

## *The Mind-Body Problem*

Some of the most important questions concerning mind, brain, and consciousness had their origin in philosophy, but they are no longer the exclusive property of philosophy. Empirical and theoretical developments in psychology, the neurosciences, and artificial intelligence have a direct bearing on them.

What are the mind and body, and what is the relationship between them? Is mind a product of brain activity, able to exist only within a living brain? Or is mind some immaterial thing, perhaps spiritual and eternal? The *mind-body problem* has implications for one's fundamental beliefs about human nature and selfhood, as well as for the psychology of consciousness. A related issue, *reductionism*, concerns the question whether it is possible to explain mental phenomena in terms of neurophysiological events.

How can we determine whether another being is conscious? After all, we cannot directly see or feel another person's conscious experience. What if the other being is a child too young to talk, or an animal, or an intelligent robot? The problem of finding an objective criterion by which to recognize the presence of mind or consciousness is known as the *other-minds problem*. Its solution is tied to our understanding of the concept of consciousness, and the question of its place in science.

In this chapter I will review some of the major viewpoints on these controversial issues: the mind-body problem, reductionism, and the other-minds problem. Besides being interesting in their own right, these issues are di-

rectly related to some of the topics in the following chapters, particularly Chapters 5 and 6 on the brain and consciousness.

## APPROACHES TO THE MIND-BODY PROBLEM

To attempt to solve the mind-body problem in detail, in all of its philosophical, psychological, and neurobiological aspects, is the most difficult and complex intellectual adventure of all. The problem may be unsolvable. The philosopher Schopenhauer called it the "world knot."

What are the essential natures of the mind and the body, and how do they relate to each other? In philosophers' terms, the mind-body problem is an *ontological problem*, that is, a question about what things really exist, and what their essential nature is. Some writers call it the "mind/brain" problem, since we now know that the brain is the part of the body that is most critical for mind, though there once was a time when people believed that the mind or soul was in the heart.

To begin with, let us grant that the body, including the brain, is made of material (physical) substance, complexly organized, and operating largely according to principles of cause and effect. I will dismiss as too mystical the notion that the world, including my body, is merely an idea in someone's (God's?) mind (*idealism*, or mentalistic monism).

The critical issue, then, is whether mind is produced by the material brain. Or is mind a product of some immaterial substance, capable of existing independently of the brain? And exactly how do mind and brain relate to each other, for example, do they operate in parallel, or are they identical, or do they interact—and if they interact, how do they do it? There are many specific theories on the mind-body problem, but they can all be classified into two broad categories: dualism and materialism.

*Dualism* is the belief that the mind and brain/body are made of different substances.<sup>1</sup> The brain is made of material substance, whereas the mind is made of some sort of immaterial or nonphysical substance. Dualism implies that the mind can exist independently of the brain. Dualism is consistent with most religions, in which mind is identified with the soul. If you believe in some sort of continuation of your mental life after the death of your body, then you are a dualist.

Materialism (materialist monism) is the belief that there is only one type of substance, namely material substance. Thus, mind and consciousness are functions of complexly organized matter, probably (but not necessarily) limited to organic brains. If your brain dies, your mind dies. Materialism is the viewpoint adopted by most scientists, although there are some noteworthy exceptions.

It could be argued that there are at least two mind-body problems, one concerning mind in the sense of the control of behavior by thought processes, and another concerning consciousness in the sense of subjective experience (Weimer 1976). This distinction was not made in most of the older philosophical discussions of the mind-body problem, since it was thought that mind and consciousness are synonymous. However, as Nagel (1979) suggested, "Without consciousness the mind-body problem would be much less

interesting. With consciousness it seems hopeless" (p. 166). I am more optimistic than Nagel, but I agree that consciousness makes the mind-body problem especially difficult.

There are at least five varieties of dualism, and a comparable number of varieties of materialism (P. M. Churchland 1988). I will describe only a few of the most influential viewpoints on the mind-body problem. But first I must say a bit more about the importance of the problem.

### Why the Mind-Body Problem Is Important

Science is concerned with discovering the principles of organization and cause-and-effect relationships that govern natural phenomena. Only if you adopt a materialist viewpoint can you be confident that a science of mind is possible. If mind is dependent upon the physical brain, then it is reasonable to assume that both behavior and mental phenomena, including consciousness, operate according to some reasonably consistent general principles that can be discovered by scientific methods. But if you adopt a dualist viewpoint then there is no reason to believe that there are any general principles, or if there are, then you cannot assume that they will be discoverable by scientific methods. If you adopt a compromise position, saying that behavior is a function of the material brain but that consciousness is not, then you have to settle for a limited psychology that does not even attempt to answer questions about consciousness with scientific methods.

The materialist viewpoint is also valuable for clinicians—such as neurologists, neuropsychologists, psychiatrists, and clinical psychologists—who deal with abnormal mental functioning. The materialist view tells us that it is reasonable to search for the causes of such abnormalities in the physical world. The ancient theory of demonic possession has been replaced by concepts such as brain damage or dysfunction, inadequate learning experiences, and psychologically stressful environments.

Finally, your stand on the mind-body problem is fundamental to your view of what it means to be human. Are we entirely the creation of our biology and our culture, with intimate ties to the natural world because of our origin in it? Or are we somehow fundamentally different and separate from the natural world, due to the presence of some immaterial mind-soul? Will our self-awareness die when our brain dies, or will it continue, immortal, in some nonphysical realm?

## DUALISM

The most important variety of dualism is *interactionist dualism*. Its best statement was made by the seventeenth-century French philosopher René Descartes (1596–1650), and so it is often called *Cartesian dualism*.

Descartes asserted that the behavior of animals can be explained by assuming that they are merely complex reflex machines. If their behavior can be explained in this way, then there is no reason to believe that animals have mind or consciousness. However, Descartes believed that human behavior is governed by rational thoughts, not reflexes. Rational thinking is a product of

mind, or soul, which is something entirely different from the material body. (Note that Descartes lived over two hundred years before Darwin published the theory of evolution and Freud published his psychoanalytic observations. Thus, Descartes was unaware of the evolutionary continuity between animal and human minds/brains, and he was unaware of the role of nonconscious and nonrational mental processes that influence human behavior.)

There were two reasons why Descartes believed in mind for humans but not for animals (P. M. Churchland 1988). First, there was the *argument from religion*. It was consistent with Descartes' Catholic religious beliefs to hold that humans have an immortal soul, and that the soul is responsible for our thoughts. Second, there was the *argument from introspection*. Descartes introspected on his conscious experience and came up with his famous dictum, "*Cogito, ergo sum*" ("I think, therefore I am"). What he meant was that he observed that he thinks, and therefore somebody must be doing the thinking. Thoughts seem to be immaterial phenomena. He could not conceive how his physical brain/body could be responsible for his thoughts. Therefore, he concluded that mental phenomena *cannot* be caused by mind/brain processes (*argument from irreducibility*), and so he attributed his thoughts to an immaterial entity, the soul. Furthermore, introspection led him to conclude that the actions of his physical body depended upon his mind's rational thoughts, and that the body's senses kept the mind informed of the body's actions. Mind and body are of different substance, but the different substances interact: hence, Cartesian dualism is *interactionist* dualism.

The German philosopher Leibniz (1646–1716) proposed *parallelist dualism* (also called psychophysical parallelism), the theory that an immaterial mind/soul has perceptions and ideas that are correlated in time with the body's actions, but that the mind/soul does not actually control the body. The parallelist dualism idea has never had many supporters, however, compared to interactionist dualism.

**Critique of Cartesian dualism.** Paul M. Churchland (1988) summarized several of the materialist arguments against dualism. First, there is the *principle of parsimony*, a fundamental principle of scientific theorizing which says that if two competing hypotheses are equally good at explaining a phenomenon, then we should choose the simpler one. Another version of the same principle is known as "Ockham's Razor," after the medieval philosopher who first stated it. Ockham said "Do not multiply entities beyond what is strictly necessary to explain the phenomena." By this rule, materialism is to be preferred because it postulates only one kind of substance (material), whereas dualism postulates two types (material and spiritual).

Dualism can be rejected prior to the application of Ockham's Razor because of its *explanatory impotence*. Dualism makes no attempt at a detailed explanation of how mind accomplishes things such as perception, memory, thinking, emotion, and control of the body's movements. Materialistic neuroscience, on the other hand, has made considerable progress toward an explanation. Though a detailed explanation of how the brain produces mental phenomena is not yet available, there is evidence that firmly supports the materialist viewpoint, namely, the fact that thinking and consciousness can

be severely altered by brain damage and by drugs that affect the transmission of neural impulses in the brain. (I will be reviewing some of this evidence in later chapters.)

A major problem with interactionist dualism is that it cannot explain how an interaction between a nonphysical mind substance and material brain substance could occur. Something nonphysical could not produce a physical effect without violating the laws of conservation of mass, energy and momentum. For example, how could something nonphysical push or pull a physical object to make it move?

Dualism's strongest card is the *argument from introspection* combined with the *irreducibility argument*. Introspectively, the subjective qualities of our sensations (sensory "qualia"), emotional feelings, mental images, and the meaningful contents of our thoughts seem to be entirely different in kind from physical substance. Hence, dualists argue that it will forever be impossible reductively to explain mental phenomena in terms of physical, neurophysiological events. Churchland refuted the argument from introspection:

The argument is deeply suspect, in that it assumes that our faculty of inner observation or introspection reveals things as they really are in their innermost nature. This assumption is suspect because we already know that our other forms of observation—sight, hearing, touch, and so on—do no such thing. The red surface of an apple does not *look* like a matrix of molecules reflecting photons at certain critical wavelengths, but that is what it is. The sound of a flute does not *sound* like a sinusoidal compression wave train in the atmosphere, but that is what it is. The warmth of the summer air does not *feel* like the mean kinetic energy of millions of tiny molecules, but that is what it is. If one's pains and hopes and beliefs do not *introspectively* seem like electrochemical states in a neural network, that may be only because our faculty of introspection, like our other senses, is not sufficiently penetrating to reveal such hidden details. Which is just what one would expect anyway. The argument from introspection is therefore entirely without force, unless we can somehow argue that the faculty of introspection is quite different from all other forms of observation (1988, p. 15).

To continue this argument by analogy, just as the eye cannot directly see what it itself is made of or how it works, the mind cannot directly perceive what it is made of or how it works. For the eye to see itself, it needs a mirror. For the mind to perceive itself it needs tools more complex than mirrors, such as the tools of brain science, neuropsychology, and experimental psychology. Introspection alone is not sufficient. With the argument from introspection defeated, the irreducibility argument loses most of its force. Whether mental phenomena are reducible to neurophysiological events is a matter to be decided by scientific research and clinical observations, not by *a priori* arguments.

### Parapsychological Experiences: Evidence for Dualism?

Aside from the arguments from religion and introspection, is there any independent evidence to support a dualist viewpoint on the mind-body problem? The answer depends largely on your prior beliefs. Over the centuries, many people have claimed to have had *parapsychological experiences*—ex-

periences that *seem* to be contrary to the laws of nature as understood by orthodox scientists (Irwin 1989). People have reported seeing apparitions (ghosts), having houses haunted by poltergeists (noisy ghosts), and communicating with dead loved ones through mediums (people supposedly able to make contact with the spiritual world). Some have claimed that they themselves are reincarnated souls and that they have memories of a prior life. Also, *psi experiences* (clairvoyance [extrasensory perception, ESP], telepathy, precognition, psychokinesis) have been reported by many people. For example, many people have reported that they somehow "knew" that a loved one had died at the moment of death, even though the loved one was far away and the death was unexpected. More mundane types of psi experience, such as remote perception (clairvoyance) of events such as the cards turned up in a randomly shuffled deck, have been studied experimentally by parapsychologists.

In the absence of ready explanations of parapsychological experiences by materialist scientists, some dualists have claimed such experiences as evidence in proof of some sort of *paranormal* processes or things, such as immaterial mind-stuff or souls, that have characteristics that are contrary to the physical world as we know it. In fact most scientists ignore claims for parapsychological experiences, dismissing them as either without adequate proof or, at worst, downright fraudulent (Moss & Butler 1978; Randi 1982). Admittedly, parapsychological experiences are so contrary to current materialist scientific beliefs about how the world works that most scientists do not want to spend time examining the evidence. The evidence that is available is highly controversial, at best. Whether people accept the available evidence as proof of the reality of paranormal phenomena depends largely on what they want to believe. If psi experiences represent genuine psi phenomena, then they cannot be explained by current materialist, scientific theories. On the other hand, dualist theory cannot explain psi phenomena, either. There is merely the assertion that, since these strange events cannot be explained by materialist theory, therefore by default they support dualism (Irwin 1989).

In recent years there has been increased attention to two types of parapsychological experiences, the out-of-body experience (OBE) and the near-death experience (NDE), that have been offered as evidence for a dualist viewpoint. Compared to psi and other parapsychological experiences, the OBE and NDE have been taken more seriously by scientists. Rather than disputing the reality of the subjective experiences, scientists have attempted to explain them in materialist terms.

In the *out-of-body experience* (OBE), the center of awareness appears to the individual to occupy temporarily a position that is spatially separate from his or her body (Irwin 1985, 1989). For example, while lying in bed you might have the experience of floating out of your body and then looking down on it from a place near the ceiling. It has been estimated that OBEs have occurred in about 8 to 15 percent of the general population. Estimates among college students are greater (20 percent or more), perhaps because college students are more willing to report such experiences. According to introspective verbal reports by OBEers, in about 80 percent of OBEs the things seen by the (allegedly) out-of-body consciousness appear to be naturalistic. Fantastic or transcendental experiences occur mainly in cases where the

OBE is a component of a near-death experience. In typical cases, OBEers say that their mental state during the OBE was one of relaxed alertness, mental clarity, effortless concentration, peacefulness, and emotional detachment. Some people claim to be able to control the content of their OBE experience by shifting their attention. If the OBE is truly what it subjectively seems to be—a separation of consciousness from the body—then it would support a dualist position on the mind-body problem. However, scientific explanations have been offered.

About 90 percent of OBEs occur while the individual is physically inactive, usually while lying down. Also, most OBEs occur under conditions of reduced sensory stimulation. This suggests that reduced attention to body senses (kinesthetic [movement] and proprioceptive [position]) may be an important condition for OBEs. Supporting this inference are reports that shifting attention to bodily processes brings the OBE to an end (Irwin 1989). It has been suggested that OBEs occur in the hypnagogic (drowsy, presleep) state, or in REM sleep, when vivid dreams occur. LaBerge (1985) found OBEs to be associated with lucid dreams—REM dreams where the dreamer is aware that he or she is dreaming. However, while some OBEs may occur in hypnagogic or sleep states, there are many exceptions. H. J. Irwin (1985) argued that the OBE depends on a combination of physical factors (relaxation, sensory deprivation) and cognitive factors (absorbed imagination, inattention to body processes). However, other factors may be involved when an OBE occurs in the context of a near-death experience.

The *near-death experience* (NDE) is a special sort of mystical experience that has been reported by many people who have narrowly escaped death. For example, NDEs have occurred in cases of people nearly killed by drowning, automobile accidents, or heart attacks, or during or after major surgery. Though not all NDEs are the same, Raymond Moody (1975) reported that most NDEs have some of the following characteristic features (cited in Irwin, 1989): (1) *Positive affect*, with feelings of peacefulness or joy; or simply emotional detachment, with freedom from fear or pain. This is the most common feature of NDEs, occurring in almost all cases (Ring 1980; Sabom 1982). (2) *Out-of-body experience*. For example, some people have reported that they could see their body, and the activities of people trying to revive them, from an objective viewpoint. Unfortunately there have been few attempts to verify the accuracy of these descriptions. OBE is one of the most common features of NDE; estimates of OBE in NDE range from 37 percent to 99 percent. (3) *Floating through a dark and empty space*, sometimes described as a tunnel. (4) *Emerging from the darkness into a realm of brilliant light*, which seems to signal entry into a nonphysical realm. The light has a reassuring quality. Sometimes it is experienced without first passing through darkness. (5) *Encountering some sort of "presence,"* described by Moody (1975) as a "being of light." The being's presence may be known intuitively, rather than sensed directly. Sometimes there is a discussion with the being over whether the individual is to die or return to the physical body. Encounters with a "presence" occur in about 40 percent of NDEs. Some studies (such as Ring 1980; Sabom 1982) have found that there is no relationship between experiencing the "presence" and the individual's prior religious beliefs. On the other hand, "it is feasible that even the most avowed atheist becomes an instant 'believer'

when confronted with imminent death, if only for the duration of the threat to life" (Irwin 1989, p. 189). (6) *Panoramic life review*, involving a sequence of vivid visual images of events from the individual's life. The panoramic review occurs spontaneously, without any effort of recall by the individual. It occurs in about 25 percent or less of NDEs. (7) *Entering a transcendental (unearthly) realm*, usually described as a pastoral paradise with lush green grass, trees, beautiful flowers, and vivid colors. Some individuals report encountering the spirits of deceased loved ones, who reassure them. Interestingly, the transcendental realm in NDEs is rarely described as being like Biblical descriptions of heaven. In rare cases it is described as more like hell. The transcendental realm is estimated to occur in 20 to 54 percent of NDEs.

NDEs often have profound aftereffects on peoples' attitudes. Ring (1984) found that the NDE can serve as a "spiritual catalyst": core NDEers (people who had had profound NDEs) reported that afterward they felt closer to God, less materialistic, more appreciative of life and other people, more self-confident and secure, and that they had a greater belief in the underlying unity of all religions. NDEers reported a reduced fear of death, and an increased belief in some sort of afterlife. Furthermore, these attitude changes were not due merely to the close brush with death *per se*. Among people who had nearly died, the attitude changes were greater for those who had had an NDE than for those who had not had an NDE.

Near-death experiences would seem, on the surface, to provide strong support for the dualist view that consciousness can exist independently of the body and can survive the death of the body. However, alternative interpretations of NDEs have been offered by scientists who take a materialist viewpoint on the mind-body problem (Irwin 1989; Shaver 1986). NDEs may be hallucinations produced by physiological and/or psychological states that occur in near-death situations (Siegel 1980). *Hallucinations* are mental images, usually vivid, that occur spontaneously, without voluntary control by the individual, and that the individual interprets as real sensory perceptions. Hallucinations may be produced by oxygen deprivation (or equivalently, by an increase in carbon dioxide) in the brain, or by high fever. A sudden massive release of endorphins—morphine-like neurotransmitters that reduce pain and produce a sense of psychological well-being—could account for some aspects of NDEs. Some NDEs can be explained as hallucinations produced by surgical anesthetics or other drugs. NDE hallucinations might result from a general disinhibition of cerebral activity, or more specifically, from seizure-like discharges in the temporal-parietal region, which is involved in mental imagery and in the body image. None of these physical hypotheses accounts for all aspects of NDEs, though each is probably relevant to some cases of near-death experience. Psychological factors, such as fear of death and strong desire to survive, probably interact with physical factors to produce the NDE in many cases. Some aspects of hallucinations, such as a dark tunnel or a bright light, may be produced by physical factors, but the individual's interpretation of the experience depends on psychological factors. In general, NDEs can be explained as vivid hallucinations resulting from brain state changes that produce vivid mental images combined with a decreased ability to distinguish imagination from reality, where the specific nature of the imagery and/or its interpretation is influenced by psy-



chological factors such as fears, desires, and memories associated with the near-death situation.

Logically, the availability of a naturalistic interpretation of NDEs does not disprove the dualistic interpretation, but it makes the dualistic interpretation unnecessary. It is worth noting that persons who reported the near-death experience *did not really die*. Thus, we have no first-hand reports of life after death. Some of the questions raised by the NDE may never be answered because it is impossible to do experimental research on this topic. Research ethics committees at universities and hospitals will not allow psychologists to cause research subjects to nearly die, so the NDE cannot be studied under controlled laboratory conditions. Meanwhile, acceptance of the naturalistic interpretation of NDEs does not prevent us from agreeing with Ring's idea (1984) that the NDE may be a critical, age-old source of spiritualistic religious beliefs.

## VARIETIES OF MATERIALISM

The many varieties of materialism are united only by their assumption that mental phenomena are produced by organized physical substance. I will briefly summarize some of the most influential of the materialist viewpoints on the mind-body problem.

### Epiphenomenalism

Epiphenomenalism is the view that consciousness exists, and that it is produced by the brain, but that it plays no role in controlling the actions of the body. (The notion that consciousness has no causal efficacy brings to mind Leibniz's parallelist dualism.)<sup>2</sup> The epiphenomenalist viewpoint was adopted by radical behaviorists in the 1920s. The behaviorists wanted to establish psychology as an empirical science that studies only objectively observable behavior, and not subjective mental states. "Behaviorism claims that consciousness is neither a definable nor a usable concept" (Watson 1924, p. 3). Since they did not know how to study mental events objectively, the behaviorists justified their ignoring of mental events by denying that mental events are important for controlling behavior. More recently B. F. Skinner (1971, 1974) restated the behaviorist viewpoint on mental events:

We do not need to try to discover what personalities, states of mind, feelings, traits of character, plans, purposes, or other perquisites of autonomous man really are in order to get on with a scientific analysis of behavior. . . . We do feel certain states of our bodies associated with behavior. . . ; they are by-products and not to be mistaken for causes (Skinner 1971, pp. 13-14).

We will see that there are reasons to believe that conscious states are something more than mere useless by-products of brain processes.

### Identity Theory

Identity theory says that *mind and brain are one*. Or to be more precise, for each unique mental state there is a corresponding brain state.<sup>3</sup> From this

viewpoint *all* mental phenomena can, in principle, be reduced to specific brain phenomena. Hence, identity theory is sometimes called *reductive materialism* (P. M. Churchland 1988).

To identify mind with brain is to say that they are inseparable. They exist in the same time and space. We have the impression that they are different, and we describe them with different vocabularies, because we have two different perspectives—subjective and objective—on the same thing.

From the objective perspective, when we examine another person's brain (or our own brain, with our skull opened up, using a mirror) we see only a mass of tissue, and we do not see any mental processes. Even if we slice up a brain and look at it under a microscope, or measure its electrical activity with implanted electrodes, or measure its bloodflow with a PET scanner, we see only physical structures and events, not mental processes. We describe what we see in the language of neuroscience, using words like "neuron" and "synapse" and "neurotransmitter."

Subjectively, as intact living and functioning people, we know our brain from a different perspective. We perceive objects and we have mental images and verbal thoughts. These mental events are brain events perceived from the subjective viewpoint, and from that viewpoint they appear to be very different from neural events. Because mental events are phenomenally so different from objective neural events, we use a different vocabulary to describe mental events.

The identity theory is particularly popular among neuroscientists. The promise of identity theory is that ultimately it will be possible to translate all statements about mental states or processes into statements about brain states or processes. Insofar as this has not yet been accomplished, it must be admitted that the identity theory rests on a reductive materialist faith. This faith is sustained by the progress that neuroscience is making in understanding how the brain works.

What is the difference between identity theory and epiphenomenalism? If consciousness is identical to brain events, can it have any role in controlling behavior? D. L. Wilson (1978) gave an identity theorist's reply to epiphenomenalism:

According to the identity theory, consciousness is both a brain process and an active force in the behavior of higher organisms. That active force can be viewed from the personal side, as the experiencing self does, or it can be viewed, in principle, by the neurophysiologist with implanted electrodes and so forth. This is not a dual-aspect position, but is merely two perspectives on the same brain events. Those who question why consciousness should have evolved and what survival value it has apart from brain events are assuming a dualistic view in order to attack a monistic position. Within an identity theory paradigm, to ask about the Darwinian survival value of consciousness is to ask about the survival value of the brain processes that *are* conscious events. Since these are likely the processes involved in situation-analysis and decision-making in the brain, their survival value is obvious. Furthermore, as the identification of mind with brain allows for a causal role of mental functions in human behavior, it allows for personal dignity, freedom of choice, and other humanistic attributes, despite opinion to the contrary (p. 13).

The main complaint against identity theory is the dualist argument from introspection and the irreducibility argument, which says that subjective experience is so different from brain processes that they could not be two aspects of the same thing. I have already presented Churchland's (1988, p. 15) refutation of the dualist argument.

Irving Kirsch (1985; Kirsch & Hyland 1987) proposed the principle of *causal isomorphism* to account for the relationships between mental events, neurophysiological events, and behavior. The main idea is that the causal relationships between two mental events are *functionally equivalent* to the causal relationships between the corresponding physiological events. The principle of causal isomorphism, according to Kirsch, rests on three related assumptions:

- (a) There is (in principle) a physiological counterpart to any instance of a mental event; (b) the relation between a mental event and its physiological substrate is better described as an identity relation than as a relation of cause and effect . . . ; and (c) then for any causal sequence of mental events, there must be a corresponding sequence of physical events (Kirsch & Hyland 1987, pp. 421-22).

Point (b) deserves emphasis. According to identity theory it is not strictly correct to say that mental states are *produced by* physiological states; rather, mental states *are* physiological states, as they are subjectively experienced. Likewise, it is not strictly correct to say that mind and body *interact*, since such phrases suggest a dualist view that mind and body are separate and different things. Rather, physiological states interact with each other, and some of them correspond to mental states.

Figure 4.1 shows some diagrams that Kirsch and Hyland (1987) used to illustrate the principle of causal isomorphism. Figure 4.1A illustrates causal relationships in psychosomatic illness, in which mental states such as "feeling inadequate" are said to cause illness. But "feeling inadequate" corresponds to Brain State 1, which is part of the sequence of physiological events that leads to illness. Figure 4.1B illustrates an example of a drug placebo effect, in which a person is given an inert substance (a placebo) and told that it is a stimulant, with the result that he feels more tense. Strictly speaking, the placebo and instructions produce a brain state corresponding to expected arousal, and that brain state produces physiological responses (increased pulse rate, etc.) and a brain state that corresponds to a subjective feeling of tension. Figure 4.1C shows the events in an emotional reaction, in which perceived danger (such as the presence of a lion) is followed by feelings of fear. Strictly speaking, a brain state corresponding to "perceived danger" is part of a sequence of physiological events leading to increased pulse rate and other body and brain responses that corresponds to the feeling of fear. This sequence illustrates the important point that causal sequences of physiological events often include some nonconscious physiological events that do not correspond to any conscious mental state.

In conclusion, to reiterate Wilson's point (1978), to deny the adaptive significance of mental events is to deny the importance of the brain events to which they correspond. It is appropriate to develop theories about the causal

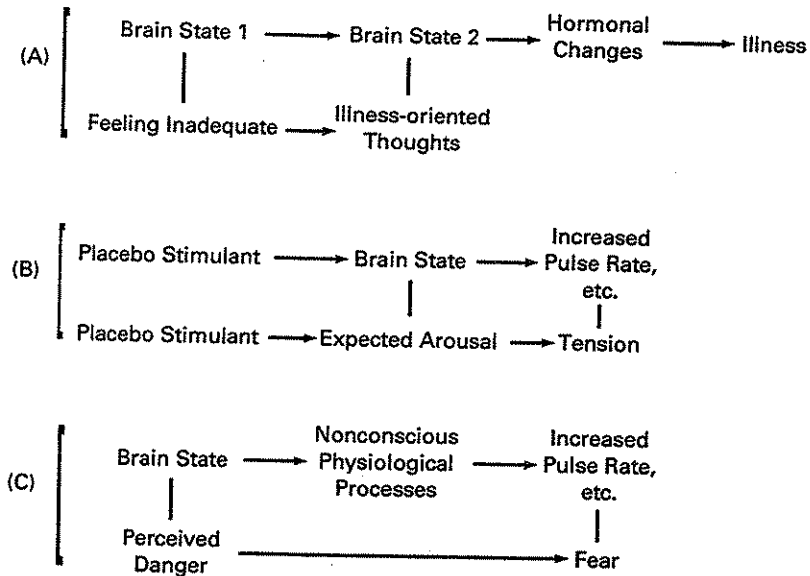


FIGURE 4.1 Examples of causal isomorphism between mental states and physical states. Functionally isomorphic mental and physical states are linked by vertical bars. Causal sequences are shown by horizontal arrows. (A) Psychosomatic illness. (B) Placebo effect. (C) Emotional reaction. [From Kirsch, I. & Hyland, M. E. (1987). How thoughts affect the body: A metatheoretical framework. *Journal of Mind and Behavior*, 8, 417-34. By permission of the publisher.]

efficacy of mental states, since mental states correspond to physiological states that have causal efficacy.

### Emergent Interactionism

Roger Sperry (1969) argued that consciousness cannot be understood merely by analyzing the physiological and molecular processes of the brain. Rather, consciousness is an *emergent* phenomenon. An emergent phenomenon is one that appears as a result of a unique relationship among the parts of an organized system, and which cannot be predicted from a knowledge of the parts alone. Life itself is an emergent phenomenon: its characteristics cannot be predicted from a knowledge of the properties of organic molecules. Nor can the characteristics of complex organisms be predicted from a knowledge of their individual cells.

Sperry's emergent interactionism emphasizes the causal efficacy of consciousness as an emergent phenomenon:

Consciousness awareness, in the present view, is interpreted to be a dynamic emergent property of cerebral excitation. As such, conscious experience becomes inseparably tied to the material brain process with all its structural and physiological constraints. At the same time the conscious properties of brain excitation are conceived to be something distinct and special in their own right. They are 'different from and more than' the collected sum of the neuro-physico-chemical events out of which they are built.

Compared to the elemental physiological and molecular properties, the conscious properties of the brain process are more molar and holistic in nature. They encompass and transcend the details of nerve impulse traffic in the cerebral networks. . . . It is the emergent dynamic properties of certain of these higher specialized cerebral processes that are interpreted to be the substance of consciousness. . . . The subjective mental phenomena are conceived to influence and to govern the flow of nerve impulse traffic by virtue of their encompassing emergent properties. . . . (Sperry 1969, pp. 533–34).

Some critics (such as Bindra 1970) have accused Sperry of being a dualist, since they interpret Sperry's hypothesis as implying that once consciousness emerges from cerebral excitation it has an autonomous existence and can act independently of the underlying neural organizations. In reply, Sperry (1970, 1980) argued that his position is both mentalist and materialist, but not dualist. Sperry's position is materialist in asserting that mental forces are identical with the holistic properties of cerebral excitation, and "in denying that these mental forces can exist apart from the brain process of which they are a direct property." It is mentalist "in accepting the existence of potent mental forces that transcend the material elements in cerebral function" (Sperry 1969, p. 534). Recently, Sperry (1987) reemphasized the importance of "downward control" of more elementary physiological brain processes by higher-level conscious mental processes, and suggested that downward control is a revolutionary scientific concept and a major reason for the renewed interest in consciousness in psychology and neuroscience.

There are three major differences between Sperry's emergent interactionism and identity theory (Sperry, 1976). First, identity theory argues that a reductive explanation of all mental processes in terms of brain processes is possible in principle. Sperry denies that reduction is possible, even in principle, in regard to higher-order, conscious mental processes. He was specific on this point:

The process of reducing an entity to its material components, physically or conceptually, inevitably destroys the space-time components at the affected level. . . . The spacing and timing of the parts with reference to one another largely determine the qualities and causal relations of the whole but the laws for the material components fail to include these space-time factors (1980, p. 203).

Second, Sperry draws a sharper distinction between conscious and nonconscious processes than identity theory does. There is more of a continuity between conscious and nonconscious processes in identity theory. Third, the concepts of interaction and emergent downward control are not used in identity theory, insofar as they imply that consciousness is something different than physiological brain processes. Of course, neuroscientist identity theorists acknowledge that the complex organization of the brain is critical for mental processes, including consciousness, and that hierarchically-organized control systems are important. But rather than saying that consciousness interacts with elementary brain processes, identity theory says that various brain processes interact with each other, and some of those pro-

cesses are identical to conscious experiences. (See Natsoulas 1987 for a discussion of criticisms of Sperry's interactionism.)

Sperry has been heavily criticized by neurophysiologists for denying that a reductive explanation of mind in terms of brain is possible, in principle. But this very point has made Sperry's hypothesis popular among many psychologists. Most psychologists take a materialist viewpoint, in view of the obvious effects of drugs and brain damage on thinking and consciousness. But many psychologists do not believe that it is fruitful to try to develop detailed explanations of mental phenomena in terms of brain processes. Psychologists have proceeded to develop theories that try to explain learning, memory, thinking, emotion, consciousness, and behavior in terms of higher-level, psychological constructs, without paying much attention to neuroscientific evidence on how the brain works.

I must admit that I am torn between the identity theory and emergent interactionism. I lean toward the identity theory, since it seems to be more consistent with the notion that consciousness can exist in a variety of forms, and it is clear that some degree of mind-brain reductionism is possible, particularly in regards to the "lower-level" processes of sensation and motor control and some aspects of emotion. On the other hand, while I am in principle a materialist regarding the entire mind-body problem, I admit that the higher-level aspects of thinking, consciousness, and personality are so complex that it seems unlikely, as a practical matter, that they will be explained reductively in very much detail. There is still a very practical role for psychological theories, since they can explain mental processes in ways that are useful for education, psychotherapy, interpersonal relationships, self-understanding, and self-control.

### Functionalism

According to modern functionalist theory, mental states have meaning only in a particular context. They are not independent entities. In this view, a mental state (a belief or feeling, for example) can be defined according to its *functional role* in a cause-and-effect network involving (a) stimulus inputs to the organism, (b) other mental states, and (c) the organism's actions (P. M. Churchland 1988). For example, the mental state of fear is elicited by certain objects (such as grizzly bears), and it serves the functional role of producing certain actions (such as running away or preparing to defend yourself).

Though functionalism is materialist, it is not particularly concerned with the microscopic details of the physical bases of mental processes. If the important thing about mental states is their functional role, then it does not make much difference what their physical basis is. Mind might be produced by either a brain or a computer, in the modern functionalist view. Thus, modern functionalism disagrees with the identity theory argument for *type physicalism*, in which all mental states correspond to specific neurophysiological brain states. Instead, functionalism supports a *token physicalism*, in which all functional mental states correspond to specific physical states in *some* physical system—but the physical system does not have to be an organic brain of the type that has evolved on Earth over the last 500 million years (Fodor 1981).

The token physicalism principle has made functionalism particularly popular among cognitive psychologists and artificial intelligence researchers. Cognitive psychologists feel justified in developing theories about cognitive processes described in abstract terms, without worrying about their neurophysiological basis. Artificial intelligence (AI) researchers write computer programs to carry out the functional equivalent of mental processes—that is, to make the same kinds of decisions that a human mind could make. From the functionalist viewpoint, it is conceivable that computers (perhaps a future generation of parallel-processing computers) could have mental states, including consciousness. Mental states and processes must be produced by physical matter, but the important thing about them is what they do, not the type of physical substance that does it.

Fodor (1981) sees functionalism as a combination of the best ideas of identity theory and logical behaviorism. *Logical behaviorism* says that mentalist terms have *implicit operational definitions*, that is, they can in principle be defined in terms of people's *dispositions* to behave in particular ways in particular situations. For example, the mental state of pain can be defined in terms of people's dispositions to behave in certain ways in response to certain types of stimuli, such as when they smash their toes against chair legs while walking barefoot across a room in the dark. But identity theory, unlike logical behaviorism, says that mental states (being identical to neurophysiological states) have causal efficacy in themselves. Functionalism combines the idea of implicit operational definitions of mental states with the idea that they have causal efficacy.

The main criticism of functionalism is that it does not account for subjective consciousness experience (such as sensory qualia and feelings). Functionalists argue that the exact nature of subjective experience is not important; what is important is the functional role that it plays in regulating the person's behavior. For example, consider the *inverted spectrum thought-experiment* (P. M. Churchland 1988). If we both see a light of 700 nm wavelength, and I have a subjective experience of "red" and you have a subjective experience equivalent to what I would call "blue," it does not make any practical difference so long as you call it "red" and react to it in the same way that I react to a red light (hit the brakes!). And maybe Martians do not feel exactly the same way that we do when they stub their toes, but they still hop around on one foot and say bad words.

So far, OK. But functionalism's problem with conscious experience is more serious. There is the *absent qualia problem*. Modern functionalist theories can get by with no subjective qualia at all. In principle, twenty-second-century androids might have mental processes that would be functionally equivalent to those of humans, from an objective viewpoint, but the androids would not subjectively feel anything at all. They would be nonconscious automatons. But humans are not nonconscious automatons. Consciousness is an intensely real aspect of our existence.

Another complaint against modern functionalism, with its emphasis on computer modeling of mental processes, is its assumption that all mental processes are computational processes (P. S. Churchland 1986). There is growing evidence that the computational model is inadequate to account for mental processes, including consciousness. Minds are like computers only in

a very loose analogy, and the analogy cannot be carried very far toward understanding the mind/brain system. The "new wave" in cognitive theory is to design "connectionist" models of mental processes based on complex multiple and parallel connections between simple elements—models more like a brain than like a computer (Bechtel & Abrahamsen, 1991; Martindale 1991). If connectionist models succeed better than computer models of mental processes, the implication is that the token physicalism principle is wrong. The physical basis of mental processes is not arbitrary (Thagard 1986). Brains, or brain-like structures, are necessary to produce mind and consciousness.

## REDUCTIONISM

The mind-body problem is intimately interwoven with the question of reductionism. Is it possible, in principle, to explain mind and consciousness in terms of neurophysiological processes in the brain? Dualism says no. Identity theory says yes. Emergent interactionism takes an intermediate position.

Dualism insists that the subjective "qualia" of experience are so totally different that it is inconceivable that they can be explained in neurophysiological terms. If we cut open a brain and examine it microscopically we cannot see anything that resembles subjective experience. Nor do chemical processes at neuron synapses resemble subjective experience.

Methodologically, the best that we can hope for is to find detailed *correlations* between specific subjective experiences and specific patterns of neural activity in the brain. Much progress has already been made in this direction. But we will not be able to observe anything that looks like a direct cause-and-effect relation. We will not be able actually to *see* that a specific neural action causes a specific subjective experience in a manner analogous, for example, to the way we can see that if we push a book over the edge of a desk then the book will fall to the floor. The problem is that we cannot see or feel another person's subjective experience. Thus, methodologically, it would seem to be impossible for empirical science to prove that conscious experience is caused by brain processes. Reductionism is not possible, in principle, by this argument. But the modern view of reductionism escapes from this methodological problem.

In the modern view, what is required is *intertheoretical reduction*. It must be possible to make a direct statement-by-statement translation between a psychological theory of mind and a neurophysiological theory of brain processes. If a psychological theory could explain everything about thinking and behavior in psychological terms, and a neurophysiological theory could explain everything in neurophysiological terms, and all statements in the psychological theory could be shown to correspond to statements in the neurophysiological theory, then the psychological theory would be successfully *reduced* to the neurophysiological theory. In principle, the neurophysiological theory would make the same predictions about behavior that the psychological theory would make.

An extreme modern view on the mind-body problem, *eliminative materialism*, says that intertheoretical reduction will not, and cannot, succeed (P. M. Churchland 1988; P. S. Churchland 1986). In particular, eliminative material-



ism argues that folk psychology—popular conceptions of mind based on introspection and mentalist concepts (desire, belief, intention)—is inaccurate, misleading, and unsuitable for intertheoretical reduction. Nor are current formal psychological theories suitable candidates for reduction. Not only do they fail to provide a full account of thinking and behavior, but they are not expressed in terms that are suitable for translation into neurophysiological theory terms. Therefore, in the eliminative materialist view, psychological, and especially mentalist, concepts should be—and will be—eliminated as scientists learn more about the brain and develop better neurophysiological explanations of human behavior.

In reply, there are two points: First, eliminative materialism concentrates on neurophysiological explanations of behavior but it ignores conscious experience as something worthy of being explained. Thus, like behaviorism, eliminative materialism would eliminate a major aspect of human nature from scientific study. Second, it could be argued that intertheoretical reduction *might* be possible at some time in the future, if brain scientists and psychologists were more mutually aware of each other's work, and if neurophysiological and psychological theories were to evolve together (P. S. Churchland 1986; Hatfield 1988). However, intertheoretical reduction is more likely to succeed for relatively simple functions (such as early visual processing, motor control, classical conditioning) than for more complex ones (such as complex perception and learning, language, creative problem solving). In any case, it seems likely that psychological theories could be improved by taking neuroscientific findings into account. To that end, in the next two chapters I will discuss the brain and some neuropsychological research that has implications for understanding consciousness.

#### Thought Experiment:

Imagine that this is the twenty-third century and that you are a scientist doing field work on a distant planet. You have contracted to spend three years there studying the life cycles of the primitive life forms. You have been entirely without human companionship for a year. Then one day a supply spaceship arrives from Earth. The ship has a crew of one person, an exceptionally beautiful, witty, charming person of your preferred sex. You are immediately attracted to each other. You walk, you talk, you hold hands, you become lovers, and so forth. He/she agrees to stay with you for the next two years, and sends the supply ship back to Earth on autopilot. Your spirits soar, you have a renewed enthusiasm for work and for life, since you have a new companion and lover. Then one day, while exploring your beautiful lover's sleeping body, you find something that you hadn't noticed before. Lining the inside of his/her navel there is a silver ring, and imprinted on it are the words, "Made in Japan." Your lover is an android, a synthetic human! The question is, would you still accept this "person" as a conscious, feeling being, as you did before you discovered the ring? Why did you accept him/her as a conscious being in the first place? And would you still love him/her, after you discovered the ring?

## THE OTHER-MINDS PROBLEM

The only conscious experience that we can ever directly know is our own. We cannot directly know the conscious experience of other adult humans, much less preverbal children or animals. How, then, can we know whether another being is conscious at all? This is the *other-minds problem*. It is one of the two main *epistemological* problems (having to do with the nature of knowledge) regarding mind, the other being the problem of how—and what—we can know about our own minds through introspection (P. M. Churchland 1988).

To clarify the problem we must ask what we mean by consciousness in regard to other minds. We can distinguish three aspects of conscious experience that are particularly relevant to the problem at hand: perceptual awareness, intelligent thought (particularly conceptual thought and volition—the flexible control of behavior by thought processes), and self-awareness. We can consider the other-minds problem in regard to each of these three aspects of consciousness. There are three main approaches to the other-minds problem: (1) the argument from analogy; (2) the argument from behavioral criteria; and (3) the argument from hypothetico-deductive theory.

### The Argument from Analogy

We can never directly know another being's conscious experience. The best we can do is to *infer* that another being probably is conscious, and to infer the nature of its conscious experience, based on some sort of objective evidence. The practical problem is, on what basis can we make plausible inferences about consciousness in other beings?

The simplest solution is the argument from analogy: If one creature is known to be like another creature in *some* characteristics, then the first creature may be assumed to be like the second one in *other* closely related characteristics. (A corollary of the argument is that the more similar the creatures are in some characteristics, the more similar we would expect them to be in other characteristics, too.) For example, you know that you are a human being and that most of your behavior is accompanied by conscious experiences, such as perceptions, emotional reactions, and thinking in words or images. You observe that your friend has a human body, and that his or her behavior suggests emotional reactions of anger, fear, surprise, and joy in the same situations in which you would feel these reactions. Your friend also shows complex behaviors, such as speech, learning, and problem-solving abilities, that you know to be accompanied by conscious thought in your own case. Noting that your friend's appearance and behavior are in many ways similar to your own, you infer—by analogy—that your friend has conscious experiences similar to your own.

The argument from analogy seems reasonable if we limit it to cases in which there is a high degree of similarity between ourselves and another creature in both body and behavior. In a looser application it could be extended to children or animals. For example, I could argue that my cat, Farfel, is a conscious being as follows: Disregarding the form of the body and concentrating on the brain, and taking an identity theory approach that identifies mental processes with brain processes, I say that to the degree that there

is a similarity of brain and behavior, there is likely to be a similarity of consciousness. Differences in brain and behavior imply not the absence of consciousness, but differences in consciousness. Thus, from my knowledge of cat brains and my observations of Farfel's behavior I infer that she is a conscious being, though a stupid one.

There are, however, some arguments against the argument from analogy (P. M. Churchland 1988). First, in inferring by analogy that another creature is conscious, I am trying to make an inductive generalization based on only a single case. I believe that there are certain relationships between my brain, behavior, and consciousness, and I assume that such relationships are true of all other creatures, too. Thus, I am ready to attribute consciousness to other beings whose brain and behavior are similar to mine. The problem is that I know only one consciousness, my own, and I might be wrong in assuming that my case of brain-behavior-consciousness relationships is typical. Conceivably, another creature might have a similar brain and behavior, but not be a conscious being. Perhaps it is a nonconscious automaton. Thus, it could be a mistake to infer consciousness in other beings based on their apparent similarity to myself. But what is the alternative? Perhaps *solipsism*, the belief that only one consciousness exists, namely my own. But solipsism is too arrogant a notion to be acceptable. It seems to me to be more reasonable and plausible to argue by analogy that some other beings are conscious.

Going beyond the inference *that* another being is conscious, the greater problem is to infer in any detail the *nature* of that being's conscious experience. The problem is greater, the greater the difference between that being and ourselves. Thus, a second argument against the argument from analogy is that we can reasonably infer in other minds only what we find in our own conscious experience. Presumably a color-blind person could not infer color perception in other people. In an essay titled "What Is It Like to Be a Bat?," Nagel (1979) argued that we *cannot* know, we cannot even imagine it, because bats are too different from us. However, inferences about consciousness in other beings need not be based entirely on direct analogies to our own experience. Using special instruments and behavioral tests, scientists have determined that bats can navigate by sonar, and that snakes can see infra-red wavelengths and bees can see ultra-violet, even though humans do not have those abilities. The analogy to our experience is looser here—at least we can hear echoes and see wavelengths in between infra-red and ultra-violet. This is not to deny that there are severe limits on our inferences about the nature of consciousness in other beings, but the limits may not be quite as severe as is sometimes supposed. In any case, an argument from analogy is still useful for inferring *that* another being is conscious.

The main problem with the argument from analogy is that it is too vague. It is generally applied informally to make plausible guesses about consciousness in other beings. Sometimes one sort of similarity to ourselves is emphasized, and sometimes another, depending on the situation and the creature at hand. Perhaps this is excusable, insofar as consciousness in other beings may bear only a loose family resemblance to our own. Just as children of the same parents may be similar to each other in some respects but not others—one pair having similar eyes and hair, another pair having similar face shape—different creatures in the family of conscious beings may be sim-

ilar to each other in some features of behavior and consciousness but not in others. However, in order to develop scientific theories of consciousness, it would be desirable to develop specific objective criteria for inferring consciousness that could be applied in every case.

### The Argument from Behavioral Criteria

There is no general agreement among scientists about specific objective criteria for inferring consciousness in other beings, but two general approaches can be outlined: (1) introspective verbal reports, and (2) intelligent behavior. Both of these approaches involve inference by analogy, but they attempt to be more specific than the usual "argument from analogy."

Here, the distinction of three aspects of consciousness becomes more important. There is little problem in developing criteria to infer sensory-perceptual awareness. For example, operant conditioning procedures can be used to train animals like pigeons and cats and monkeys to make different responses in the presence of different stimuli. This implies basic sensory awareness. Further, operant methods have been used to train laboratory animals to recognize simple perceptual concepts, in which they have to distinguish the similarities between pictures that differ in their details (for example, recognizing pictures of humans, as distinct from pictures of various other animal species). When the animals respond correctly to new examples of a concept (such as a picture that they haven't seen before), it implies that they have learned the perceptual concept (see review in Roitblat 1987).

The problem of developing objective criteria is more acute for intelligent thought and self-awareness. Here I will be especially concerned with the problem of intelligent thought and self-awareness in intact humans and animals. (In the next chapter I will discuss the problem of intelligent consciousness and self-awareness in the nonspeaking disconnected right hemisphere in human split-brain patients.)

**Introspective verbal reports.** The most widely accepted behavioral evidence for intelligent thought and self-awareness is the introspective verbal report (IVR): a verbal description of one's conscious experience. The IVR criterion comes from certain introspective observations, some assumptions, and an argument from analogy. First, I note (introspectively) that most of my conscious thoughts are verbal. Verbal thinking is a fundamental feature of consciousness as I know it. Therefore, I assume that *all* conscious creatures have verbal conscious thoughts. Second, I can make introspective verbal reports describing my conscious experiences. Not only can I describe my verbal thoughts, I can also describe—with varying degrees of accuracy—my nonverbal conscious experiences, such as things that I see and feel. I assume that all conscious creatures can make introspective verbal reports on their conscious experience. Therefore, a creature that cannot make an introspective verbal report is not conscious. Third, I assume that *only* a conscious being could make a verbal description of conscious experience. Therefore, if another being makes a verbal report that sounds like an introspective verbal report of conscious experience, I will accept that being as a conscious one.

Setting aside the problem of the limited accuracy of introspective ver-

bal reports, by the IVR criterion we can accept adult humans, and also children with good verbal skills, as conscious beings.

Nonetheless, there are three problems with the IVR criterion of consciousness. First, it assumes that *only* conscious beings can make verbal reports that sound like reports of conscious experience. What if a cleverly programmed computer or android made such reports? Conceivably, a computer might someday answer even your trickiest questions about its conscious experience in a manner indistinguishable from the way a human would answer such questions. Logically, by the IVR criterion we would have to accept such a computer as a conscious being. This view is consistent with the functionalist approach to the mind-body problem, though anthropocentric (human-centered) bias will make it tough for computers to gain general acceptance as conscious beings.

A second and more serious problem for the IVR criterion is the assumption that *all* conscious beings can make introspective verbal reports. If we grant, temporarily, the assumption that verbal thinking is fundamental to consciousness, there is still the possibility that some beings might have verbal thoughts but be unable to make IVRs. For example, adults who have suffered strokes on the left side of their brain, making them *aphasic* (unable to speak), could not make IVRs. Nor could young children who have learned to talk but who cannot yet understand a request to describe their conscious experience, or who don't have a rich enough vocabulary and an understanding of metaphor to describe their thoughts. Both adult stroke victims and young children may do many things that make us suspect that they are conscious beings, but by the IVR criterion we could not accept them as conscious. An IVR advocate might try to solve this problem by saying that having the *potential* for someday making IVRs, or having had the ability to make them in the past, could be accepted as tentative evidence for consciousness, but such a solution is merely an escape from the problem of specifying a clear behavioral criterion for consciousness.

The third and most serious problem with the IVR criterion is its assumption that verbal thinking is a fundamental characteristic of consciousness in all conscious creatures. This anthropocentric view would deny, *a priori*, the possibility of consciousness in nonhuman animals and preverbal children. But there are other important aspects of consciousness besides verbal thought, and there is abundant evidence that thinking without words occurs in both humans and animals (Weiskrantz 1988b).

**Intelligent behavior.** What behavioral criteria can we use as evidence for conscious thinking in beings that cannot make introspective verbal reports? Donald Griffin (1984) suggested as a general criterion "versatile adaptability of behavior to changing circumstances and challenges" (p. 37). To expand on Griffin's definition, intelligent behavior—behavior influenced by conscious thought—would be expected to show the following characteristics: (1) *Adaptiveness*: the behavior is suited to adaptive goals. It is appropriate to the organism's needs. (2) *Spontaneity*: actions are initiated autonomously by the organism, rather than being mere reflexive or conditioned responses to stimuli. (3) *Flexibility*: responses to stimuli change according to changing circumstances, rather than being mere inflexible habits. The

organism can initiate appropriate actions in novel situations. (4) *Conceptual learning*: behavior is influenced by acquired knowledge (not necessarily verbal) that can be utilized in a variety of situations. Conscious thought is more than perceptual awareness; it can represent objects and events not currently present and use knowledge gained in past experience to deal with the present situation. (5) *Anticipation*: adaptive intelligent behavior suggests that the organism anticipates at least the short-term future. The ability of conscious thought to represent things not currently present can be extended to anticipation of future needs, actions, and events.

I am suggesting that an organism that is capable of conscious thinking will show these features in some of its behavior—not necessarily in all of its behavior. The fact that the flexible behavior of animals is sometimes maladaptive or stupid is not evidence against conscious thinking (Griffin 1984). We don't deny consciousness to humans who sometimes do stupid things. Also, the fact that some adaptive behaviors of animals can be explained as simple nonconscious instincts, reflexes, or conditioned responses is not evidence against the possibility that some animal behaviors are influenced by conscious thinking.

Griffin (1981, 1984) gave numerous examples of flexible, adaptive animal behaviors that suggest conscious thinking. It is beyond the scope of this book to go into any detail on this evidence, and I refer the interested reader to Griffin's books. But to list briefly just a few of his examples: optimal nesting-territory selection by marsh-dwelling redwing blackbirds; learning to open aluminum-foil milk bottle tops in England in the 1930s by two species of birds; cooperative group hunting by lions and hyenas; dam-building by beavers; tool-using, such as chimpanzees using sticks to extract termites from a mound, and sea otters using rocks to dislodge and crack open shellfish; and imitative learning in apes and porpoises and others. Most important of all is the abundant evidence for communication among many animals by gestures and cries; for example, they can signal threat, appeasement, invitation to mating, the location of food, and warnings about predators. "Because communicative behavior, especially among social animals, often seems to convey thoughts and feelings from one animal to another, it can tell us something about animal thinking" (Griffin 1984, p. 38).

Besides the naturalistic observations by ethologists, we can add the studies by psychologists who have taught apes to communicate by sign language (such as Gardner & Gardner 1969; Miles 1983; Terrace 1979). Chimpanzees, gorillas, and an orangutan have learned to make simple requests and answer simple questions by signs, and they have shown some degree of inventiveness in making new combinations of signs. Whether the communication accomplishments of apes really deserve to be called a language in the human sense—with novel ideas expressed in grammatically correct sentences—is a matter of controversy (see Terrace, Petitto, & Bever 1979). Nonetheless, the evidence from these studies strongly suggests that the apes are consciously thinking when they use sign language.

Griffin (1984) suggested that there would probably be a lot more evidence for conscious thinking in animals if ethologists had been looking for it in their field studies. But such evidence has often been ignored because most ethologists do not believe that animals are capable of conscious think-

ing. Ethologists and behavioral ecologists have tried to explain most natural animal behaviors in terms of blind instincts—behaviors programmed mainly through heredity rather than through specific learning experiences. Behaviorist comparative psychologists have tried to explain complex animal behaviors in terms of conditioned reflexes and chains of conditioned operant responses. Both groups of researchers have denied that animals do any conscious thinking.

Historically, the denial of animal consciousness derives from two ideas. The first is *anthropocentrism*, the belief that humans are the most significant beings in the universe, which carries with it the attitude that humans must be absolutely unique in some important way. The denial of the possibility of animal consciousness reflects the same sort of anthropocentric thinking that led to Descartes' dualist philosophy. The second is the principle of parsimony and the denial of anthropomorphism. In the early 1900s there was a reaction by scientists against some writers' over-indulgence in *anthropomorphism*, the attribution of human-like mental processes to animals. Attributing mental processes to animals was thought to be against the principle of parsimony (using the simplest explanation) when animal behavior could be explained in mechanistic terms. But as Griffin (1984) suggested, our increased knowledge of the complexity and flexibility of animal behaviors now makes it seem more parsimonious to suggest that at least some behaviors of some animals are controlled with the help of conscious thinking. The alternative—that animals are nonconscious automatons whose flexible behavior in changing circumstances is due entirely to genetic-preprogramming of behavior and/or complex conditioned-response chains—is becoming harder and harder to believe. Griffin went so far as to suggest that not only behaviors based on learning, but also those with a large hereditary or instinctive component may be accompanied by conscious awareness and thinking. He proposed a "cognitive ethology," a science of mental processes in animals. His proposal is highly controversial among animal behavior researchers, but it is consistent with the liberal attitude toward consciousness that I am advocating in this book (see Griffin 1978 and accompanying commentaries).

**Self-awareness.** Earlier I defined self-awareness in terms of an awareness of one's individuality and having a self-concept, including knowledge of one's appearance, one's abilities, and one's personal history. It would seem that a conscious animal must be aware that most of its actions are undertaken to serve its own needs. However, it is hard to conceive what sort of specific evidence might indicate that an animal had self-awareness. Griffin (1984) cited some evidence for a rudimentary self-awareness in some animals, limited to awareness of its own physical appearance, and knowledge of how it appears to others. For example, it has been suggested that a predator's attempt to conceal its body behind bushes as it sneaks up on prey suggests some degree of self-awareness, as does the prey's attempt at self-concealment. Gallup (1977) showed that chimpanzees can learn to recognize themselves in mirrors. He first let the chimpanzees become familiar with their own mirror image, then he marked their faces with rouge while they were under anesthesia. The chimps subsequently showed startled reactions when they recog-

nized their altered appearance in the mirror. Other attempts to show mirror self-recognition in animals have been unsuccessful more often than not, though it is not clear whether the failures were due to methodological problems or the animals' limitations. In any case, mirror self-recognition is an unnatural task, and it is conceivable that some animals might have a rudimentary self-awareness without necessarily being capable of mirror self-recognition. On the other hand, it seems likely that a well-developed self-concept depends critically on reflective-awareness and linguistic thinking—capacities well-developed in humans, but little or none in even the most intelligent of nonhuman animals. In any case, the problem of finding a satisfactory objective criterion for self-awareness in the absence of introspective verbal reports remains unsolved.

### **The Argument from Hypothetico-Deductive Theory**

A third approach to the other-minds problem is to use conscious mental states (beliefs, feelings, volition, and so forth) as hypothetical explanatory constructs in a theory about how the mind works. From theoretical assumptions about how the hypothetical mental states work, deductions can be made to predict the organism's behavior. To the extent that the predictions about the organism's behavior are successful, it suggests that the organism has the mental states specified by the theory.

P. M. Churchland (1988) argued that this is exactly what we do in "folk psychology" theories, that is, our everyday assumptions that we use to explain why people do what they do, and to predict what they will do next. For example, I believe that when people are insulted they feel angry, and they are likely to retaliate if they think they can get away with it. Folk psychological theory is useful. For example, it leads me to predict that if I insult my boss, then he will retaliate against me. Bandura (1986) developed a formal social-cognitive theory that uses mental states as constructs to explain human behavior. Formal psychological theories are more successful than folk psychology theories at predicting human behavior.

In principle, one could develop mentalist theories to explain the behavior of nonhuman animals, such as dogs, horses, and chimpanzees. The rules of the game are the same as for mentalist theories of human behavior. You would describe certain mental states and processes and explain how they interact with each other and with the immediate situation to produce the animal's behavior. The animal's hypothetical mental states might be quite different from those of humans. To the extent that the theory succeeded in predicting the animal's behavior, the argument for the existence of the hypothetical mental states would be supported. The argument from hypothetico-deductive theory does not depend upon introspection or an argument from analogy. Though people sometimes apply informal, folk psychology theories to familiar animals, psychologists have not developed elaborate mentalist theories for animals because they are not convinced that such theories would be any better than the mechanistic theories of behaviorism. It remains to be seen whether this attitude will change, in view of the renewed interest in cognitive processes in animals (Burghardt 1985).



### Concluding Comment

The other-minds problem is difficult, but it will not disappear. People who live or work closely with animals recognize that apes and horses, dogs and cats, are conscious beings, though their consciousness is very different from our own. I would argue that our recognition of consciousness in other beings is not a mere projection of human-like attributes onto unconscious automatons. Rather we, as conscious beings, can *intuitively* infer consciousness in other conscious beings. Future theoretical and methodological developments will lead to the development of more objective, scientifically acceptable means of recognizing consciousness in other beings. Progress on this topic might be aided by a better understanding of intuitive inference processes.

In the next chapter we will encounter a special case of the other-minds problem, the question of dual consciousness in split-brain patients. How can we determine whether the nonspeaking, disconnected right hemisphere is conscious in its own right, independently of the speaking left hemisphere?

### SUMMARY

The *mind-body problem* asks what is the relationship between the mind (or consciousness) and the body (or brain). The two major positions are dualism and materialism. *Interactionist dualism* (Cartesian dualism) holds that mind and body are made of different substances: the body is material but the mind is some immaterial soul stuff, and the mind interacts with the body to control human behavior. Parapsychological phenomena (clairvoyance, telepathy, precognition) have been offered in support for dualism, but their reality status is a matter of controversy. Out-of-body and near-death experiences have also been offered in support of dualism, but alternative, naturalistic explanations of these experiences are available. Dualism's strongest card is the argument from introspection combined with the irreducibility argument: introspectively it seems that conscious experience is quite different from brain processes, and therefore consciousness cannot be a product of brain processes. But introspection is limited; it cannot reveal the true nature of things.

Materialism is the view that mind and body are inseparable: mental events are produced by brain events. Four varieties of materialism were discussed. *Epiphenomenalism* is the view that consciousness is a side-effect of brain activity but it has no role in controlling behavior. *Identity theory* says that mental events are identical with brain events; they are different viewpoints on the same events. For each mental event there is a corresponding brain event. The principle of causal isomorphism says that the causal relationships between successive mental events are functionally equivalent to the causal relationships between corresponding physical (brain) events. Mental events have causal efficacy in that they are identical with brain events that have causal efficacy. *Emergent interactionism* is the hypothesis that consciousness is an emergent phenomenon: it is produced by brain processes, but it has holistic properties of its own and it exerts downward control on brain processes. Emergent interactionism differs from identity theory by

claiming that consciousness cannot be fully understood by reductionist analysis of brain processes; its holistic or field properties must be considered. *Functionalism* is the view that the functional characteristics of mental processes (what they do) is their critical feature, and it doesn't make any difference whether the physical substrate is a brain or a computer.

The *other-minds problem* is the question of how we can recognize consciousness in other beings. Three approaches were discussed. The *argument from analogy* says that we infer consciousness from similarities of structure and function between other beings and ourselves. The closely related *argument from behavioral criteria* requires that specific behavioral criteria for consciousness be developed. Introspective verbal reports on conscious experience are widely accepted as evidence of consciousness, but they are limited to adult humans and verbally skilled children. Several signs of intelligent behavior were suggested as a criterion for recognizing consciousness in animals. These criteria are controversial because some scientists argue that "intelligent" behaviors might be produced without consciousness. Finally, the *argument from hypothetico-deductive theory* suggests that we develop mentalist theories to explain behavior. When a theory successfully predicts behavior then the argument for the actual existence of the mental processes specified by the theory is supported.

## ENDNOTES

<sup>1</sup>To be more precise, I am talking here about *substance* dualism. Another type of dualism is *property* dualism, the view that while there is only one type of substance (physical substance), the brain has special properties (mental properties) that are possessed by no other type of physical object. In property dualism it is assumed that mental properties, such as conscious sensations, thoughts, and feelings, cannot be reduced to or explained solely in terms of the physical properties of the brain (P. M. Churchland 1988).

<sup>2</sup>Epiphenomenalism is sometimes classified as a type of *property* dualism (see endnote 1) because of its claim that while consciousness is produced by the brain, it has the special property—unlike other brain events—of playing no further role in the cause-and-effect processes of the brain. However, I classify epiphenomenalism as a variety of materialism because of my emphasis on the source (or substance) of consciousness, where epiphenomenalism acknowledges that consciousness is a product of brain activity and cannot exist without the brain. Epiphenomenalism has a closer kinship with the materialist identity theory than it does with Cartesian dualism.

<sup>3</sup>By "mental state" or "conscious state" I mean here the relevant set of conscious contents at a particular moment, including percepts, feelings, inner speech, and so on.