CHAPTER TWO

THE PHILOSOPHICAL APPROACH

Enduring Questions

WHAT IS PHILOSOPHY?

Philosophy in its broadest sense is the search for wisdom and knowledge. It is the first approach we will tackle in our voyage through the different disciplines of cognitive science. There are good reasons for beginning here. Philosophy plays a vital participatory role in cognitive science. It does this not by generating results, since it is a theoretical rather than experimental discipline, but by "defining problems, criticizing models, and suggesting areas for future research" (Garfield, 1995, p. 374). More than any other discipline in cognitive science, philosophy is not limited by its subject matter or a particular theoretical stance. It is, therefore, free to evaluate and contribute to the remaining disciplines in a way the others cannot. This approach is also the oldest of the different approaches, tracing its origins back to the ancient Greeks. It is thus fitting that we begin our tour here.

The translation of the word *philosophy* yields "love of wisdom," indicating the philosopher's concern with knowledge and understanding the universe. Philosophy as a formal discipline studies a wide range of topics. In fact, there is no topic that is not fair game for a philosopher; he or she may examine politics, ethics, esthetics, and other subjects. We concern ourselves here with two branches of philosophy. **Metaphysics** examines the nature of reality. The mind–body problem is a metaphysical one at heart because it seeks to understand whether the mental world is part of the physical material world. **Epistemology** is the study of knowledge and asks questions such as, What is knowledge? How is knowledge represented in the mind? How do we come to acquire knowledge?

CHAPTER OVERVIEW

In this chapter, we survey philosophic thoughts that center on a couple of vexing issues, which can be summed up in terms of "this"

Learning Objectives

After reading this chapter, you will be able to:

1. Define the mind–body problem.
2. Compare monism and dualism as potential solutions to the mind–body problem.
3. Assess functionalism as a framework for studying the mind.
4. Formulate an example syllogism that is logically sound.
5. Describe the "hard problem" of consciousness.
6. Contrast reductionism and emergence.
7. Describe and critique Searle's Chinese room thought experiment.

versus "that." Such terminology suggests that the debates that have arisen from these issues have polarized the arguments and that there are only two possible answers to a problem. We will see that this is actually not the case and that there are multiple ways to conceptualize the issues. These issues are the mind–body and nature–nurture debates. In addition, we discuss the question of consciousness and its relation to cognitive science.

This chapter is motivated by great questions in the philosophy of mind. In the first section on the mind–body problem, we ask a very fundamental question: What is mind? Is it something that is physical or not? Is a body necessary to have a mind? These questions are primarily metaphysical in nature. In the second section on functionalism, we get more specific. If a mind is physical, then what sort of substrate can it be based on? Can a mind only be based on brains or can minds emerge from other things like computers? In the third section, we address the issue of knowledge. How does information get into our "heads"? Are we born knowing certain things or is information primarily learned? These are epistemological questions, since they concern information. In the fourth and final section, we look at perhaps the most contentious and debated issue in the modern philosophy of mind—consciousness. Here we will address questions like What is consciousness? Can it exist in other animals? What exactly is happening in our brains when we have a particular conscious state?

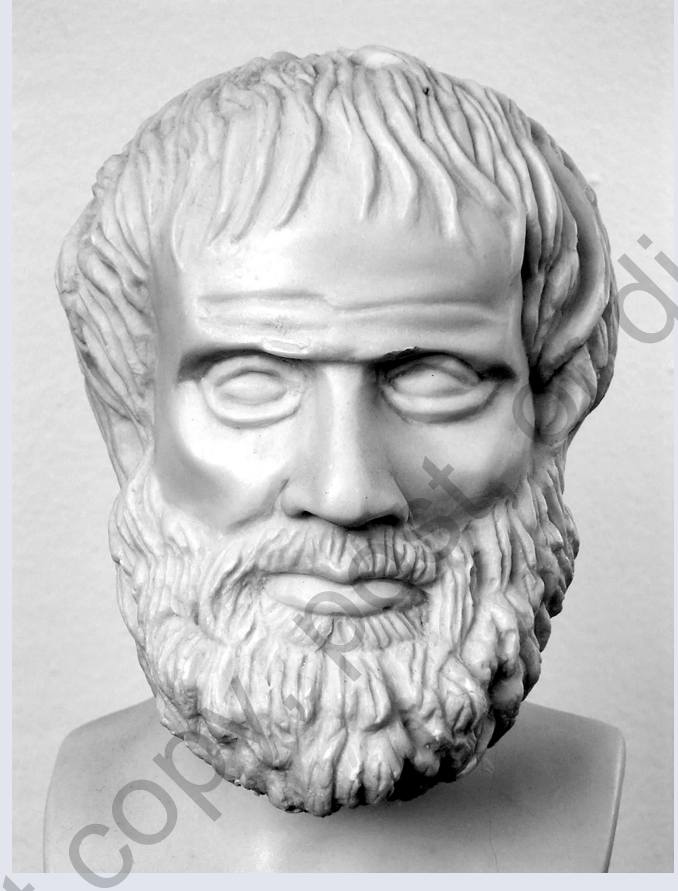
THE MIND–BODY PROBLEM: WHAT IS MIND?

The mind–body problem addresses how psychological or mental properties are related to physical properties. The debate stems from a fundamental conception about what the mind is. On the one hand, we have the brain that is material and physical. It is made up of substances that we can measure and understand. The mind could be thought of in the same way, as simply a physical thing. On the other hand, there are those who argue that the mind is something more. They say we can't equate our subjective conscious experiences, such as beliefs, desires, and thoughts, with something as mundane as the human nervous system. They say that the mind is nonphysical and consists of something resembling a soul or spirit. The mind as a nonphysical entity inhabiting the brain or some other physical system is sometimes called "the ghost in the machine."

The first question of the mind–body problem is metaphysical and refers to the nature of what mind is. Is the mind physical or something else? A second and more specific question concerns the relationship between these two things. If we were to assume that there are two such entities, then what is the causal relationship between them? Does the mind control the body, or does the body control the mind? There are many theories supporting the directions this control takes. Some theories argue that only one exerts control; others state that they both control each other; additional theories state that the two work in parallel but that neither has any causal influence.

Our discussion in this section is structured around basic conceptions of the nature of mind. According to **monism**, there is only one kind of state or substance in the universe. The ancient Greek philosopher Aristotle (384–322 BCE) was a monist (see Figure 2.1). He characterized the difference between mind and body as the difference between form and matter. One way to think of his notion is to consider a lump of clay. It

Figure 2.1 **A bust of the early Greek philosopher Aristotle. He believed that there was no substantial difference between mind and matter.**



is made up of physical matter, and we can think of it as corresponding to the brain. We can shape the clay with our hands into different forms—for example, we can roll it into a ball or flatten it into a pancake. The shapes the clay can assume, Aristotle implied, are like the different thoughts the mind can take on when it undergoes different patterns of activity. These shapes are just different physical states and do not constitute any non-physical or spiritual substance.

In **dualism**, one believes that both mental and physical substances are possible. Plato, another Greek philosopher (427–347 BCE), was a dualist. Plato was Aristotle's

teacher, but the two held quite different views. Plato believed that the mind and the body exist in two separate worlds. Knowledge of the mind, he thought, exists in an ideal world of forms, which is immaterial, nonextended, and eternal. The body instead resides in a world that is material, extended, and perishable. There are crucial differences between the objects of one world and those of the other. Mental ideas such as "circle" that reside in the ideal world of forms are perfect, and according to Plato, the circles of this world are always perfectly round. Concrete examples of circles that we find in the real world are always imperfect. If we examine an actual circle, at some level of magnification, the circle's edge will lose its curvature.

A possible way out of the mind–body problem is to rely solely on behaviors as indicators of mental states. Behaviors such as talking, running, and laughing are external and more easily measured than brain activity. According to **philosophical behaviorism**, mental states are dispositions or tendencies to behave in certain ways under certain circumstances. Happiness, for instance, can be thought of as the tendency to smile or laugh. There is support for this idea in that there are strong connections between mental states and behavior. Many people do smile when they are happy or cry when they are sad. However, we cannot conclusively prove any causal link between the two. A "tough guy" or stoic might not cry when sad for fear of being perceived as weak. Alternatively, somebody might exhibit nervous laughter before going on stage, in which case anxiety and not happiness is causing the laughter. For these reasons, we cannot consistently rely on behaviors as indicators of mental states. Please note that philosophical behaviorism is not the same as psychological behaviorism, discussed in Chapter 3. The psychological version focuses on the laws that govern the relationship between stimuli and responses and how we can shape responses by manipulating stimuli such as rewards and punishments. Psychological behaviorism ignores mental states altogether.

MONISM

If we are monists, we are left with two fundamental choices. Either the universe is mental or it is physical. It is difficult to take the first argument, called **idealism**, seriously—although it cannot be falsified. Imagine that everything you know isn't real but is simply an illusion of reality. The world you perceive and understand exists only "in your head"—although, in this universe, you don't have a real head, so we can say this only colloquially. A supreme being such as God could be responsible for this. Alternately, it could be that aliens have taken our brains and are feeding them information in such a way as to make us think that we live in the world we do. This is known informally as the "brain in a vat argument." In either case, we could never get outside our own conscious experience to prove what reality is really like.

On the flip side of our metaphysical coin is **physicalism**, or materialism. The origins of this view go back to the Greek philosopher Democritus (ca. 460–370 BCE), who believed that all things were composed of atoms. The attributes and behaviors of the atoms, he said, can explain the differences between things. Physicalists are also monistic and believe that the universe is composed of a single substance. However, they regard this substance as physical and material rather than spiritual or ethereal. Physicalism is,

thus, the doctrine that everything that exists is physical. The operations of the human mind are seen here simply as the operations of the human nervous system.

We can consider identity theory to be a form of physicalism. According to **identity theory**, the mind *is* the brain. Specifically, mental states are physical states of the brain. Just as salt is the same as its chemical structure of sodium chloride, in this perspective, we say that being happy is identical to a particular type of brain activity. In fact, for every type of what we think are mental states, there is a type of brain state. These physical brain states cannot be discovered through simple observation or an examination of the linguistic meaning of the terms. They can be understood only through proper scientific analysis.

Anyone who believes in identity theory will be left with more terminology than is needed. On the one hand, we have the common everyday subjective terms we all use to refer to mental states. These are words such as "feeling good" that we use to denote our subjective experiences. On the other hand, we have objective scientific phrases such as "activation of dopaminergic neurons in the meso-cortico-limbic system" to describe the brain states for these terms. Because the scientific descriptions are more accurate and better reflect what is actually going on, it might make sense to abandon our mental terms and rely solely on the scientific terminology. In effect, this is what **eliminativism** is about. Eliminativists reject not only the terminology but also the mental states they supposedly describe. According to eliminativism, once you have a good biological explanation for a psychological phenomenon, then you can eliminate the psychological explanation and rely solely on the biological one. Eliminativists don't believe in mental states at all. Only physical brain states are acknowledged as existing.

Theories of mind that use subjective terms such as those mentioned above and that use commonsense or intuitive reasoning are collectively referred to as **folk psychology**. Here is an example of a folk psychology theory: Cookies taste good. Robert likes to feel good. If Robert walks by the cookie store on his way home from school, he will buy and then eat some cookies. Folk psychology helps us understand and predict the behavior of people around us. In this sense, it is very useful. However, folk psychology is far removed from rigorous scientific explanations that are tested using experimental methods.

Churchland (1981) argues that folk psychology has outlived its usefulness and that we should eliminate it and replace it with neuroscience. He argues that folk psychology has not yielded any new insights into human behavior. It fails to inform us on a wide range of issues such as mental illness, creativity, and memory. Unlike valid scientific theories that tend to support one another, no folk psychology theories support one another or support scientific theories. Folk psychology, however, may still prove its worth in that it may serve as a basis for generating scientifically testable hypotheses. This would work only if the translation between the ill-defined subjective notions could be translated effectively into well-defined hypotheses that can be subjected to scientific testing.

Evaluating the Monist Perspective

Monist views have one clear advantage: They are simpler than dualist accounts. Rather than partitioning the universe into two types of things, they allow for only a single self-consistent universe. This follows the principle of Occam's razor, which states that all things being equal, the simplest explanation is usually the correct one. As mentioned

above, idealism does not really qualify as a legitimate theory because it cannot be proved true or false. As a result, it doesn't serve as a legitimate subject of scientific scrutiny. On the other hand, there is abundant evidence in favor of physicalism. Indeed, most of the experimental results in this book support the notion of the brain as being the core engine of mind. For example, neurological studies that look at behavior following brain damage show that when particular parts of the brain are damaged, relatively specific types of cognitive deficits can follow. This suggests that those brain areas are important physical mechanisms for those particular mental phenomena.

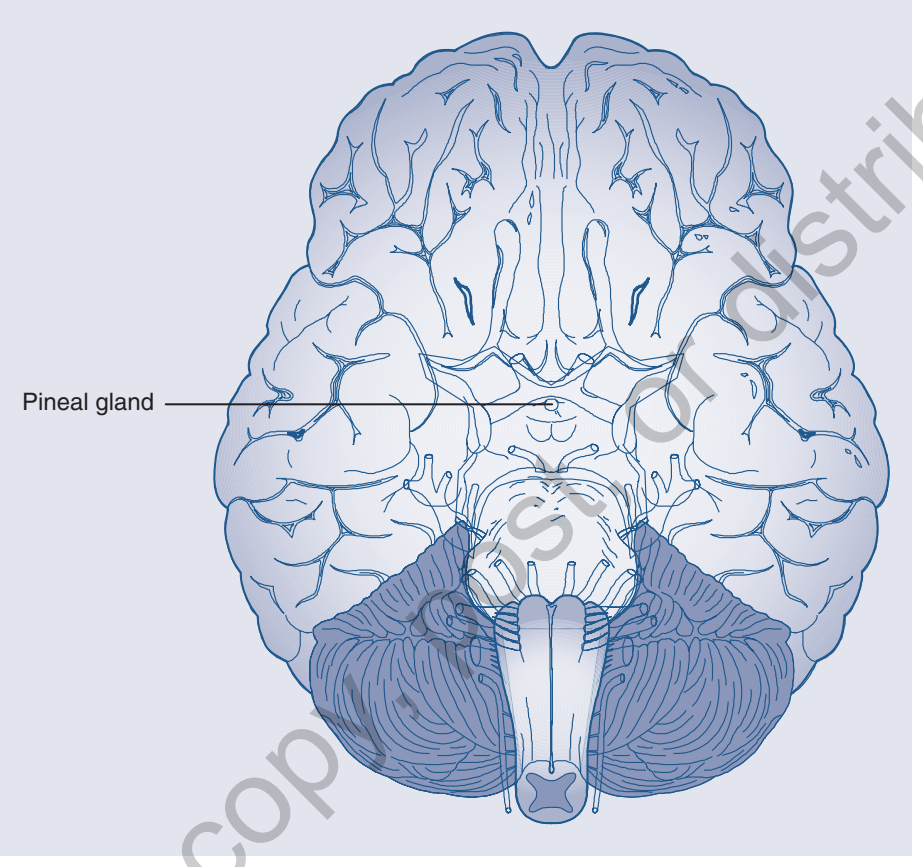
Physicalism, however, also has its share of critics. Some allow that physical processes can determine mental ones but deny that they can explain them. So, they might argue, changes in the brain may very well correlate with and produce fear, but they do not explain different kinds of fear, how a person becomes fearful, and so on. These critics acknowledge that the world is physical but indicate that there is no physical explanation for many phenomena. In these cases, they believe, it is perhaps better to explain using mental terms.

One argument that can be directed at identity theory goes by the name of **multiple realization**. Simply put, any given mental state, such as pain, can be instantiated or realized differently in different creatures. Pain in humans may be identical to activity in c-fibers, pain in squid may be identical to activity in d-fibers, while pain in aliens may be identical to activity in e-fibers (Ravenscroft, 2005). Furthermore, it is quite likely that differences exist not only between species but also between individuals within a given species. You and I may both feel pain when there is activity in c-fibers, but the particular pattern of activity in those fibers and in other areas of the brain could produce radically different mental perceptions of pain in each of us. Identity theory must, therefore, account for all the subtle mental differences that can exist between individuals.

DUALISM

Now that we have reviewed the various forms of monism, let us turn our attention to its logical alternative. Dualists believe both mental and physical realms are possible but differ in the way they think these two interact. **Classical dualism** originated with the French philosopher René Descartes (1596–1650). Descartes was a revolutionary philosopher for his time and introduced theories on many of the ideas that underlie cognitive science. He believed in a one-way causal link, with the nonphysical mind controlling the physical body but not vice versa. Descartes thought the mind exerted its control on the body through the pineal gland, perhaps because it is one of the few anatomical structures not duplicated on either side of the brain (see Figure 2.2). In this view, the mind is like a puppet master, the body is like a puppet, and the pineal gland is like the puppet strings by which the former controls the latter. Classical dualism conforms to most people's commonsense notion of the mind–brain relationship, which is that our thoughts control our actions. For instance, when we feel hungry, we get up and eat a snack. It certainly seems as though the feeling of hunger comes first and causes the action of getting up to eat.

Figure 2.2 Descartes believed the pineal gland to be the location where the mind influenced the body. This belief was itself influenced by the fact that the pineal gland is located in the center of the brain.



Substance Dualism

Descartes's views conform to what we now more commonly call **substance dualism**, where mind and body are composed of completely different substances. By *substance*, we mean the actual "stuff" that makes them up. In this view, there are two sorts of stuff: mental substances and physical substances. The physical substances can be the atoms, molecules, and cells of the natural world as we know it. It isn't clear what the basic elements or building blocks of mental substance might be.

There is a logical argument in favor of substance dualism that takes the form of several propositions. It goes as follows: Minds can do X. No physical object can do X. Therefore, minds are not physical objects. X can be any mental ability that has not been fully realized in some alternate physical form. Candidates include language, reason, and consciousness. The problem with this argument is that many of the abilities that we thought were exclusively human have now been replicated in machine form. Artificial intelligence (AI) is now capable of pattern recognition, language comprehension and production, and even artistic creativity (Friedenberg, 2008).

The arguments against substance dualism are stronger. One early attack came from Princess Elizabeth of Bohemia (1618–1680). She pointed out to Descartes that if mind and body are of two different substances, then they should not be able to causally interact. Saying they interact at the pineal gland specifies the location at which this might occur but not how. Descartes was unable to come up with a successful reply.

At least two other arguments against dualism involve explanatory power. First, biology can give us a very good account of mental states. If we can explain these states completely in terms of neurons, action potentials, and synapses, then there is no need to even provide a nonphysical account. Second, substance dualism does not give us any kind of alternate explanation. What are mental things made of? How do these things interact or operate? These questions are left unanswered.

Property Dualism

Another type of dualism may overcome these difficulties. According to **property dualism**, the mind and the body can be of the same stuff but have different properties. By *properties*, we mean characteristics or features. To illustrate, a golf ball and a tennis ball are both made of atoms, yet they differ in their properties. A golf ball is smaller, white, and dimpled. A tennis ball is larger, green, and fuzzy.

A property dualist believes that mental states are nonphysical properties of the brain. He or she would say that the brain has certain properties, such as being gray and wrinkled, weighing a certain number of pounds, and so on. But then he or she would also say that the brain gives rise to another set of properties, such as seeing red, being in pain, and feeling happy, that are wholly different from any of the currently understood properties of the brain.

Very little is gained in the dualist switch from substance to property. Nonphysicality still fails to provide an explanation for mental states, whether those are characterized as features or not. Again, we can ask, How do physical brain processes give rise to mental features? We are still left with better and more complete physical accounts of mentality. Therefore, in general, we can level the same arguments we have used against substance dualism also against property dualism.

Evaluating the Dualist Perspective

One critique of dualism comes from the philosopher Gilbert Ryle. Ryle's (1949) argument centers on our conceptualization of mind and its relation to the body. He believes that the mind is not any particular component of the brain but, rather, all the parts

working together as a coordinated, organized whole. He illustrates with a story: Imagine a visitor from a foreign country arriving at a large university. He is shown around the campus, and the various parts of the school are pointed out to him, including the dormitories, departments, and lawns. The visitor, who has never seen any of this before, is puzzled. He says, "Well, I've seen all this, but I haven't yet seen the university." We would have to explain to him that the university is not any of the individual sites he has viewed but all the sites together and the interconnections among them (see Figure 2.3). Ryle thinks philosophers fall into the same trap as the visitor, mistaking the part or parts for the whole. He argues that the mind belongs in a conceptual category different from that of the body, just as the university is in a category different from those of the things that make it up.

Andy Clark (2001) summarizes several other critiques of dualism. These would apply to Descartes's conception as well as other perspectives. Clark says that dualism is uninformative and tells us what the mind isn't rather than what it is. If the mind isn't the brain and isn't physical, then what is it? Dualists are remarkably silent on this matter,

Figure 2.3 Where is the university?

Source: Courtesy of Peter Finger Photography (2003).

often conceding that it is something nonphysical that we can't understand yet. As a theory, dualism also is inelegant because it postulates two worlds that must be coordinated. An explanation that does not violate the principle of Occam's razor would involve a single type of world, not requiring coordination.

There are further problems with dualism. One has to do with the dependence of the mental on the physical. Factors that affect the brain, such as head trauma or drug use, have direct and dramatic mental effects. We can see that damage to a certain part of the brain—say, from a motorcycle accident—results in somewhat specific forms of mental disruption—for example, language deficits. Taking a drug such as marijuana, which alters brain chemistry, results in altered mental states. In addition, the evolutionary approach shows us that there is an approximate correlation between intelligence and the number of neurons in the cerebral cortex across species, such that species with more cortical neurons tend to have greater cognitive capacity. It is obvious from these observations that the mental is integrated with the physical, that the mind depends on the brain.

Some dualists, in response to attacks on their positions, have stated that the mind exhibits extraordinary abilities and that it would be impossible for a physical system to duplicate such abilities. For instance, how can a physical system, be it a brain or a computer, write a novel or negotiate a peace treaty? The truth is that as our technological sophistication increases, many of these abilities are becoming better understood and implemented computationally. There are now computers that can beat the best chess champions and successfully diagnose medical disorders. These are capacities once thought to be the exclusive domain of humans.

Dualists and other philosophers also argue that our subjective experiences—things such as thoughts, beliefs, and desires—are not equivalent to physical brain states. They base this conclusion primarily on introspection. When we examine what is inside our heads, they say, these subjective experiences seem to be something more than just physical. The problem with this argument is that introspection is a weak form of evidence and can be wrong (as are many of our ideas). What is required is objective proof that such experiential states are not physical.

FUNCTIONALISM: ARE MINDS LIMITED TO BRAINS?

The most influential philosophical theory of mind in cognitive science is functionalism. For this reason, we will discuss it in considerably more detail than any of the theories already discussed. To get an idea of what functionalism is about, we need to make a distinction between two ways of classifying things. **Physical kinds** are identified by their material composition only. In this view, jellyfish and carpets are different because they are made up of fundamentally different physical substances. **Functional kinds**, however, are distinguished by their actions or tendencies. Here, we could say that all automobiles fall under the same functional category because they do the same things—namely, transport goods and people—even though they may be made up of different physical substances.

So far so good, but things get interesting when we extend these ways of classifying to the idea of mind. If we think of mind as a physical kind, then minds must be the same

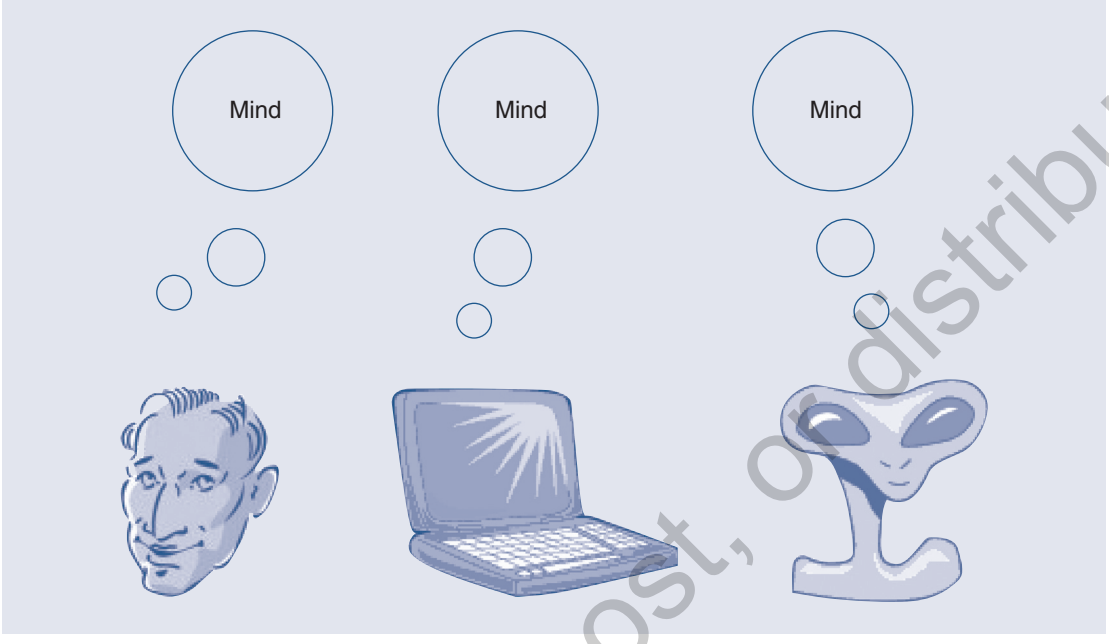
things as brains, since, as far as we know, minds cannot exist apart from physical brains. To many, this seems too exclusive. It is possible, they argue, that computers might develop minds and that there might be alien species with minds (see Figure 2.4). Neither computers nor aliens need have brains in the sense that we know them. It is more fruitful, they say, to identify minds as functional kinds and, thus, to define them by the sorts of processes they carry out rather than the stuff they're made of. According to **functionalism**, mental states are not just physical states but also the functioning or operation of those physical states. According to this view, a mind could conceivably be implemented in any physical system, artificial or natural, capable of supporting the appropriate computation.

Functionalism has several significant implications (Garfield, 1995). One is that the same mental state could be realized in quite different ways in two separate physical systems. This can be illustrated with computing devices. Two such different devices, say a desktop computer and a smartphone, can both compute the same result, such as displaying a page of text, but in different ways. The same might also be true for human computation. If we examined the brains of two people thinking exactly the same thought, we would in all likelihood not find exactly the same neural processes at work. This is the idea of multiple realization that we first introduced when discussing identity theory.

There are several schools of thought in functionalism, ranging from conservative views that advocate direct connections between physical and computational states to more liberal ones that emphasize computation over physicality. The liberal schools give two reasons for their stance. They say that for both computers and thinking organisms, the number of possible computational states always exceeds the number of possible physical states. Take, for example, all the different possible beliefs one could hold concerning politics, the environment, one's friends, and so on. Mathematically, the number of such beliefs could be practically infinite (Garfield, 1995). The number of possible physical states the brain can assume, though, is probably finite. According to that argument, the computational level of description thus becomes a richer and more diverse way of describing the mind and should be the preferred level. Second, liberal functionalists argue that psychological states such as beliefs are defined more by their relations to other such states, to inputs from the environment, and to behaviors than by their relations to physical states. A belief such as "patriotism" usually manifests itself in behaviors—for example, in flag waving. It will elicit predictable reactions to environmental stimuli—for example, feeling upset when one's country has been compromised—and will produce external behaviors such as marching or protesting. In functionalism, rather than just a brain state being what explains a patriotic state of mind, the full set of relations between the relevant thoughts and actions is what explains a patriotic state of mind.

To summarize, functionalism implies that mental states might not be reduced to any particular physical state. This argument does not quite require one to be a dualist. It is not saying that mental states don't conform to physical ones, only that there may be a wide variety of possible physical states capable of producing any given mental state. One last point of clarification needs to be made. As was the case with behaviorism, functionalism takes on two slightly different sorts of meaning within the philosophy of mind and psychology communities. Whereas the philosophical conception stresses multiple realizability, the psychological conception stresses the functions themselves. We elaborate more on this latter definition in the next chapter.

Figure 2.4 According to functionalism, different physical substrates can in principle all give rise to mind.



Evaluating the Functionalist Perspective

Although functionalism has been the dominant view in cognitive science since the 1970s, it is not without its deficiencies (Maloney, 1999). Remember that a tenet of functionalism is that minds that are not based on brains can exist. They can exist in objects such as computers as long as the physical substrates of those objects allow for the relevant computations. Critics have argued that, although it is possible that minds can exist in the absence of brains, this does not make it plausible. There is no current empirical evidence to justify this claim. We have yet to see something mental in the absence of a brain. Also, some have argued that the failure to identify mind with a physical kind can itself be considered a reason to do away with the concept of mind—rather than give it special status as a functional kind.

An additional problem with functionalism is that it cannot account for the felt or experienced character of mental states—a phenomenon known as **qualia** (quale, singular). Examples of qualia include the subjective experience of what it is like to feel hungry, to be angry, or to see the color red. It would seem that these kinds of experiences cannot be replicated as purely functional processes. A machine could be programmed to detect the color red, even mimicking the same human functional process, but this machine might not have the same conscious experience of what it is like to see red that a person has.

What is more, two individuals having the same conscious experience often do not experience it subjectively in the same way. A number of experiments have shown this to be the case with color perception. Participants looking at the same color will describe it differently (Chapanis, 1965). If asked to point out on a color spectrum what pure green looks like, one person may select a slightly yellowish green, another a slightly bluish green. This is the case even though the functional operations of their respective brains as they view the color are approximately equivalent. In this case, the neurophysiological operations behind color perception tend to be the same across most individuals.

THE KNOWLEDGE ACQUISITION PROBLEM: HOW DO WE ACQUIRE KNOWLEDGE?

A fundamental question asked by even the earliest of philosophers was, How do we acquire knowledge? Clearly, you are not born knowing everything—otherwise, you would not need to go to school and wouldn't be reading this book. But are we born knowing anything at all? Is the mind completely blank, or do we start with some rudimentary understanding of the world? One way to frame these questions is within the **nature–nurture debate**. This debate centers on the relative contributions of biology and experience in determining any particular capacity. The term *nature*, in this context, refers to traits that are genetically or biologically determined. These are coded for in our genes and so are "hardwired," meaning they are present at birth or appear at a prespecified time during development. The term *nurture* refers to traits that are learned through experience and interaction with the environment. We will examine theories of knowledge acquisition that argue for the greater influence of one or the other.

According to **nativism**, a significant body of knowledge is innate or "built into" an organism. In this sense, nativism is a theory of knowledge that favors nature over nurture. Plato was the first to outline a nativist theory of knowledge. He thought learning was a matter of recollecting what is already known—these concepts existing in the ideal world of forms and being part of our immortal souls. **Rationalism** must be subtly distinguished from nativism. Descartes was the progenitor of this perspective. Rationalists also tend to believe in the existence of innate ideas. These basic concepts include ideas such as "God" and "triangle." However, they additionally emphasize the existence of innate reasoning powers. These include certain logical propositions, such as knowing that something cannot both exist and *not* exist at the same time. We can use these a priori rational powers to form new ideas that are not given to us innately. Descartes would agree that we are not born with the idea of "table" but can acquire it given our innate ability to perceive, think about, and interact with objects. By applying formal logic to the propositions we assume are true, rationalism allows us to combine them to generate new knowledge that must also be true.

For example, let us assume the following premises are true:

1. If a person cheats on their taxes, that person is a criminal.
2. Dan cheats on his taxes.

Aristotelian logic allows us to generate a new conclusion for this syllogism and add to our list of premises-assumed-to-be-true the following conclusion:

3. Dan is a criminal.

Implied in this particular logical syllogism is a kind of causal relationship between tax cheating and being a criminal. But sometimes the directionality of the causal relationship can get confusing. Occasionally, human reasoners will simply treat Premise 1 as though it is a bidirectional correlation, such that not only are tax cheaters criminals, but criminals are also tax cheaters. In such instances, if Premise 2 were replaced with "Dan is a criminal," then a person using that flawed interpretation of Premise 1 might conclude that "Dan cheats on his taxes." This is called making the logical error of *confirming the antecedent*. Just because Dan is a criminal doesn't allow us to use Premise 1 to conclude that he is a tax cheater. While all tax cheaters are criminals (because cheating on one's taxes is a crime), some people fail to consider the counterfactual (alternative) possibility that some criminals may not cheat on their taxes—they just commit other crimes. Therefore, not all criminals are tax cheaters. The directionality of the causal relationship might be easier to identify in this next example.

Let us assume that Premises A and B are true:

A. If one smokes cigarettes, one increases one's likelihood of getting cancer.

B. Joe smokes cigarettes.

Like before, we can apply logic to generate a new conclusion:

C. Joe is increasing his likelihood of getting cancer.

The directionality of the causal relationship in this example is easier to see because we know that multiple things can cause cancer (not just cigarettes). So if Premise B were replaced with "Joe has cancer," most people would not falsely confirm the antecedent and jump to the conclusion that Joe is a smoker. (Although they might ask.) When reasoning about causality, it is important to both consider counterfactual circumstances (such as "If Joe is not a smoker, could he still get cancer?") and to avoid confirming the antecedent.

In contrast to nativism (and rationalism), **empiricism** sees knowledge as acquired through experience and observation: It favors nurture over nature. Rather than relying on knowledge that you might start out with (e.g., a list of premises-assumed-to-be-true) and then using rationalism to logically combine them and generate new knowledge, empiricism requires new knowledge to be acquired by testing your hypotheses in the world. For instance, if you think Joe in the example above might have cancer, then run some medical tests on him. Or if you think Dan might be a criminal, then test that

theory by collecting evidence and presenting it in a court of law. In this view, knowledge gets into the head through interaction with an environment, meaning it is learned. The senses provide the primary channels via which knowledge of the world is generated. Our knowledge of the concept "lemon" in this account begins with looking at a lemon, touching it, and tasting it. The British philosopher John Locke (1632–1704) is credited as the founder of the empiricist movement. He used the phrase *tabula rasa*, which literally translates as "blank slate" (essentially a chalkboard with nothing written on it). Locke believed that we are born as blank slates, lacking any knowledge, and that over time experience puts writing on the slate, filling it up.

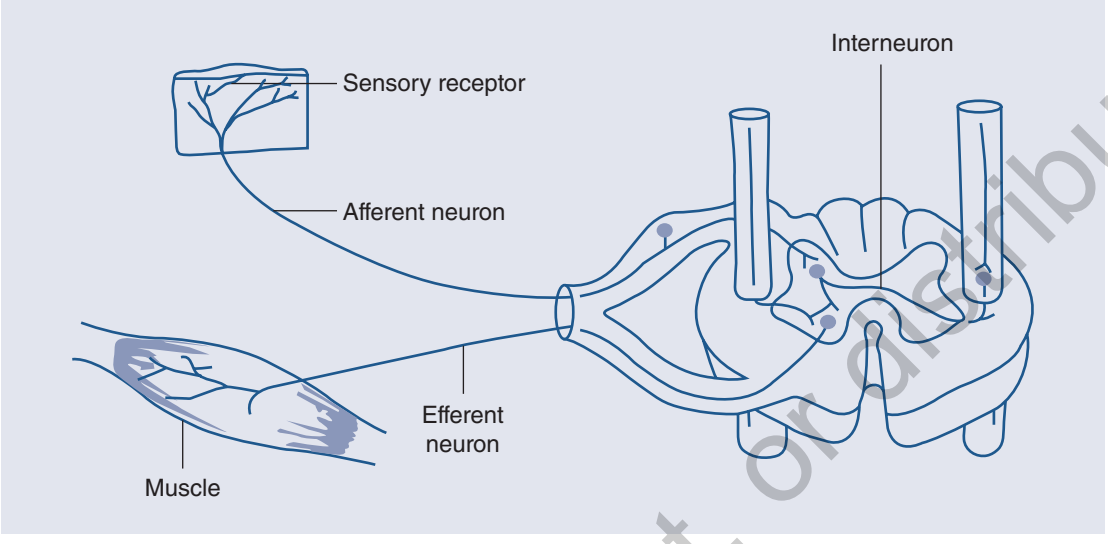
Locke had a more fully developed theory of learning. He differentiated between simple ideas and complex ideas. **Simple ideas** are derived through sensory input or simple processes of reflection. They are received passively by the mind and cannot be reduced to simpler ideas. Looking at a cherry would generate the simple idea of "red." Tasting a cherry would produce the simple idea of "sweet." **Complex ideas** are formed from the active mental combination of simple ideas. They are created through reflection only and can be reduced to parts, their component simple ideas. The idea of "cherry" would result from the associative combination of such simple ideas as "red," "sweet," and other commonly occurring sensations derived from one's experiencing cherries. This cluster of simple ideas is naturally associated because each time we experience a cherry, many of those simple ideas are generated. For this reason, Locke and others who have proposed similar notions are sometimes known as the associationists.

Evaluating the Knowledge Acquisition Debate

One might be tempted to immediately dismiss the doctrine of innate ideas put forth by the nativists and rationalists. After all, it seems absurd that we should be born knowing factual information such as the content of the Gettysburg Address. But the scope of knowledge is broader than this. Think back to the previous chapter, in which we defined declarative knowledge for facts and procedural knowledge for skills. There is quite a bit of research supporting the notion that some forms of procedural knowledge are innate. Newborn infants, for instance, come into this world with a variety of different skills. These skills are universal across the human species and manifest themselves so soon after birth that they couldn't possibly have been learned. Therefore, they qualify as examples of innate knowledge. Let us examine a few of these.

All healthy infants demonstrate a set of **reflexes**. These reflexes include the grasping reflex, in which the fingers tighten around a touch to the palm, and the rooting reflex, in which the infant turns his or her head and begins sucking an object placed near the mouth. Reflexes serve a clear adaptive function. Grasping and sucking, along with behaviors generated by other early reflexes, are important for survival. The physiology behind reflexes is simple and fairly well understood. A stimulus triggers one or more sensory neurons that then activate intermediary neurons. These in turn activate motor neurons, causing the resulting behavior. It is easy to see how such a simple mechanism could be hardwired at birth to enable the infant to respond effectively to its environment. Figure 2.5 shows the anatomy of a spinal reflex.

Figure 2.5 The neural connections in a spinal reflex. A reflex is an example of innate procedural knowledge.



Smell preference is another example of innate behavior. Steiner (1979) found that newborns tend to agree with adults in terms of which odors they consider pleasant or unpleasant. He found that odors such as strawberry and banana elicited agreeable facial expressions from young infants—for example, smiling. Unpleasant odors, such as rotten eggs, elicited expressions of disgust. As is the case with reflexes, these smell preferences have survival value. Babies who find the smell of fruit attractive will eat the fruit and thereby gain nutrients; those who are repulsed by spoiled food will reject the food and avoid getting sick. The neural mechanisms behind such preferences are probably not too complex either. They need to involve little more than a mapping between the odor and the emotional response.

Given the above examples, we see that it is not so far-fetched for us to be born with some basic procedural knowledge. This knowledge is in the form of simple neural circuits that map stimulus inputs to appropriate behavioral outputs. This knowledge can even be represented using the conditional rules we talked about in Chapter 1. A computational coding of a smell preference might look something like this: "If *smell is rotten eggs*, then *disgust*." The odor, if it satisfies the first part of the conditional, would then trigger the response in the second part.

But how did these circuits get there in the first place? The early nativists and rationalists either did not specify the source of innate knowledge or attributed it to God. Evolutionary psychology offers us another explanation. It attributes such capacities to generations of selection pressures acting on a species. Basically, over many generations,

individuals who happened to not be born with those disgust-based procedural rules found themselves eating things that led to them dying before they could contribute their DNA to the next generation. Those who were born with those rules succeeded in contributing their DNA to the next generation. The result is that eventually the entire species is generally born with those procedural rules innately programmed into their nervous systems. These pressures can promote the development of adaptive (survival-related) cognitive abilities. Evolutionary psychologists can be considered modern-day nativists. See the evolutionary approach chapter for more on their views.

The phrasing of the nature–nurture debate as allowing for only a single alternative, either one or the other, is misleading. Although some traits may indeed be mostly the product of nature or of nurture, there is a large middle ground consisting of traits or cognitive abilities that can result from the complex interaction of the two. In these cases, nature may set constraints or limits on environmental influence. Take memory, for example. Tsien, Huerta, and Tonegawa (1996) engineered a mutation in a gene that affects a particular type of receptor in the hippocampus, a brain area responsible for the learning of new information. Rats with the mutation did poorly in a memory task as compared with normal rats in a control group. Tang et al. (1999) did something even more remarkable. Through genetic manipulation, they increased production of a particular subunit in the hippocampal receptor. This change increased the receptor's effectiveness. Rats with this "enhanced" version of the gene outperformed rats with normal receptors on a spatial memory test.

This research is exciting because it shows that memory in these animals is at least partially under genetic control. However, it is also well documented that human memory capability can be improved through training and the use of memory strategies (Roediger, 1980). The way in which these genetic and environmental factors interact to determine memory in any given individual is sure to be complex. A person with an innately poor memory could probably improve their memory skills to better than average by using a variety of memory strategies. And a person with an innately strong memory could make their memory skills look almost superhuman by using those memory strategies.

THE MYSTERY OF CONSCIOUSNESS: WHAT IS CONSCIOUSNESS AND HOW DOES IT OPERATE?

Consciousness is a complex concept and has no single agreed-on definition. In its broadest sense, we can think of it as the subjective quality of experience (Chalmers, 1996). It may be thought of as our individual subjective awareness of mental states. These states include sensation, perception, visual images, conscious thought processes, emotions, and sense of self, just to name a few. But these states assume that a person is in a normal, awake, and alert frame of mind. The issue becomes more complex when we think of other types of consciousness—for example, being unconscious, asleep, in a drug-induced state, hypnotized, or meditating. There are clinical cases representing other states of consciousness as well. In dissociative identity disorder, a person can alternate between separate personalities. Each personality can possess unique skills and may or may not be aware of the others. In split-brain patients, one half of the brain can

possess an awareness of an object that the other half does not possess. For simplicity, we do not consider these alternate states of mind.

An interesting aspect of consciousness is whether it is unitary or divided. Subjectively, our consciousness seems to be unitary. That is, one recognizes himself or herself to be one person, experiencing things in the present moment. When one studies the brain, though, one finds that there is no single place or even time where consciousness seems to happen. Instead, the brain in action is a case of activity going on all over the place. Furthermore, the brain may even be processing different aspects of a single experience at different times. How can we reconcile this objective evidence with our subjective experience? See the "Interdisciplinary Crossroads" section for one theory on this apparent contradiction.

Chalmers (1996) makes a distinction between phenomenal and psychological concepts of mind. The **phenomenal concept of mind** is essentially the idea of mind as a conscious experience. Mental states in this view need to be explained in terms of how they feel. The **psychological concept of mind** sees mental states only in terms of how they cause and explain behavior. Here, mind is characterized by what it does—how it feels is irrelevant. Philosophers have concerned themselves primarily with the former, psychologists and cognitive scientists with the latter. To make this distinction clear, imagine biting into a candy bar. A phenomenal investigation would attempt to explain why you experience the mental states of "sweetness" or "chocolate" and why you might perceive them differently than somebody else does. A psychological investigation would concern itself with the neural circuits that become activated during the taste, how they might be represented computationally, and how this explains when you might stop eating. In this section, we concern ourselves with the phenomenal concept of mind and its relation to consciousness, since the psychological view is in most cases the topic of the remainder of this book.

Chalmers (1996) also differentiates between what he calls the easy and hard problems of consciousness. **Easy problems of consciousness** are those that can be solved by cognitive science and explained in terms of computational or neural mechanisms. His examples include the ability to discriminate, categorize, and react to environmental stimuli; the focus of attention; and the difference between wakefulness and sleep. Obviously, these correspond to the psychological concept of mind. The **hard problem of consciousness** involves subjective experience. Here we would need to explain why we have a visual or auditory experience when we look at or listen to something. In this case, we are now talking about the phenomenal concept of mind. Whereas science can give us answers to the easy problems, it may not be possible to provide answers to the hard problem of consciousness. The fact that subjective human experience may not be fully explained by an objective account using physical and mechanical processes is known as the **explanatory gap** (Levine, 1983).

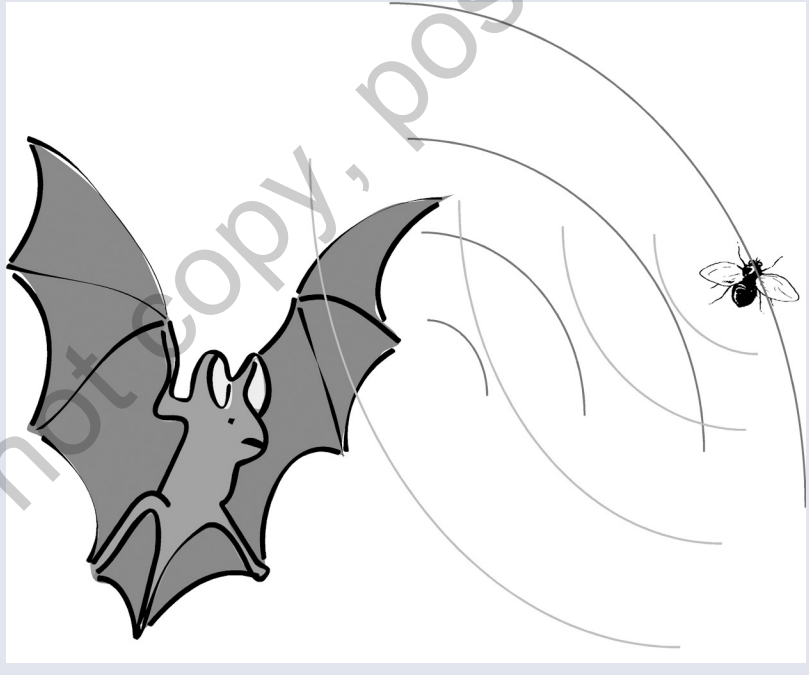
The What-It's-Like Argument

Nagel (1974) says that there is "something that it is like" to have a conscious mental state. When you bite into a candy bar, you have a subjective conscious experience of

tasting it. The candy bar, of course, has no such experience. There is nothing that "it is like" for the candy bar being bitten. This is one way of describing consciousness—that organisms that possess it can be described as having some sort of experience. Things incapable of supporting consciousness cannot.

But what is this experience like? Nagel (1974) asks us to imagine what it must be like for a bat to navigate by echolocation. In echolocation, the bat emits high-pitched sounds. The sound waves bounce off an object in the animal's path, and the animal uses the reflection time as a measure of the object's distance (see Figure 2.6). Neuroscientists have developed an in-depth understanding of the neuropsychological mechanisms that allow a bat to perceive its environment this way (Ulanovsky & Moss, 2008). And engineers have built adaptive sonar machines that can compute echolocation in a way similar to how a bat does. But none of this tells us what it is like for the bat to experience the world in the way it does. We have seen this argument before in our evaluation of functionalism. There, we said that a functional description of a cognitive process does not account for the qualia, or subjective experience, of the process.

Figure 2.6 What is it like to be a bat?



The problem here is that science can provide only an objective account of a phenomenon, and consciousness is an inherently subjective state. As organisms capable of supporting consciousness, we can introspect and analyze what it is like to have or to experience a mental state. Unfortunately, for cognitive science, this is not what is needed. Cognitive science must, instead, have a scientific and objective account of what consciousness is. Frank Jackson (1982) aptly illustrates the difference between objective and subjective accounts of a conscious experience. He asks us to think about a neuroscientist named Mary who is well trained in the physical mechanisms underlying color vision. She understands everything there is to know about how the eye and brain process color information. Mary, however, is colorblind. Imagine now that we take away her colorblindness and allow her to look at a tomato. Interesting questions arise. Does Mary learn anything new by this experience? Does the scientific community gain anything by Mary's (or anybody else's) description of what it is like to see red? Jackson argues that we do gain something and that science needs to explain this new information.

In contrast to this position, some state that subjective knowledge is not factual knowledge at all and, therefore, does not constitute any kind of an explanation. Personally knowing what it is like to taste a candy bar or to see red is not the same thing as objectively and factually knowing it. Adopting this position, we as investigators would be forced to ignore introspection and any other form of subjective description. Our focus would be on only legitimate objective techniques for studying the mind, such as observation, experimentation, brain imaging, and computer simulations.

Mind as an Emergent Property

Consciousness is a "hot" topic in contemporary cognitive science. In the past 15 or so years, there have been renewed interdisciplinary efforts to understand it. A number of well-known authors have published books for academic and layperson audiences that outline their definitions and views on the subject. These authors' theories are too numerous to mention here. We instead describe one popular theory in this section and a more detailed description of another in the "Interdisciplinary Crossroads" part of the chapter. Before we begin this, however, we must articulate two different conceptions of how complex systems such as the brain might work.

Reductionism is the belief that everything about a whole can be explained entirely by understanding its parts. If this were true, we could say that Phenomenon X is "nothing more than" Phenomena Y and Z combined. If it was a simple additive combination of X = Y + Z, then the reduction would be easy, and talk of X could be eliminated because Y + Z tells us all we need to know. If we were to let X be understanding the mind, Y be the connectivity among neurons, and Z be the activation of those neurons, then X = Y + Z would mean that we could completely understand the mind by a linear combination of neural connectivity and neural activation. Would that it were so easy. Few, if any, natural phenomena actually lend themselves to a linear additive reduction like Y + Z. Instead, the subcomponents that combine to generate some coherent event, like Phenomenon X, are usually combined in rather nonlinear (nonadditive) ways that involve processes akin to multiplication, division, and exponents. There are very few cases of successful linear reduction in the sciences. Thermodynamics has not been fully

reduced to statistical mechanics, nor has chemistry been fully reduced to quantum mechanics. In part, this is because it is not enough to understand the parts; one must also understand how they interact or relate to one another.

In **emergence**, the features of a whole are not completely dependent on an additive combination of the parts that make it up. Some of those features emerge over and above what the additive combination would predict. Holland (1998) lists several characteristics of emergence. Emergent systems are made up of interacting parts. The function of these systems is rule governed. The rules stay the same even though the parts change over time. And the emergent system's output often contributes to its next input. This feedback loop gives rise to novelty and unpredictability. It is difficult or impossible to anticipate what such systems will do, even if some or all of the rules that govern their operation are known. A problem with the emergent account is that it doesn't constitute an explanation. All it says is that the behavior of the whole is more than the sum of its parts. What is needed is a causal or scientific explanation of how part interactions give rise to emergent features.

John Searle (1992) uses the concept of emergence in his book *The Rediscovery of the Mind*. He argues that consciousness is an **emergent property** of the brain. An emergent property of a system, as we have mentioned, is realized through the interaction of the system's parts. Searle says if we have a given emergent system, S, made up of elements a, b, c, and so on, then the features of S may not be the same as the features of a, b, c, and so on. This is because the features of S arise from the causal interactions of the parts. Water, for example, has the features or properties of liquidity and transparency. The H_2O molecules that make it up do not share these properties. The causal interactions of these molecules give rise to these properties. In the same way, Searle says, consciousness is a property of the brain but not of its parts. If we take neurons to be the relative parts, then they have their own properties, such as being able to communicate with one another via electrochemical signals. These properties that are inherent in the way the neurons interact give rise to consciousness, but the properties of individual neurons need not be those of a conscious mind.

Searle is very careful to point out that he is not a reductionist. He does not believe consciousness is reducible to its parts. In reductionism, explanation goes downward and a phenomenon is directly explainable in terms of what is happening at a smaller scale. In emergence, explanation goes upward. The large-scale phenomena are more than just what is happening in and around the small-scale parts and cannot be explained solely by an account of what the parts are doing. This idea is similar to the concept of a gestalt in perception. Gestalts are discussed in the next chapter ("The Psychological Approach").

Searle seeks to avoid the monism–dualism dichotomy of the mind–body problem. He does this by talking about consciousness as a property rather than a substance. He likens consciousness to an emergent characteristic of what brains do in the same way that digestion is what stomachs do or photosynthesis is what plants do. He sees consciousness as a natural process and a by-product of the brain's nature. However, he does classify conscious mental states as separate from physical ones. He states that they constitute a unique and novel category of phenomena, with an independent reality and a distinct metaphysical status.

EVALUATING THE EMERGENT VIEW OF MIND

As appealing as this formulation is, it still leaves us with some vexing questions. The reformulation of consciousness as a property, and a nonphysical one at that, still begs the question: What is a property? If a property is not physical, then of what substance is it? Although attempting to avoid the mind–body debate, Searle seemingly ends up as a property dualist. Restating consciousness as a nonmaterial property of a material brain doesn't get us any further toward understanding what this type of property is. Also, it is not clear how emergence happens—that is, we do not yet have an understanding of the relationship between microscopic and macroscopic properties. In the case of water, we can say its large-scale properties have something to do with the three-dimensional shape of the H_2O molecules and other conditions, such as the surrounding temperature. For consciousness and the brain, this relationship between the microscopic and the macroscopic is far more ambiguous.

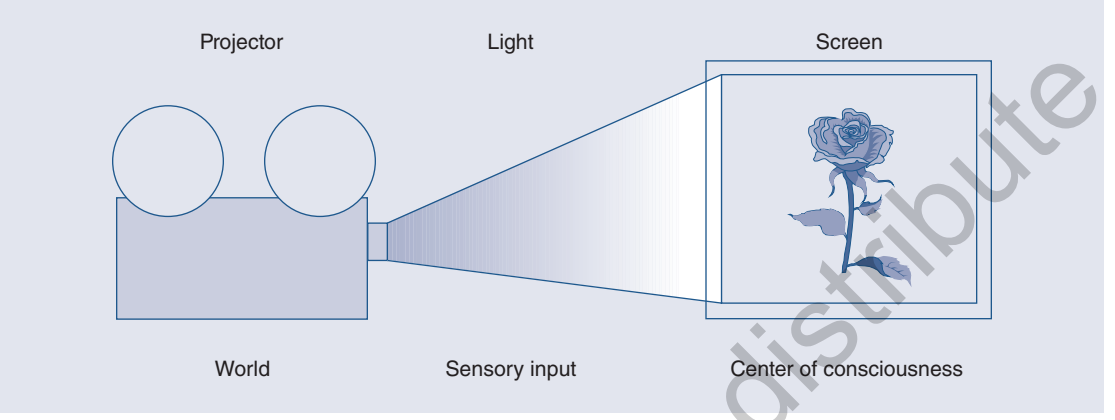
Searle's reason for believing in a nonmaterial consciousness is based on his conception of the difference between physical and mental things. For physical things, we can make a distinction between appearance and reality. A piece of wood may subjectively appear a certain way to us—as brown, as having a certain length and weight, and so on. These characteristics can also be measured objectively; we can put the wood on a scale to determine its weight, use a ruler to determine its length, and apply a wavelength detector to measure its color. For mental things, this distinction between the subjective and the objective goes away. Regarding mental experience, Searle believes that appearance is the same as reality and that our subjective introspections are objectively correct. But if this were true, we would have to trust our intuitions about the mental world as metaphysically special and nonmaterial.

In opposition, Paul Churchland (1995) points out that this reliance on the infallibility of introspection is an outdated notion. He notes that introspection often does not give us direct and accurate knowledge of the mental. Our inner assessments of mental states can be quite often and, notoriously, wrong. It is commonplace for us to err in judging our thoughts, feelings, and desires. Many of the early psychologists relied on introspection as a means to study the mind. The next chapter provides a more elaborate discussion of the problems they encountered.

Consciousness: One or Many?

In his book *Consciousness Explained*, Dennett (1991) outlines an interesting theory on the nature of consciousness. He begins by refuting the classical view of consciousness. The classical view, promoted by Descartes, posits a single place in the brain where all information funnels in. This area is a supposed center of consciousness, where we experience the world or the contents of our thoughts in a coherent, unified way. Dennett calls this center the "Cartesian theater." It is as though our consciousness is the result of a projector displaying information on a movie screen. The individual sitting in the theater watching the screen then has a single conscious experience of what is playing. Figure 2.7 gives a representation of the Cartesian theater.

Figure 2.7 The Cartesian theater explanation of consciousness.



There are a number of problems with the Cartesian theater. To start, linked modes of information do not arrive within the brain simultaneously. Light from an event precedes the arrival of sound. The sight of a fireworks display reaches the mind prior to the sound of the explosion, yet we experience the two in unison. This suggests that our consciousness is constructed; the visual experience is kept in check or delayed until arrival of the sound, at which point the two are integrated into a unified percept of the fireworks. This example and others imply that consciousness does not occur in real time—but (in many instances) several fractions of a second or so after an event. Our experience of consciousness as direct and immediate seems to be an illusion.

Another problem with the Cartesian theater is that anatomically, it is difficult to find a solitary brain region that, all by itself, links incoming sensory inputs and outgoing motor outputs. There is no **central processing unit (CPU)** in the brain as there is in a computer. The task of a computer's CPU is to schedule and coordinate ongoing activity. Furthermore, the Cartesian theater analogy requires an observer in the audience watching the screen. This observer is the subjective self who experiences the screen's contents. But how is this person inside one's head interpreting the image and having the conscious experience? To explain this, we would need to posit another mechanism or theater inside this person's head with another even smaller person and so on, ad infinitum. This is known as the homunculus problem in psychology and philosophy. **Homunculus** translated means "little man." An effective theory of consciousness must avoid the logical conundrum of homunculi nested inside each other.

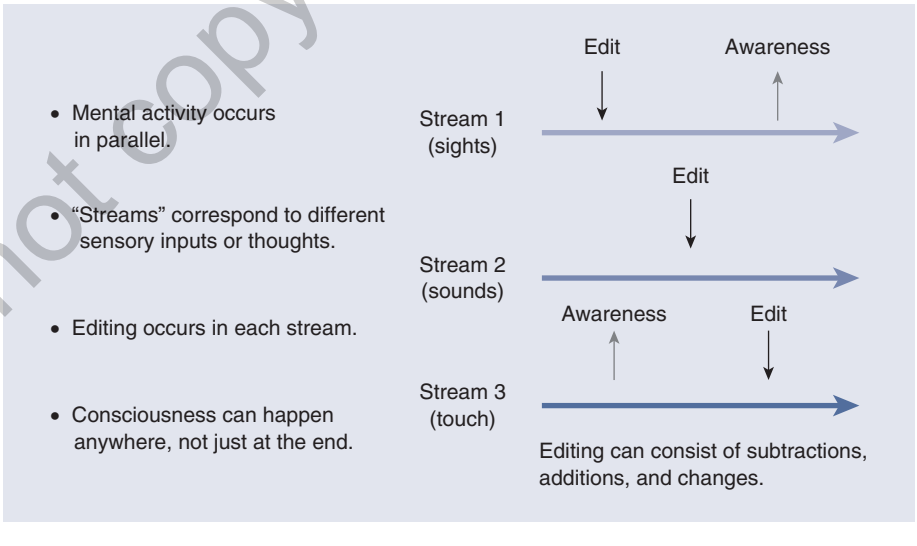
Dennett replaces this problematic formulation with a multiple-drafts model of consciousness (see Figure 2.8). In this model, mental activity occurs in parallel. Rather than projecting to a single location for processing in unison, different ongoing streams of information are processed at different times. Each of these streams can correspond

to different sensory inputs or thoughts. Processing or editing of the streams can occur, which may change their content. Editing can consist of subtractions, additions, and changes to the information. Awareness of a stream's content can happen before or after editing takes place. To illustrate, take our fireworks example. One mental stream would contain the visual experience of the fireworks, while another would contain its auditory representation. The visual stream would undergo editing in the form of a delay to synchronize it with the auditory stream. Then the information from both streams could be tapped to produce awareness.

There is abundant evidence in support of the multiple-drafts model. Take, for instance, the organization of the visual system. It adopts a "divide-and-conquer" strategy. The visual system carves up different aspects of an object during pattern recognition. These aspects are each processed separately in different parts of the brain by anatomically distinct pathways. This information is later combined to yield a unitary percept, but we are not aware that the information has been separated and then united.

A famous experiment by Loftus and Palmer (1974) also provides support for Dennett's idea. In this study, participants viewed a film of a car crash. Afterward, they were asked to estimate the speeds of the cars. The crucial manipulation was in how the question was asked. Some were asked how fast the cars were going when they "bumped" into each other. Others were asked how fast they were going when they "smashed" into each other. As you might predict, those who were queried with a mild descriptor such as "bumped" estimated that the cars were moving more slowly. Those queried with a more severe descriptor such as "smashed" estimated the speeds as considerably higher. In fact, even simply asking "what was happening?" in the event—instead of asking "what happened?"—can elicit more violent descriptions of the car crash. Essentially, the past

Figure 2.8 Dennett's multiple-drafts model of consciousness.



progressive verb form places your memory's perspective in the middle of the event, whereas the past perfect verb form places your memory's perspective outside of the event (Matlock, Sparks, Matthews, Hunter, & Huette, 2012). These results suggest that our memories of an event are not perfectly preserved "snapshots" of what happened but are actively edited over time. Subtle differences in the posing of a question, and other subsequent experiences after the event, can cause the memory to be edited and changed.

Dennett's theory also allows for different levels of awareness. Some information that is part of a stream may be available to conscious awareness and could be verbally described by the individual experiencing it. Other data streams we may be only vaguely aware of, but they can persist and influence additional mental processes. Yet other information may simply fade into the background. We may never be aware of this information. These three levels of awareness are comparable to Freud's conscious, preconscious, and subconscious aspects of mind, discussed in the next chapter.

In summary, Dennett's theory is more logically coherent and captures some of the empirical evidence on conscious experience. It suggests that there is no central place where consciousness happens but that multiple mental events occur in parallel. These events may be edited and changed in such a way that consciousness need not take place in real time. We may or may not be aware of these events.

Consciousness and Neuroscience

What does the brain have to do with consciousness? Is there some part of the brain or some particular pattern of neural activity that gives rise to consciousness? What is the neural correlate of conscious experience? Although philosophers have been debating the relation between the brain and mental phenomena for millennia, recent advances in neuroscience have yielded more specific insights into these questions. Let's examine some of them here.

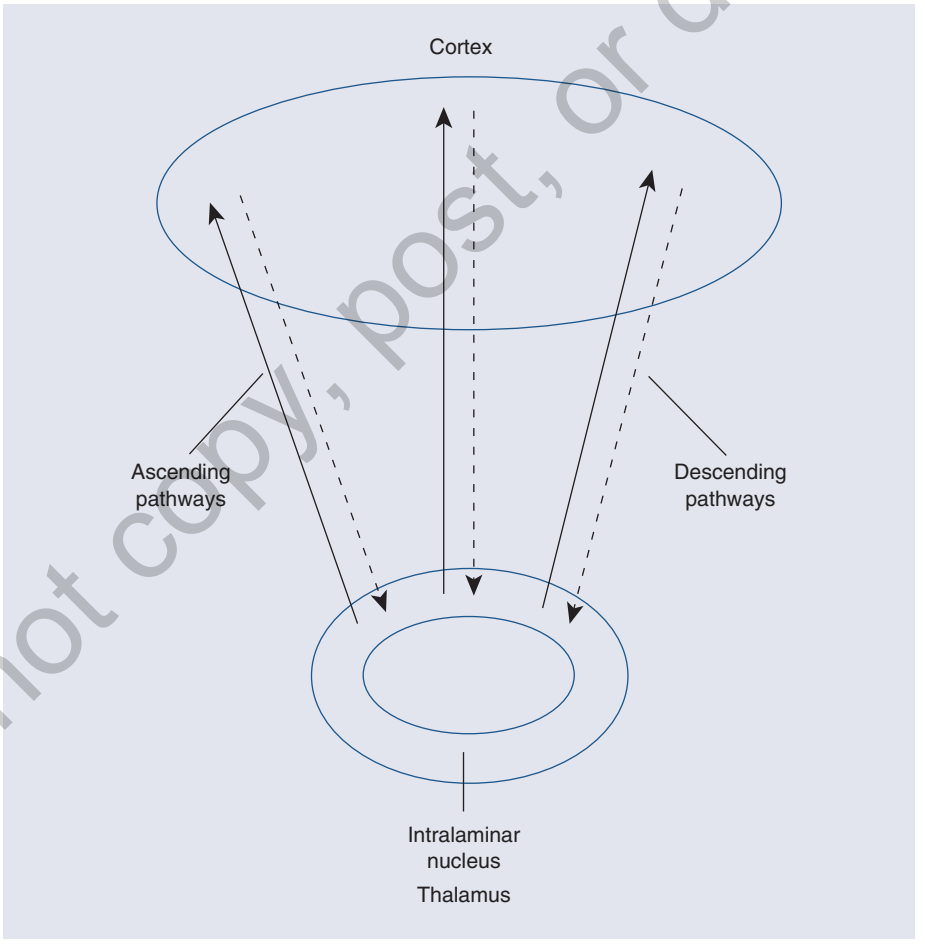
In general, the neuroscience view is that consciousness results from the coordinated activity of a population of neurons in the brain. Popper and Eccles (1981) see consciousness as an emergent property of a large number of interacting neurons. A different idea is that there are neurons specifically devoted to producing consciousness. Crick and Koch (1995) believe that these are located throughout the cortex and in other areas associated with the cortex. Activity in at least some subset of these neurons produces conscious experience. They believe that these neurons are special and that they differ from other neurons in terms of their structure and function. A similar but slightly different conception is that any cortical neuron may contribute to a conscious experience; however, different groups of cortical neurons mediate different types of conscious experience.

If there were special consciousness neurons, where might they be located? It has been proposed that one area is the intralaminar nuclei of the thalamus (Purpura & Schiff, 1997). The thalamus is a relay center for incoming sensory information. It sends information from each of the different sensory modalities, such as vision, audition, touch, and taste, to specialized areas of the cortex devoted to processing the information. Lesions of or damage to this brain region results in coma and loss of consciousness. It may be that these thalamic neurons, because they have widespread projections to many cortical areas, serve to activate or arouse other cortical neurons. Specific activity in different cortical

regions may then account for specific forms of consciousness. For example, activation or arousal in the occipital visual regions may correspond to visual awareness, while activation of the somatosensory cortex may produce awareness of different parts of the body.

Churchland (1995) formulates a neurocomputational theory of consciousness that focuses on connections between the intralaminar nuclei of the thalamus and disparate cortical areas. The circuit consists of ascending projections from the thalamus to the cortex, as well as descending pathways from the cortex to the thalamus. Figure 2.9 shows the anatomical layout of this area. These pathways are recurrent, meaning that a signal can be sent back and forth inside it. In this case, information coming into the thalamus

Figure 2.9 The consciousness circuit proposed by Paul Churchland. Activity in these pathways may give rise to conscious experience.



can be passed to the cortex, while the cortex can also pass information back to the thalamus. Recurrence is an important network property because it allows for feedback and learning. Recurrent activity in a network may sustain information over time and be the basis for conscious mental awareness. Recurrent artificial neural networks and their properties are described in Chapter 7 ("The Network Approach").

Churchland believes that the characteristics of this network can account for a number of different features of consciousness. One such feature is the capacity of consciousness to hold information over time—the equivalent of a short-term memory in which we are aware of the current moment in relation to the past. This is in keeping with the network's recurrent nature, since information can be held over time as it is cycled back and forth. Churchland also shows that this network can maintain activity in the absence of sensory inputs—for example, when we are daydreaming or thinking with our eyes shut. It can additionally explain why we lose consciousness during sleep, why it reappears during dreaming, and a host of other such features.

Churchland is quick to acknowledge, however, that it is the dynamical properties of this recurrent network and not its particular neural locus that make consciousness possible. He admits that a consciousness circuit may exist in places that have been suggested by other researchers. One such area is within the right parietal lobe (Damasio, 1994). The neuroscientist Rodolfo Llinas (2002) has suggested that consciousness may arise within the layers of the primary sensory cortex itself. He has written a comprehensive book that lays out his views.

Koch (2004) has also adopted a neurobiological approach to consciousness. He defines the **neural correlates of consciousness (NCCs)** as the minimal set of neural events and structures sufficient for a specific conscious experience. Since the entire brain appears to be sufficient for consciousness, the term *minimal* in the above definition is used because we would like to know what specific brain areas or activity is necessary. There should be as many distinct NCCs as there are different types of conscious experience. For instance, one pattern of activity may underlie the taste of chocolate, another the feel of wind on your face, and yet another the feeling of jealousy. If we change this pattern of activity, there should be a corresponding change in the subjective experience, and if we suppress it, the associated experience should disappear. Furthermore, it should be possible to induce a particular experience artificially by creating an NCC through an electrode or some other form of stimulation.

INTERDISCIPLINARY CROSSROADS: PHILOSOPHY, NEUROSCIENCE, AND BINOCULAR RIVALRY

Imagine an experiment where you are asked to put on glasses with one red lens and one blue lens. On a screen in front of you, there is a pattern consisting of vertical red bars crossed by

(Continued)

(Continued)

horizontal blue bars. The glasses present a different image to each eye: One sees the red stripes; the other the blue stripes. Under normal circumstances, the information from both eyes is fused to form a single unitary sense of the world with depth. Under these new circumstances, though, the two eyes compete for conscious awareness. You might first see the horizontal blue bars for a few seconds; then there would be a shift, and for the next few seconds, you would perceive the vertical red bars. This phenomenon, where each perception alternately vies for visual consciousness, is called **binocular rivalry** (Alais & Blake, 2005).

Now in this study, we ask you to push two buttons, each corresponding to one of the possible percepts. You are instructed to push the left button when you see the vertical red bars and the right button for the competing perception. While this is happening, the researchers are measuring your brain activity using an imaging technique known as magnetoencephalography. In this procedure, a helmet containing a number of sensors that detect very small magnetic fields induced by neural activity is placed over your head. This allows the investigators to see which parts of your brain are active during the experience. In this way, they can determine the neural correlates of these conscious experiences.

The results from this work show a particular pattern of neural action when participants reported not being consciously aware of the images. This pattern is characterized by widespread action in the visual areas of the occipital lobes as well as in the frontal lobes that underlie more high-level cognitive processing. This pattern, however, changed dramatically once an observer reported seeing one of the two possible sets of bars. Awareness of the pattern produced an increase of 40% to 80% neural responding. There was much reentrant activity in this state, with looping feedback found between different brain regions. But perhaps what is most interesting is that in these studies, no two observers had the same pattern, even when they reported seeing the same thing. These sorts of individual differences pose a problem for cognitive science because they imply that the brain does not always rely on the same areas to produce a given function. Fortunately, brain imaging does show some localization of function across individuals, as well as variability.

In another similar study, participants were presented a picture of a house to one eye and a picture of a face to the other eye (Tong, Nakayama, Vaughn, & Kanwisher, 1998). When the observers reported seeing a house, functional magnetic resonance imaging showed bilateral activation in the parahippocampal place area (PPA) and less activity in the fusiform face area (FFA). When they perceived the face, this pattern of activity was reversed, with greater activation in the FFA and less in the PPA. The FFA, as we will describe in more detail in the neuroscience chapter, is that portion of the temporal lobe that seems to respond selectively to facial stimuli.

Experiments such as these can tell us what parts of the brain or what patterns of brain activity underlie conscious experience. Ultimately, the use of such imaging techniques may be able to give us a complete picture of brain action for any given conscious experience in any individual. This will certainly tell us a lot about the brain, but it falls short of explaining the subjective nature of conscious experience. That is because, in the end, we will be left with an objective description of an inherently subjective phenomenon. In fact, that is all that science as an objective method can provide. At that point, we will either need to be content with our physicalist account of consciousness at the neural level or continue our investigation further by looking for more physical correlates at different levels, such as the synaptic or molecular.

Consciousness and Artificial Intelligence

Researchers in AI design algorithms to perform real-world computational tasks such as language comprehension and problem solving. Many of these algorithms can be judged as successful from a behavioral standpoint because they adequately perform their tasks, some under a variety of different conditions. If we define thought as computation in some physical substrate, as functionalists do, then we can also, without much risk, say that the execution of these programs is equivalent to "thinking." But does this correspond to consciousness? This is a much riskier proposition, since, as we have seen, consciousness can imply more than computation. It seems to involve subjective experience and perhaps other things. In this section, we address the question of whether a machine can be conscious. This is perhaps the most interesting philosophical issue in AI today.

There are a variety of different perspectives on whether or not a machine can become conscious (Freedman, 1994). These may be generally classified into two categories. The **strong AI** view asserts that consciousness can arise from a purely physical nonbiological process. Followers of this perspective believe that, eventually, as we create machines with greater complexity and computational power, we will see consciousness emerge in them. Proponents of **weak AI** claim that consciousness is itself either not a physical process, and so can never be reproduced, or a physical process that is so complex that we will never be able to duplicate it artificially.

Let us examine the arguments both for and against strong AI. Daniel Dennett (1998) raises several points in its defense. He mentions that many phenomena that used to have mystical and supernatural explanations now have scientific ones. Consciousness should be no different, he argues. Some have claimed that consciousness may be possible only in an organic brain. Dennett concedes that this may be true but notes that science has already been able to mechanically reproduce small-scale biochemical processes. An alternate counterargument is that consciousness is simply too complex to be artificially replicated. In response to this, Dennett says that consciousness of a more basic form may not require a sophisticated artificial substrate. Dennett ends by noting that any conscious machine will probably have to develop this capacity through an extended learning process, just as humans do. From a practical standpoint, this is not a barrier, since a number of machines that learn from experience have been designed.

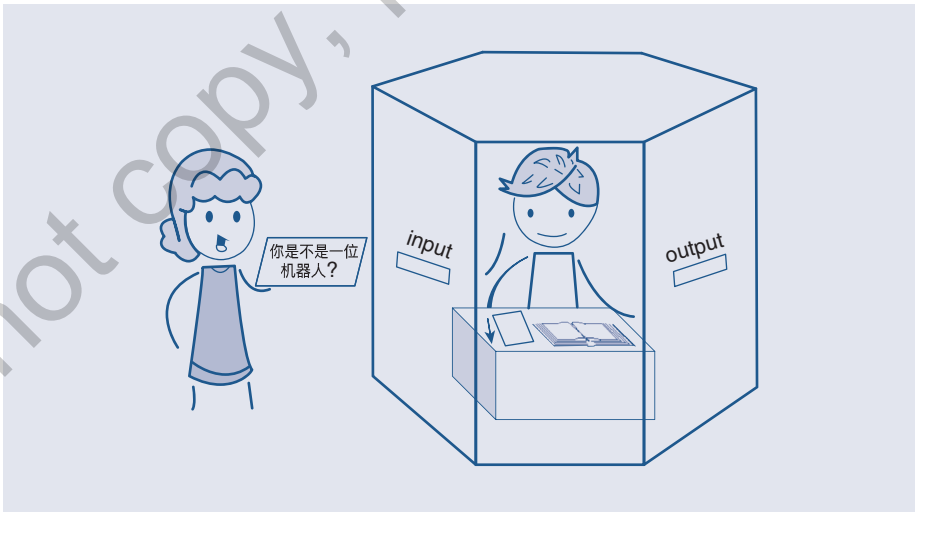
Perhaps the most persuasive and well-known argument against the strong AI position is the **Chinese room scenario** (Searle, 1980). In this hypothetical situation, a man who doesn't speak or read Chinese is in a room by himself. Outside the room is a person who writes a question or comment in Chinese and slides the paper through an input slot to the room. Although the man in the room understands no Chinese whatsoever, he has a book of rules that tells him how to relate the Chinese symbols on the incoming paper into a set of symbols constituting a suitable response (see Figure 2.10). He then traces those symbols on another piece of paper and sends it through the output slot. For example, if the outside person utters, "How are you?" in Chinese, the man in the room may, using the rule book, counter with, "I'm doing fine" in Chinese. To an outside observer, it would seem as if the person in the room understands Chinese. After all, he has a fluent reply to any question or comment that is given. But Searle's point is that the man knows

no Chinese and therefore has no understanding of the content of the conversation. He is only following a prescribed set of rules that maps one set of symbols onto another. Those symbols that he is using are not grounded in the world for him; they have no meaning for him. This is a rote execution of an algorithm, and according to Searle, it is all a machine can do. Therefore, he says, machines can never "understand," "know," or "be aware of" the information they process. They cannot be conscious of what they do. Consciousness of the human sort requires something more than just following an algorithm. To Searle, these extra ingredients are intentionality and meaning—aspects of mental representation discussed in the introductory chapter.

Boden (1990) raises a number of objections to the Chinese room argument. First, the terms *understanding* and *intentionality* are not well defined. Understanding could be operationally defined as being able to respond successfully when asked, rather than as entailing some inherent meaning on the part of the person. Second, a person who remained for some time in the Chinese room (or even a machine in the room with sufficient complexity) might eventually achieve some level of understanding. Either might eventually notice that certain combinations of characters always go together and from this learn the basic elements of syntax.

Moreover, Andy Clark's (2008) proposal that the informational tools you use become *part of your mind* (not just appendages for your cognition) would suggest that while any

Figure 2.10 Searle's Chinese room scenario.



one part of the Chinese room may not look like it understands Chinese, the room as a whole system does process and understand Chinese. Consider when you use your smartphone to have a text-based conversation with friends. Your brain knows English pretty well, but it may not always have the correct spelling for certain words, it probably hasn't memorized every single one of the emojis at your disposal, and it certainly cannot send a text message by itself. Your thumbs have learned to move in quick, articulated ways that work well with the smartphone interface, but they too cannot send a text message by themselves. And your smartphone can make loads of clever suggestions for upcoming words or emojis, and it autocorrects your misspelled words, but it better not be sending a text message by itself! Your brain, your thumbs, and your smartphone together form a whole system that fully understands how to process a text-based conversation overloaded with bizarre emojis and maintaining correct spelling. In this scenario, your brain is a little bit like the man in the Chinese room: It is only with the aid of your thumbs and the smartphone that your brain can you really generate meaningful contributions to this text-message conversation.

OVERALL EVALUATION OF THE PHILOSOPHICAL APPROACH

One of the main advantages of the philosophical approach is that it allows us to ask much broader questions than those of other disciplines. A cognitive psychologist studying memory for nouns might wonder why concrete nouns are recalled better than abstract ones. This psychologist is constrained into formulating specific questions and hypotheses by the narrow focus of the research. This very focus, of course, is an advantage since it allows the researcher to scientifically examine and understand a natural phenomenon in depth. A philosopher examining the results of this same research is free to inquire about the nature of concreteness or what it means that something is abstract. He or she could also inquire as to how concrete or abstract stimuli are processed in other cognitive systems, such as attention and language. Of course, he or she is free to ask even more fundamental questions, such as, Why do we have memory? What purpose does memory serve? What would a person be like without a memory? Philosophy thus shows us the "bigger picture." It gives us key insights into the relationships between different areas of study—within and between disciplines—and, therefore, plays a very important role in the interdisciplinary endeavor of cognitive science.

Keep in mind that philosophy is a nonempirical approach. It does not use the scientific method. Concepts in philosophy are validated through logical reasoning and argument rather than by systematic observation and experimentation. For this reason, the conclusions reached in philosophy are speculative and theoretical until tested. Philosophy is better suited to the asking of important questions—how, what, and why we should study something—than to providing definitive answers. These answers come from the scientific disciplines. It is, therefore, important that a good two-way working relationship exists between philosophers and these science-based approaches.

SUMMING UP: A REVIEW OF CHAPTER 2

1. Philosophy plays a key role in cognitive science by asking critical questions.
2. According to the mind–body problem, it is not known whether mental states are physical states. Two approaches to this are (1) monism, which argues that mind and body are unitary, and (2) dualism, which states that mind and body are separate.
3. There are two versions of monism: (1) idealism, where the universe is not physical, and (2) physicalism, where it is. Identity theorists hold that the mind is the brain, while eliminativists wish to do away completely with the concept of mind as well as with any folk psychology or commonsense explanations of the mind.
4. There are also two schools of dualism. Substance dualism holds that mind and body are composed of different substances. Property dualism holds that mental states are nonphysical features of physical bodies.
5. Functionalists believe that mental states are equivalent to the functioning or operation of physical states. It may be that different material substrates can give rise to the same mental state.
6. The knowledge acquisition problem is about how mental capacities are acquired. Nativists believe that we are born with knowledge. Empiricists believe that we attain knowledge through learning and experience.
7. Consciousness is the subjective quality of experience. The "easy problem" of consciousness is explaining qualia (subjective experience) in computational or neural terms. The "hard problem" is explaining how we have such subjective experiences at all.
8. Mind may be an emergent property of a physical brain. That is, it may not be fully explained by an understanding of its component parts and then just adding those parts together.
9. The neurobiological approach to consciousness tries to discover the neural states corresponding to different types of consciousness experience, known as the neural correlates of consciousness.
10. The strong AI view is that it may be possible some day to engineer a complex system that is conscious. In contrast, the weak AI view advocates that consciousness is either nonphysical or that it is physical but too difficult to engineer.

SUGGESTED READINGS

Bechtel, W. (1988). *Philosophy of mind: An overview for cognitive science.* Hillsdale, NJ: Erlbaum.

Blackmore, S. (2012). *Consciousness: An introduction.* Oxford, England: Oxford University Press.

Chalmers, D. (1996). *The conscious mind.* Oxford, England: Oxford University Press.

Churchland, P. M. (1986). *Neurophilosophy: Toward a unified science of the mind-brain.* Cambridge: MIT Press.

Churchland, P. M. (1995). *The engine of reason, the seat of the soul: A philosophical journey into the brain.* Cambridge: MIT Press.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension.* New York, NY: Oxford University Press.

Dennett, D. C. (1991). *Consciousness explained.* Boston, MA: Little, Brown.

Pinker, S. (2002). *The blank slate: The modern denial of human nature.* New York, NY: Viking Press.

Searle, J. R. (1992). *The rediscovery of the mind.* Cambridge: MIT Press.