 2001). For example, it oversees the output of the

## 32 ● SCOPE AND METHODS



**Figure 2.4.** A view of the human brain showing the hindbrain and forebrain structures.

pituitary gland in emotional regulation. Endocrine glands secrete hormones into the bloodstream as a result of signals from the pituitary, the master gland. These hormones affect the emotional expression of internal feelings such as anxiety, relaxation, anger, pleasure, happiness, surprise, fight-or-flight reactions, and sexual responses. For example, the adrenal medulla is an endocrine gland that releases adrenalin (also called epinephrine). This hormone acts to increase the rate and force of the heart beat, constricts the small arteries of the skin and internal organs, dilates the small arteries of the skeletal muscles, and elevates the levels of glucose in the blood. All of these prepare the body for the expenditure of energy—fight-or-flight reactions. The hormones released by the endocrine glands then provides feedback to the pituitary gland, the hypothalamus, or both so as to regulate their output.

It has long been known that the brainstem, basal forebrain, and diencephalon are essential for maintaining the basic life support mechanisms of the body. The alertness cycle of waking and sleeping as well as the sensory and motor signals for the respiratory system, the heart, the mouth, and the throat are controlled here, for example. Signals are brought to these brain regions via

nerve pathways or the bloodstream (e.g., pH, hormone, and glucose levels) to determine the state of body organs such as the heart, blood vessels, muscles, and skin. The function of these brain structures is to maintain, in a dynamic way, a condition of homeostasis in which bodily variables are kept within optimal ranges for the support of life. **Homeostasis** refers to a state of equilibrium of the internal environment of the body. When there is insufficient rest, food, water, or heat, for example, these brain structures initiate behaviors that change the internal state so that it falls back within an optimal range.

A useful metaphor for homeostasis is to compare these life support systems to a thermostat used to control an air conditioner during the summer. Temperature readings that exceed the set point used to keep the environment comfortable set off a response in the air conditioner. Other temperature readings have no effect at all. Homeostasis is thus achieved by maintaining room temperatures around the desired levels, even though it varies from moment to moment. The brainstem, basal forebrain, and diencephalon act essentially as a massive array of detectors whose values represent the state of the body from moment to moment (Damasio, 1999).

*Limbic System.* The **corpus callosum** is the next structure identified in Figure 2.4. This is the large band of fibers that connects the right and left cerebral hemispheres together. Surrounding the corpus callosum, there is a layer collectively known as the limbic lobe, shown in Figure 2.5. In ancient primitive species such as the crocodile, most of the forebrain consists of the limbic lobe (Thompson, 2000). Above the corpus callosum lies the cingulate gyrus, a band of cortex that runs from the front or anterior portion of the brain to the back or posterior portion. The fornix extends from the cerebral cortex to the hypothalamus. The cingulate gyrus, fornix, hippocampus, and other related structures form a larger functional unit called the **limbic system**.

The limbic system is characteristic of the mammalian brain. In more primitive species, such as the crocodile, the limbic forebrain is devoted to analyzing the smells in the environment and to preparing approach, attack, mate, or flee responses. Although emotional responses are still among the functions of the limbic system, in mammals there is less reliance on the olfactory sense of smell. Of even greater interest, some of the structures of the limbic system have taken on the cognitive functions of learning and memory. For example, the **hippocampus** is involved in the learning and storage of new events in long-term memory.

*Cerebral Cortex.* The remaining aspect of the forebrain is the cerebral cortex. The deep nuclei of the diencephalon and basal ganglia are surrounded by fatty myelinated fibers that appear white in color. The cerebral cortex, on the other hand, is called gray matter because of the grayish appearance of its

---

The brainstem, basal forebrain, and diencephalon are essential for maintaining the basic life support mechanisms of the body. They provide homeostatic control over variables such as internal temperature, pH, hormone, and glucose levels.

---

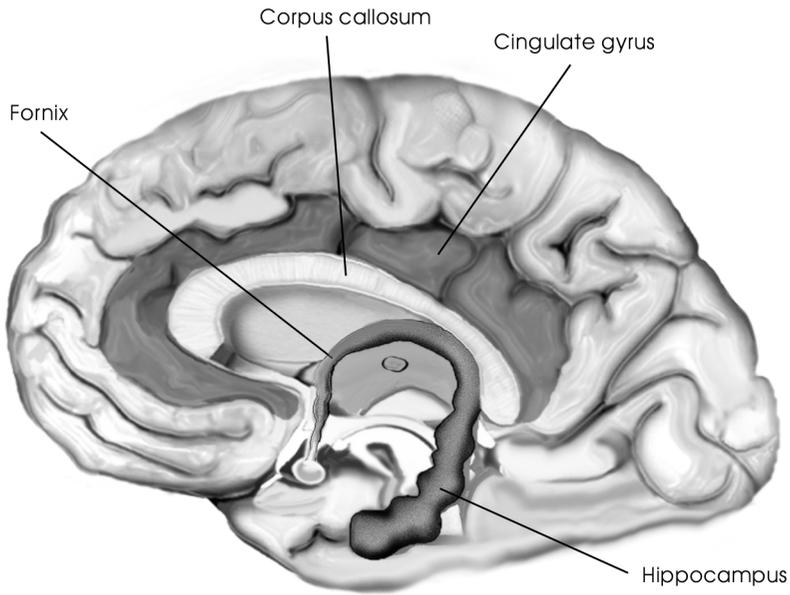


---

The limbic system consists of the limbic lobe and subcortical structures such as the hippocampus. Its functions include emotion, learning, and memory.

---

## 34 ● SCOPE AND METHODS



**Figure 2.5.** A view of some of the structures constituting the limbic system as seen in the right hemisphere with the left hemisphere removed.

unmyelinated, densely interconnected neurons. The overall thickness of the cerebral cortex averages only about 3 millimeters, arranged in layers parallel to each other and the surface of the brain (Gazzaniga et al., 1998).

The most recently evolved parts of the cerebral cortex, which is well-developed only in mammals, is called the **neocortex**. In humans, this comprises most of the cerebral cortex. The total surface area of the human cerebral cortex is 2,200 to 2,400 square centimeters, but most of this is buried in the depths of the sulci (Gazzaniga et al., 1998). To pack that much neural tissue in the small space of the human cranium is no small challenge. The evolutionary solution to this problem was to fold the cortex, creating the convoluted surface seen clearly in Figure 2.2 presented earlier. Each enfolded region is a sulcus. Cortical regions within these lobes have been mapped extensively based on how the neurons in those regions appear in structure and on how they are arranged with respect to each other.

Nearly half a century ago, brain surgeons began using direct electrical stimulation of the cortex to identify the regions that needed to be carefully spared during surgery to control epileptic seizures that failed to respond to drug treatments. The surgeons needed to remove the tissue causing the seizures while sparing the tissue that supported cognitive and behavioral functions such as perception, motor skills, and language. Because the central nervous system contains no pain receptors, patients remained awake during

the surgery and reported their subjective experiences. A small electrical current applied to the cortex during surgery caused no discomfort, but it did activate motor responses and sensations (Penfield, 1959).

The motor cortex, lying just ahead of the central sulcus, and the sensory cortex, lying just behind it, were mapped region by region. For example, the regions of the motor cortex were systematically stimulated, and the hand, arm, leg, or other movement produced in the patient was recorded. The same was done for the somatosensory cortex, with the patient verbally reporting the sensation experienced by each stimulation. In a similar fashion, the cortical regions that control speech production and comprehension were identified. Through this research, Penfield (1959) was able to preserve the regions of the brain that serve spoken language, sensation and perception, and motor behaviors.

Electrical stimulation of some regions elicited what seemed to be recollections of past experiences. Penfield (1959) found, for example, that stimulating the temporal lobes produced an auditory memory of a song playing in the patient's mind, a song heard many years before. However, the interpretation of these observations is unclear. The experiences were rare and not easy to replicate in the patient. Repeating exactly the same stimulation did not produce exactly the same memory of the past. Furthermore, how can the neurosurgeon verify that the experience reported by the individual was in fact a true memory? It is possible that the stimulation created a false memory, an experience that only seemed as though it happened in the past but was actually a new event (Loftus & Loftus, 1980).

The portions of the cortex that do not elicit a sensory or motor response when stimulated are called association areas (Gazzaniga et al., 1998). For example, there are association areas in the temporal and parietal cortex that receive inputs from the primary visual cortex of the occipital lobe. These regions, as noted earlier, process visual inputs so as to recognize objects and specify their locations. Recall that multiple regions of the brain are required for complex cognitive functions such as memory, perception, and language. Even a sensorimotor skill such as riding a bicycle depends on more than just the somatosensory and motor cortical regions. Many regions of the brain are recruited to maintain balance, to navigate, and to attend to traffic on the road. Indeed, even seeing the road and the locations of cars, buses, and pedestrians involves multiple cortical regions in the occipital, parietal, and temporal lobes.

The two hemispheres look like similar structures, but they do not perform the same functions in exactly the same way. Instead, the left and right hemispheres have evolved to specialize to a degree in particular cognitive functions (Ornstein, 1997). For example, the left hemisphere specializes in producing and comprehending language. For its part, the right hemisphere specializes in recognizing faces and processing the spatial relationships among objects.

---

Some functions are known to be localized in the regions of the neocortex such as the sensory and motor regions on either side of the central sulcus in the parietal and frontal lobes, respectively. Regions critical for language are located in the left hemisphere, whereas facial recognition and spatial processing depend on regions in the right hemisphere.

---

## ● METHODS OF COGNITIVE NEUROSCIENCE

---

The focus here is on three of the most widely used methods of studying the functions of brain structures. These are the lesion, electrophysiology, and neuroimaging methods. A treatment of all methods of cognitive neuroscience is beyond our scope here. Moreover, behavioral neuroscience studies using animals in learning tasks and recordings of single neurons in the brain, plus studies in which lesions are created in the brains of animals, fall beyond the scope of this chapter but are fundamental to the scientific understanding of cognition and the brain. For example, the model of long-term memory that is introduced in Chapter 5 rests as much on animal research as it does on human research.

### Lesions

The oldest method of studying the function of the brain is to examine individuals who have suffered damage to brain tissue through accidents, strokes, and diseases of the brain such as Alzheimer's and Parkinson's disease. For example, in the 19th century, Paul Broca reported a case study of "Tan," a man whose speech ability was reduced to saying the word "tan" repeatedly as a result of brain damage. Such tragic circumstances have provided the data for the field of clinical neuropsychology, which seeks to correlate specific lesions in the brain with specific kinds of behavioral and cognitive deficits. Lesions have also been experimentally created in rats, rabbits, monkeys and other mammals to determine the function of the damaged area. With the exception of psychosurgery performed on psychiatric patients, lesions have not been created in humans for ethical reasons. Indeed, many have questioned the ethics of treating even severely disturbed psychiatric patients with lesions in the frontal lobe and limbic system.

Until recently, clinical neuropsychology was limited to verifying the exact location of a lesion only after the death of a patient through postmortem examination of the brain. For example, Broca discovered that Tan's brain was damaged in the left frontal lobe. This became known as Broca's area when additional patients with speech disorders turned out to also suffer from lesions there. Today, the development of neuroimaging methods has allowed one to detect which regions of the brain have been damaged as the result of a stroke. This has hastened progress in using lesion case studies to understand how the brain supports cognition.

Lesion research is often based on individual case studies rather than on group results. Although most research in cognitive psychology is based on

experiments in which the results for a group of people are averaged together, this approach can cause problems in cognitive neuroscience. For example, in a group of stroke victims, the exact locations and extent of the damage vary from one individual to the next. These anatomical differences may be important for the conclusions that are reached. Consequently, it has been argued that studying the behavior and cortical damage of one individual is the best approach (Caramazza, 1992). On the other hand, the group studies support conclusions about the functions of broad areas of the brain that are likely to generalize to everyone; they are not unique to one case.

In using single cases or group studies, the investigator attempts to find two tasks that discriminate between the performance of normal controls and patients with lesions in a particular region of the brain (Gazzaniga et al., 1998). The objective is to find evidence that one cognitive function is served by one brain region, whereas a different function is served by another brain region. To reach this conclusion, the investigator seeks to find double dissociations in which the specific type of brain injury affects performance in two tasks in different ways.

In general terms, a **double dissociation** refers to situations in which an independent variable affects Task A but not Task B, and a different variable affects Task B but not Task A. One independent variable might be a lesion in the parietal cortex as compared with normal controls. A second independent variable might be a lesion in the frontal lobe as compared with normal controls. To illustrate, suppose that Task A measures planning in problem solving and Task B measures locating objects in space. If it can be shown that frontal lobe damage disrupts planning performance relative to normal controls but has no effect on locating objects in space, then a single dissociation has been demonstrated (see Figure 2.6). If, in addition, it can be shown that the parietal damage affects locating objects in space but not planning in problem solving, then a double dissociation has been established. The double dissociation isolates planning as a function of the frontal lobe and locating objects in space in the parietal lobe.

---

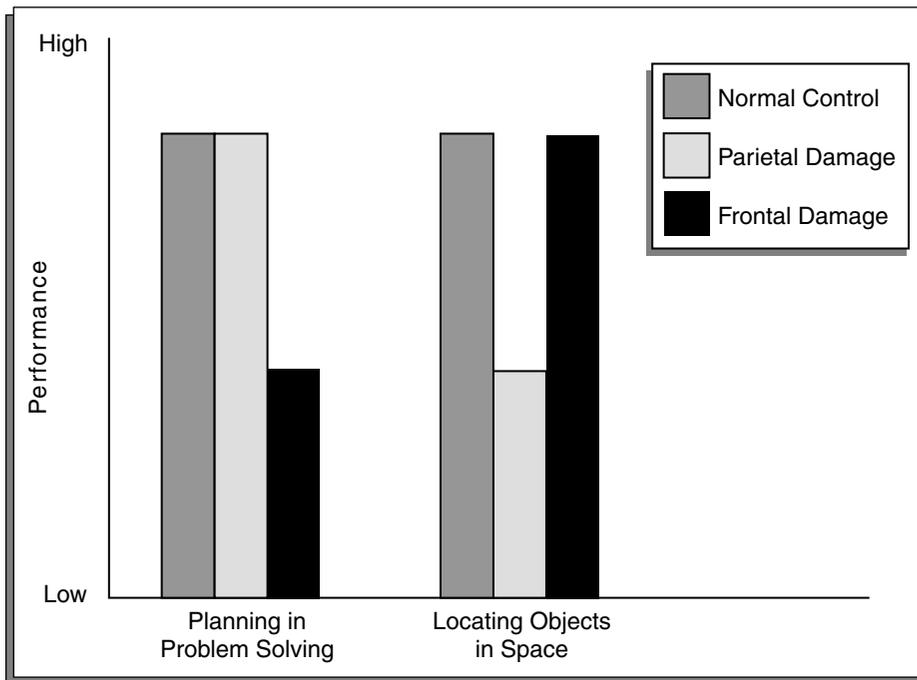
The case study method of research is a valuable tool in cognitive neuroscience. The behavior of a patient is related to the specific areas of the brain known to be damaged by a tumor, accident, or stroke.

---

## Electrophysiology

Electrophysiology reveals the activity of the brain by measuring the electric and magnetic fields that are generated by neuronal networks in the brain. As noted in Chapter 1, the electroencephalogram (EEG) is a record of the voltage changes created by the large populations of neurons activated within specific cortical regions. These brain waves can be measured with electrodes positioned on the scalp because the skull and scalp passively conduct the

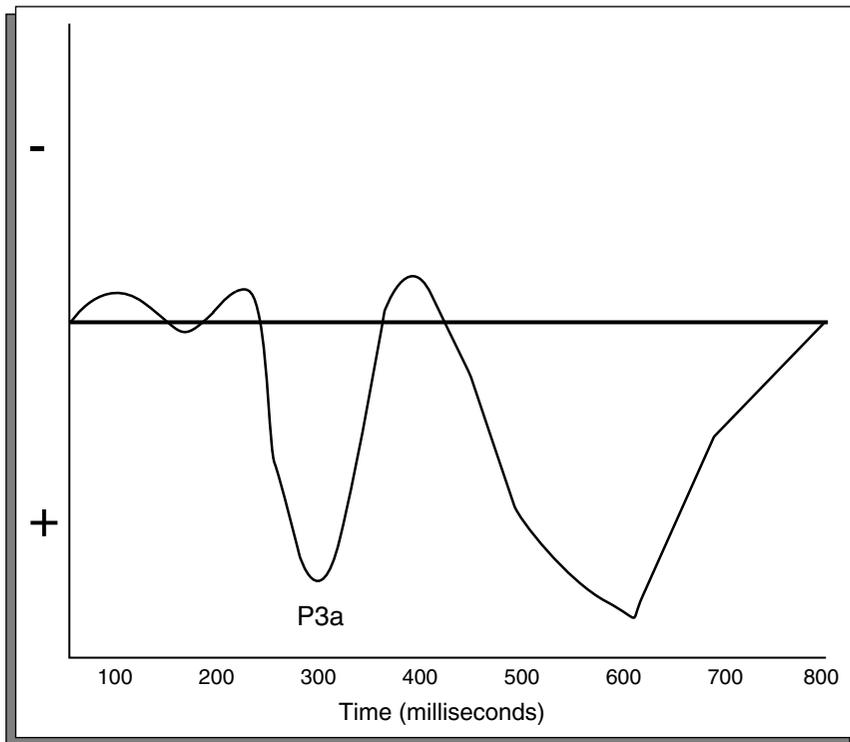
## 38 ● SCOPE AND METHODS



**Figure 2.6.** Hypothetical results of studies illustrating a single dissociation and a double dissociation.

electrical currents generated by the brain. The EEG led to the discovery that different brain wave patterns are correlated with different states of consciousness such as wakefulness, deep sleep, and dreaming.

The EEG provides a continuous measure of global changes in brain activity as a person carries out cognitive tasks. It permits one to study brain activity even in long and complex tasks. However, one drawback to the method is that changes in EEG activity caused by a particular stimulus are difficult to observe. Many responses are occurring simultaneously that have nothing to do with the particular stimulus of interest. Often times, investigators would like to know how a cortical region responds to the presentation of single stimulus such as a flash of light or the presentation of a word or picture. To this end, it is necessary to present the stimulus of interest on numerous trials. The EEG records from the trials are averaged together, making certain they are aligned with respect to the exact moment of stimulus presentation. All brain responses that are irrelevant to the stimulus are washed out of the picture through this averaging process, leaving only the response that the

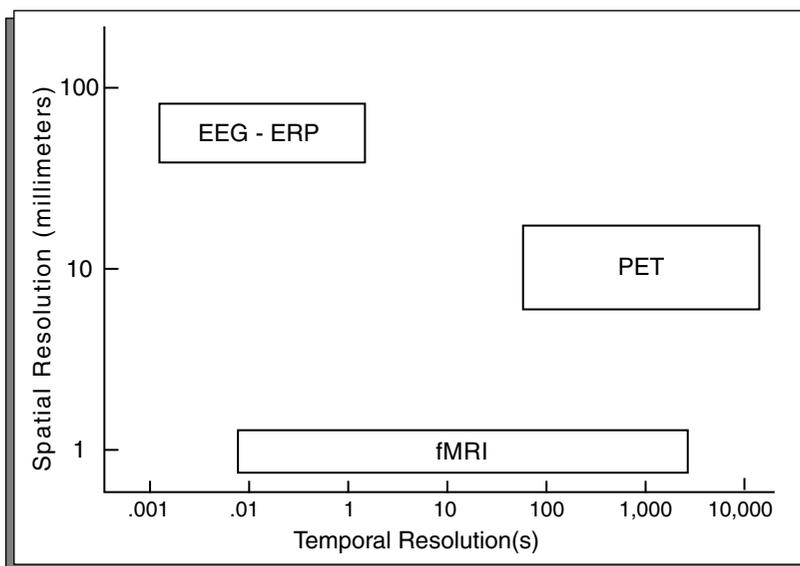


**Figure 2.7.** An idealized P3a ERP elicited 300 milliseconds after the presentation of a novel unexpected visual event. By convention, positive voltage changes are plotted below the x axis.

investigator is seeking. An EEG signal that reflects the brain's response to the onset of a specific stimulus is called an **event-related potential (ERP)** or simply an evoked potential.

To illustrate ERPs, consider the response of the brain to the presentation of a novel stimulus. An ERP called the P300 component (also known as the P3a) is the positive peak in the EEG signal that occurs 300 milliseconds after onset of an attention-getting stimulus, as shown in Figure 2.7. This component arises from an individual orienting to a novel stimulus and can be readily observed when recording from regions in the frontal lobe (Knight, 1996). Researchers use an "odd ball" task in which participants attend and count to an infrequent stimulus (e.g., red dot) while ignoring the frequent occurrences of another stimulus (e.g., green dot). In normal individuals, a novel

## 40 • SCOPE AND METHODS



**Figure 2.8.** The spatial (y axis) and temporal (x axis) sensitivity of different neuroimaging techniques.

red dot elicits a P3 ERP associated with detecting and remembering its occurrence. It turns out that this response is absent in alcoholics, however, even when they have quit drinking. Abstinent alcoholics display a diminished or delayed ERP in the odd ball task, reflecting a long-term impairment in the processing of novel information (Rodriquez, Porjesz, Chorlian, Polich, & Begleiter, 1999). The effect does not reflect alcohol intoxication per se because the participant is sober when tested.

Moreover, the novelty deficit indexed by a P300 response might not even be related to the effects of chronic alcohol consumption per se. The children of alcoholics who have not yet consumed alcohol also show the same deficit in the odd ball task. Thus, this cognitive deficit may reflect a genetic predisposition to ignore novel stimuli rather than an alcohol-produced deficit. Of great importance, the ERP deficit can, in theory, be used as a marker of the genetic disorder. Children and adolescents who display this ERP deficit are vulnerable to alcohol dependence and should avoid ever starting to drink.

EEG and ERP provide information about the temporal dynamics of neural activation in the millisecond range. Such electrophysiological measures of brain activity show excellent temporal resolution (see Figure 2.8). But it is not possible to identify the specific location, within a few millimeters, of the neuronal networks that generate the evoked potentials and fields. To pinpoint the location of neuronal activity, other methods are required.

---

An ERP measures the activation of large numbers of neurons in a cortical region by detecting positive and negative voltage fluctuations on the scalp in response to a stimulus event. Multiple ERPs occur as time passes after the event is first registered.

---



**Figure 2.9.** A PET scanner at the Washington University laboratory in St. Louis, Missouri.  
SOURCE: Posner and Raichle (1994).

## Neuroimaging

Neuroimaging provides a measure of the location of neural activation generated during a cognitive task to within 3 to 10 millimeters. Two techniques now in wide use provide an indirect measure of more localized brain activity as compared with electrical scalp recordings. The first of these is **positron emission tomography (PET)**. PET uses injections of radioactively labeled water (hydrogen and oxygen 15) to detect areas of high metabolic activity in the brain before the radioactive substance decays completely and is no longer radioactive (about 10 minutes). A person undergoing a PET scan is shown in Figure 2.9. PET images require multiple scans and allow the reconstruction of a three-dimensional picture of activated regions.

The second technique is called **functional magnetic resonance imaging (fMRI)**. With fMRI, a powerful magnetic field is passed through the head to reveal detailed images of neuronal tissue and metabolic changes. Both PET and fMRI are based on the principle that as areas of brain increase their activity, a series of local physiological changes accompanies the activity and provides a way to measure it (Buckner & Petersen, 2000). PET works by detecting increases in blood flow in the vascular network that supplies a population of neurons. fMRI works by detecting changes in the concentration of oxygen in the blood. Thus, both methods reveal how the brain

---

PET and fMRI provide neuroimages of the living brain as it processes information in a cognitive task. An increase in brain activity in a region is detected by increases in blood flow with PET and by increases in blood oxygenation with fMRI.

---

## 42 ● SCOPE AND METHODS

supports behavior in a cognitive task by measuring local changes in blood properties. Because changes in blood flow and oxygenation take a few seconds to occur, the neuroimaging methods do not provide the temporal resolution found with evoked potentials (see Figure 2.8). The color plate section of the book includes several examples of PET and fMRI images.

*Interpreting Neuroimages.* A high degree of neural activation in one region in the brain provides evidence that it is necessary for the cognitive function under investigation. It does not mean that the region is sufficient, all by itself, for the function in question. The brain processes multiple streams of data in parallel, and multiple structures are typically activated in any task. Whether all of the necessary regions turn up in a neuroimaging study depends on the control task used in the subtraction method introduced in Chapter 1. If the control task used to subtract out the “irrelevant” activation happens to tap the other supporting areas, then the very design of the study prevents them from showing up in the final results. Determining the right control task is not a trivial concern.

Once the functionality of a given brain region is known, it is possible to use neuroimaging to identify which processes are invoked by a given task (Smith, 1997). For example, it has now been established by converging evidence from lesion data, direct electrical stimulation of the cortex, and neuroimaging findings that Broca’s area mediates speech. If a task shows a 10% increase in blood flow in this left frontal area, then one can conclude that speech was produced even if it was subvocal without the participant uttering a single word. Such implicit speech might well occur, for example, when a participant silently rehearses a list of words or silently plans a solution to a problem. Changes in blood flow can detect this cognitive activity without requiring the participant to think aloud as in verbal protocols.

## ● CONNECTIONIST MODELS

---

As explained in Chapter 1, the digital computer provided a convenient analogy for understanding the architecture of the mind. Symbolic models were developed that shared key features in common with digital computers. Computations on information received by the senses were carried out in discrete serial steps such as encoding, memory storage, decision making, and response selection. A central processor used rules to process symbols similar to the rules used in computer software to process numbers and words. The digital computer helped to legitimize the study of the mind by providing an explicit model of the hidden operations of cognition that behaviorists viewed as inherently unavailable to scientists.

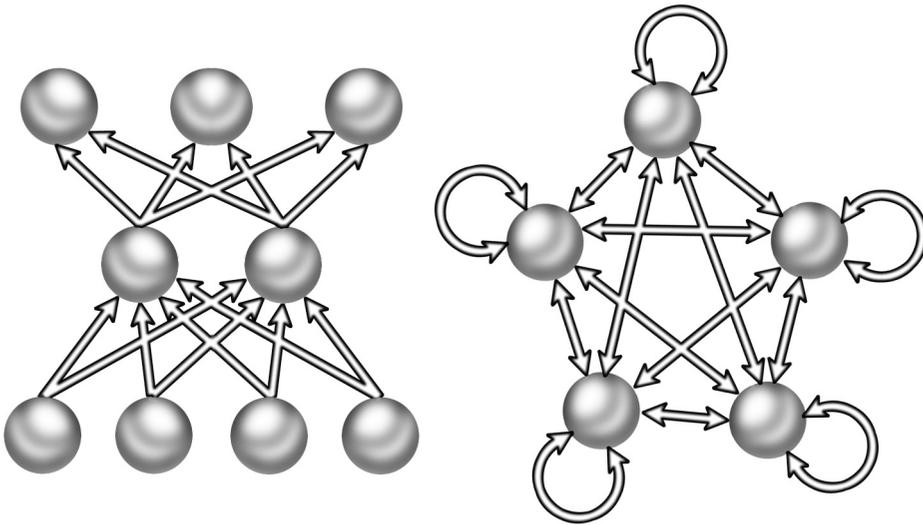
On the other hand, there are potentially important differences between the brain and a digital computer. As discussed earlier in the chapter, the brain performs computations in parallel and not in series. A single cognitive function is supported by parallel processing of multiple streams of data. Furthermore, it is not entirely clear whether the brain represents and processes symbols in the same way as a computer does. Although it makes sense for a computer to represent, say, a word as a unitary symbol, the brain may use a distributed representation. Recall from Chapter 1 that a connectionist representation of a word is distributed over multiple units, each of which codes one feature relevant to the word's meaning. Connectionist models are also called **parallel distributed processing (PDP)** models to emphasize these biologically inspired features that mimic the brain.

For example, consider how "coffee" could be represented in a connectionist model (Smolensky, 1988). It might include a node that codes "brown liquid" and another for "burnt odor." It would also include units that code features for different scenarios in which coffee appears. For example, nodes for "cup of coffee" would code for "upright container," "hot temperature," and "brown liquid contacting porcelain" as well as those already noted. Still other nodes would code features needed for a different scenario such as "can of coffee" (e.g., a node for "granules contacting tin"). Where in such a distributed representation is the symbol for "coffee"? It is everywhere and nowhere at the same time. All of the nodes that participate in coding the features of coffee together constitute the representation. Yet nowhere can one point to a specific node and say that this node, and not that one, is the symbol for coffee.

Neural networks are biologically inspired in the sense that they mimic the parallel computations of the brain and the use of distributed representations of knowledge. At the same time, neural networks are highly artificial because they are blatant but intentional simplifications of the brain. Each node is like an idealized neuron, and each connection is like an idealized synapse. They display none of the complexities of real neurons and synapses. The neural network operates with a very small number of nodes as compared with the billions found in a real brain. Finally, the network is designed to model a single function of the brain at a time. It is not intended as a complete replication of the brain, nor would this be of much value, for then the model would be so complicated that scientists would not understand it any better than the brain itself.

## Basics of Neural Networks

*Components of Neural Networks.* Connectionist models attempt to understand the architecture of human cognition by using highly simplified, idealized



**Figure 2.10.** Two types of connectionist neural networks: a three-layer feedforward network (left) and a fully recurrent network (right).

models of the brain itself. Such models are composed of many nodes or nodes that behave in ways that mimic neurons. As with neurons, the nodes gather input from other nodes. Not all nodes are connected to all other nodes. In a typical model, there is a layer of input nodes that mimic sensory receptors, receiving information from the environment (see the left side of Figure 2.10). Another layer of nodes mimic motor neurons and provide a response from the network. Sandwiched in between is a **hidden layer** that receives information from the input layer and sends forward information to the output layer.

The connection between two nodes mimics a synapse between two neurons. In some networks, the connections between nodes are unidirectional. In other cases, the connections are bidirectional, meaning that feedback is provided to the node that sends forward information. A node can also be connected to itself, providing what is called recurrent feedback. A fully recurrent network with bidirectional connections is shown on the right in Figure 2.10.

Just as some synaptic connections are excitatory and some are inhibitory, positive or negative weights are associated with each artificial synapse in the neural network. The weight for a given connection between two nodes changes in value as the network processes inputs, gives outputs, and provides feedback. Each **connection weight** represents the knowledge state of the network; mathematically, the weight is a multiplier of the output value of the sending node.

The net input to node  $i$  is given by this equation:

$$= \text{net}_i = \sum_j w_{ij} a_j$$

The weight between node  $i$  and node  $j$  is given by  $w_{ij}$  and the activation level of node  $j$  is given by  $a_j$ . To calculate the net input to node  $i$ , one sums the product of weight times activation for all node sources  $j$ .

Thus, if node  $i$  receives input from only one node whose output equals +1 and whose weight equals 0.5 (an excitatory connection), then its net activation equals 0.5. But if node  $i$  also receives input from an inhibitory node whose output equals +1 and whose weight equals -1.0 (an inhibitory connection), then its net activation would equal -0.5. In terms of the formula,

$$= [(+1 * 0.5) + (+1 * -1.0)] = -0.5.$$

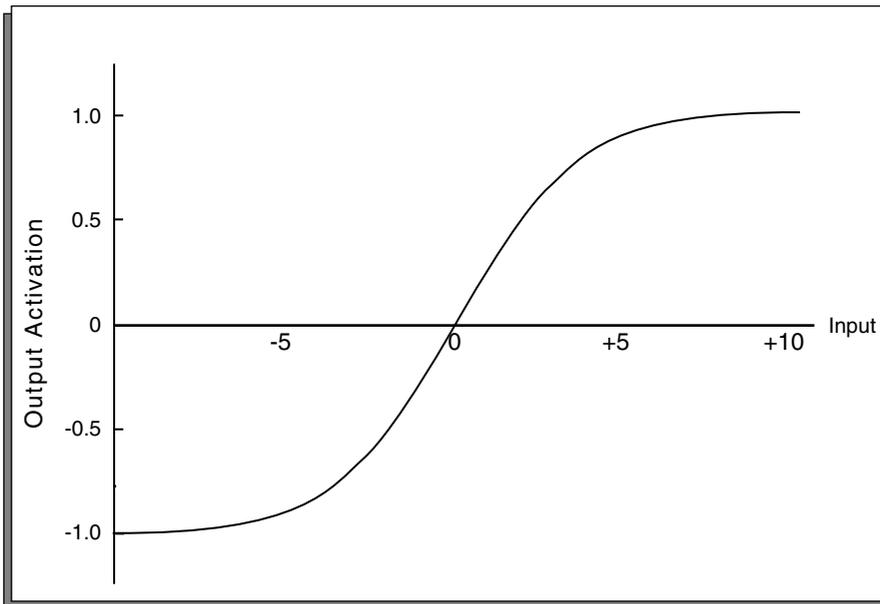
*Dynamics of Neural Networks.* Each node responds to its summed input based on an activation function. The response of the neuron is given on the  $y$  axis for different input values, ranging from negative values to positive values. A linear function, for example, would gain strength in direct proportion to the strength of the inputs. This would mean that the response is always graded, gaining strength in direct relation to the strength of the inputs. Instead, neural networks typically use a nonlinear activation function such as is provided by the sigmoid function shown in Figure 2.11. Note that it mimics the all-or-none response of real neurons for any input value less than zero and for large positive inputs. That is, for negative inputs the response is 0.0, and for large positive inputs the response is 1.0. However, graded responses, falling between 0.0 and 1.0 in value, are obtained when the inputs are small positive values, between 0 and +5. The nonlinear response of this activation function is a crucial feature of how connectionist models achieve interesting behaviors (Elman et al., 1996). Each node behaves in a categorical all-or-none fashion under certain circumstances and in a sensitive graded fashion in others.

*Logical Rules.* To grasp how neural networks behave, it is useful to consider how simplified networks implement logical rules. Suppose that there are two input nodes and a single output node, as shown in the first two cases in Figure 2.12. This is a two-layer network with no hidden nodes. Suppose further that the activation function is strictly all-or-none, assuming output values of only 0 or 1. If input activation is less than or equal to 1, then node output is “off,” taking a value of 0. If input activation exceeds 1 by any amount, then the node output is “on,” taking a value of +1.

---

The connection weights in a neural network represent its current state of knowledge; mathematically, the weights are multipliers of the output values of all nodes sending information. Some weights are excitatory (positive values), and some are inhibitory (negative values). The net input to a given node is the sum of all excitatory and inhibitory input connections.

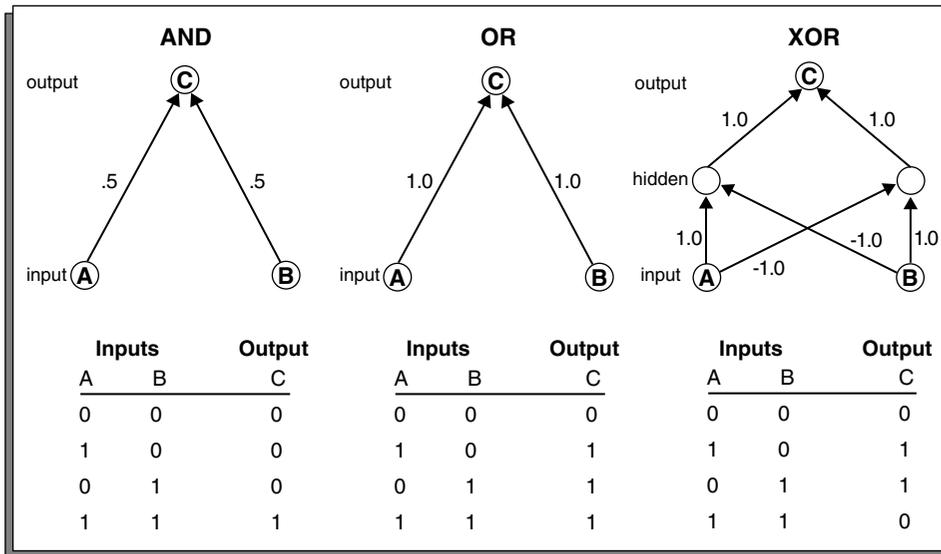
---



**Figure 2.11.** The sigmoid activation function typically used to relate inputs to output activation in each node of a neural network.

In the first case in Figure 2.12, the input patterns presented to the two nodes are shown below the two-layer network. Each input node sends an activation level to the output node equal to +1. The weight of each connection is 0.5. The net activation in this case is equal to 1.0, which triggers an “on” response from the output node. If the input from either node A or node B is less than +1, then the output is necessarily less than +1. This network models the logical AND relation. Its output is “on” if the input node on the left is +1 AND the input node on the right is +1; if one input or the other is 0 or if both inputs are 0, then the output is “off.” In the next case, the weight for each node is changed to 1.0, and now a new logical rule is implemented. The OR rule stipulates an “on” output if one input or the other is +1 or if both inputs assume a value of +1. In all three situations, the net activation of the output node will equal or exceed 1.0.

The AND and OR rules are easily modeled with two-layer networks. Input patterns that are highly similar to one another give rise to the same output in both of these rules. They differ only in whether a single input node with 0 activation is grouped with the case of both nodes being 0 (OR), on the one hand, or whether a single input node with +1 activation is grouped with the case of both nodes being +1 (AND). A much more difficult logical rule is represented in the third case in Figure 2.12. This is called the Exclusive OR



**Figure 2.12.** Two-layer neural networks can compute conjunction (AND) and inclusive disjunction (OR) logical rules. A hidden layer must be added to compute the exclusive disjunction (XOR) rule.

or the XOR rule. Now, similar inputs are not treated similarly at the output level. If one input or the other input, but not both, is +1, then the output is “on.” If neither input is 0 or if both inputs are 0, then the output is “off.” Here, then, highly dissimilar patterns must be categorized together. Take a few moments with assigning weights to the two-layer network to satisfy yourself that it fails to solve the XOR problem.

As you can verify, it is easy to get the network to produce an “off” or 0 output when the input patterns are (0, 0) and (1, 1). This can be achieved by setting the weights for each connection to 0. But this causes major problems in achieving the desired result for patterns (0, 1) and (1, 0). In these cases, we need a weight large enough so that the net activation reaches +1 when only one of the input nodes equals +1. By setting the weight at, say, +1, we solve our problem with these two patterns but foul up the results for the (0, 0) and (1, 1) cases.

The solution to the XOR problem illustrates how hidden layers can cause neural networks to behave in counterintuitive ways that are not based on similarity (Elman et al., 1996). In the third case in Figure 2.12, one hidden node is added to the network used to solve the AND problem. The connection weight between each input node and the hidden node above it is set at +1. However, the opposite hidden node is given an inhibitory connection with a weight equal to -1. The connection weights from the hidden nodes

## 48 ● SCOPE AND METHODS

to the output nodes are excitatory with weights equal to +1. Running the test patterns (0, 1) and (1, 0) now yields outputs of +1, as required by the XOR rule. For example, an input of 1 to either one node or the other results in an “on” response from the output node. Note what happens to this network when both inputs are +1, however. In this case, the inhibitory connections to the hidden layer effectively cancel the input values in computing net activation. The net of the nodes in the hidden layer is 0, resulting in an “off” response at the output node.

---

The hidden layer allows neural networks to solve the XOR problem, responding to similar inputs in different ways. It is akin to an abstract internal representation of information.

---

The hidden layer in a neural network provides abstract internal representation of the inputs. By using inhibitory connections to the hidden layer, the network treats dissimilar inputs (0, 0) and (1, 1) as alike and similar inputs (0, 0) and (0, 1) as different. The central point is that adding a hidden layer augments the power of neural networks to produce complex and often counterintuitive behaviors.

*Back-propagation of Error.* The XOR problem illustrates that, with the right architecture and weights assigned, a difficult logical rule can be modeled with a neural network. With these simplified networks, the modeler can determine the correct combination of weights that should be used to produce the desired output. As more and more nodes are added to the network and as more complex relationships between the inputs and outputs are needed, the number of computations needed for finding the right weights is too great. What is needed is a way for the neural network to learn on its own, slowly over long periods of time if necessary, a good combination of weights. So, how can neural networks learn through experience which weights should be adjusted to achieve a particular result?

A common algorithm or rule for teaching a neural network is called **back-propagation of error**. It is an illustration of supervised learning in which the specific outputs desired are known and serve as teaching values that provide feedback. Unsupervised learning can also occur in neural networks, but they fall outside the scope of this brief introduction. Another limitation of this discussion is that it illustrates only Hebbian learning. Hebb (1949) posited the following:

When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one of both cells such that A's efficiency, as one of the cells firing B, is increased. (p. 62)

Learning in neural networks is Hebbian when it takes place by altering the synaptic weights between A and B such that a future activation of A increases the probability of activating B.

The back-propagation algorithm is a procedure for training neural networks by returning an error signal from the output layer backward through hidden layers to the input layer. It aims to find the combination of weights that minimizes the error function. Back-propagation starts with the idea of comparing the weights from input nodes to those from output nodes so as to reduce the difference between the target output and the actual output (Elman et al., 1996). Because a target for the output layer is known, it is straightforward to evaluate the weights leading to these nodes and calculate how they should be changed. For the hidden layer, there is no specified target. How can one decide how much error is arising from a node weight at the hidden level if the desired output from that level is unknown? The answer is back-propagation.

The error is first calculated at the output level; the activation of an output node is subtracted from the desired output activation, called the target or teacher value. Next, the weights leading into that node are adjusted so that in future steps it is more or less activated, depending on the direction needed to reduce the error. This is done for all weights leading to the output node. If more than one output node is in the network, then these steps are repeated for each one. Finally, the “blame” for these output errors is assigned backward one level to the weights from the input layer to the hidden layer. This blame is apportioned based on (a) the errors observed on the output nodes to which a given hidden node is connected and (b) the strength of the connection between the hidden node and an output node. Thus, the error signal is propagated backward through the network to adjust the weights at the hidden layer. Although back-propagation is a useful technique, there is no guarantee that the optimal set of weights will be learned. However, a satisfactory, if not perfect, set of weights can often be found if learning takes place slowly by making only small changes to the weights with each error.

---

Back-propagation of error is a kind of supervised learning in which the specific outputs desired are known and serve as teaching values that provide feedback to the hidden layer and input layer.

---

## Modeling English Verb Acquisition

Armed with the basic concepts of neural networks, it is helpful to examine how a connectionist model explains real findings on a problem of central importance in the field, namely, language acquisition in children. It is well-known that children move through three stages in learning the correct way to produce the past tense of English verbs. Some English verbs are regular, meaning that a suffix -ed is simply added to the verb stem to produce the past tense (e.g., show/showed). Other English verbs are irregular in various ways. In some cases the past tense form is similar to the present tense (e.g., grow/grew), and in some cases it is even identical (hit/hit). In other cases, the past tense bears no obvious sound or spelling relation to the present tense

## 50 ● SCOPE AND METHODS

(e.g., go/went). Early in language acquisition, children begin to produce some irregular past tense forms correctly, saying “went” instead of “goed.” However, as children learn more and more verbs and discover that regular verbs follow a simple rule of adding -ed, overgeneralization errors begin to intrude. That is, the -ed rule is overgeneralized to the irregular verbs already learned. It is as though children unlearn the correct forms of the irregular verbs and mistakenly learn, for example, goed. Finally, with additional exposure to the language, the overgeneralization errors drop out and children produce both regular and irregular verbs with few if any errors.

Rumelhart and McClelland (1986) examined how a two-layer network learned the past tense. A set of input nodes are connected directly to a set of output nodes with no hidden layer. The input is a phonological or sound representation of the verb stem, and the output is the phonological representation of the corresponding past tense of the verb. Each node represented a specific aspect of the sound representation in both the input and output layers. So, various clusters of these nodes were able to represent the different sound patterns needed to produce the verbs and their past tenses (e.g., /g/, /o/ and /w/, /e/, /n/, /t/). For example, some output nodes represented the -ed suffix used for regular verbs. Rumelhart and McClelland used back-propagation of error to teach the network the correct English past tense for all of the verbs.

More than 400 verbs typical of daily English usage (i.e., mostly regular with some irregular) were used to train the network. The results showed that the percentage of correct past tense verbs output by the model increased rapidly for regular verb forms early in training and then leveled off. Although it continued to improve with extensive training, the gains were quite small. Because most of the verbs were regular, the network quickly settled into adding the -ed suffix. Although this was beneficial for learning the regular verbs, it caused problems for learning the irregular verbs. Early on, performance improved rapidly for the irregular verbs also but then showed a sharp reduction in the percentage of correct forms. The point where this occurred coincided with the point where the network was nearly always correct with the regular verb forms. In other words, the network was likely to produce “goed” or “hitted” by mistake, showing overgeneralization errors. These mistakes generated large error signals that decreased the likelihood of turning on the -ed output nodes the next time. So, learning the irregular verbs interfered with learning the regulars and vice versa. Slowly, and only after extensive additional training, the network also was able to learn the correct forms for irregular verbs as well as for regular verbs. Thus, the well-established phenomenon of overgeneralization of regular verb forms was duplicated in a simple neural network using a single learning process.

---

Connectionist or PDP models of the brain are implemented as neural networks where each node acts as a simplified neuron. Knowledge representations are distributed across multiple nodes, and information is processed in parallel—two features that seem to be true of the brain.

---

## SUMMARY

1. The relation of brain states to conscious states is an unsolved philosophical and scientific problem. The working assumption of many cognitive neuroscientists is materialism, which reduces mental states to brain states or regards mental states as emergent properties of the brain. Dualism is an alternative point of view that regards brain states and mental states as different entities altogether, although they may interact with one another. For example, a mental state might cause a change in the state of the brain or vice versa. To conduct research successfully in cognitive neuroscience, behavioral techniques are needed to measure mental activities (e.g., verbal reports), and neural recording techniques measure states of the brain. The aim is to relate these two parallel sets of data and not to replace behavioral measures with neurological measures. In other words, cognitive scientists adopt a methodological dualism to make progress in the field.

2. The human brain may well be the most complex structure in the known universe. The central nervous system contains on the order of 1 trillion neurons and about 1,000 trillion synaptic connections among these neurons. The organization of the brain is highly parallel, with many separate streams of data being processed to support a single function such as face recognition. Despite the complexity of interconnections, it is not the case that every neuron is connected to every other neuron through one pathway or another. Synaptic connections are either excitatory or inhibitory in their effect on the next neuron. The goal of cognitive neuroscience is to use data about the brain to help decide among alternative theories of perception, attention, memory, language, and other cognitive functions.

3. The cerebellum and brainstem are ancient structures and evolved long before mammals and primates. Lying above and surrounding the brainstem are the diencephalon and basal forebrain. These structures provide the basic life support functions of the body such as respiration and heart rate. They maintain a state of equilibrium in the internal environment of the body, called homeostasis. The limbic system lies in the next layer of neural structures and is similar in all mammals. The hippocampus is part of the limbic system and plays a critical role in emotion, learning, and memory. Surrounding the limbic system is the cerebral cortex. It appears gray in color, is arranged in layers, and averages only about 3 millimeters in thickness. Within the most recently evolved layer, the neocortex, enormous numbers of neurons are densely packed and folded, giving the brain its convoluted appearance on the surface. About 75% of the trillion neurons in the central nervous system are neocortical.

## 52 ● SCOPE AND METHODS

4. The neocortex is symmetrically divided into two hemispheres. Within each hemisphere, the frontal, temporal, parietal, and occipital lobes are distinguished. Some regions serve specific sensory and motor functions, whereas others—the association areas—play a role in numerous cognitive functions. Some functions are lateralized, meaning that one hemisphere plays a special role. For example, regions critical for language are located in the left hemisphere, whereas those involved in facial recognition and spatial processing depend on regions in the right hemisphere. It is incorrect to think of a cognitive function as completely lateralized, however. The right temporal lobe is necessary for the recognition of faces, but the visual processing of the faces in the left and right occipital lobes is also necessary.

5. Lesions or damage to cortical regions provide one way to study the cognitive functions served by the brain. Cognitive neuroscientists seek double dissociations in which one kind of lesion disrupts performance on Task A but spares performance on Task B, whereas a different kind of lesion disrupts Task B but spares Task A. Double dissociations suggest that the two brain regions damaged by the lesions support different cognitive functions, as measured by Tasks A and B. Electroencephalograms (EEGs) provide continuous recordings of the voltage changes created by large populations of neurons within a specific cortical region. An EEG signal that reflects the brain's response to a specific stimulus is called an event-related potential (ERP). Neuroimaging methods work by detecting changes in the blood supply serving the metabolic needs of activated neurons. Positron emission tomography (PET) measures blood flow and functional magnetic resonance imaging (fMRI). Using the method of subtraction, cognitive neuroscientists attempt to isolate the neural activation caused by a particular cognitive function.

6. Connectionist or parallel distributed processing (PDP) models are computer simulations that mimic basic features of the brain. The nodes of a PDP model can be activated in an all-or-none manner, and connections to other neurons can be either excitatory or inhibitory, as in real neurons. The representation of knowledge is distributed over many neurons. The connectionist architecture may include an input layer, an output layer, and a hidden layer that generates counterintuitive behaviors from the network. Neural networks can learn to provide the correct output from input received by modifying the strength of the connections among nodes. A typical way of learning relies on back-propagation of error signals from the output layer to earlier layers. Over time, the system adjusts connection weights to minimize the amount of error.

**KEY TERMS ●**

---

materialism  
emergent property  
dualism  
frontal lobe  
temporal lobe  
parietal lobe  
occipital lobe  
cerebellum  
brainstem  
thalamus  
hypothalamus  
homeostasis  
corpus callosum

limbic system  
hippocampus  
neocortex  
double dissociation  
event-related potential (ERP)  
positron emission tomography (PET)  
functional magnetic resonance  
imaging (fMRI)  
parallel distributed processing (PDP)  
hidden layer  
connection weight  
back-propagation of error

