

Chapter 1

Mind–Body Theories and Mind–Body Problems

Overview

Mind–body theories and mind–body problems form the core subject-matter of philosophy of mind. Mind–body theories offer different ways of understanding how mental and physical phenomena are related. They are divided into two broad categories: monistic theories and dualistic theories. Monistic theories claim that there is fundamentally one kind of thing. Physical monism or physicalism, as it is usually called, claims that everything is physical; everything can be exhaustively described and explained by physics. Mental monism, which is typically called ‘idealism’, claims that everything is mental – everything can be exhaustively described and explained using our prescientific psychological concepts. Finally, neutral monism claims that everything is neutral; everything can be exhaustively described and explained using a conceptual framework that is neither mental nor physical but neutral.

Unlike monistic theories, dualistic theories deny that a single conceptual framework is sufficient to describe and explain everything. Rather, a complete description and explanation of everything requires that we use both the mental and the physical conceptual frameworks. There are, then, two fundamentally distinct kinds of properties individuals can have: mental properties, which are expressed by the predicates of the mental framework, and physical properties, which are expressed by the predicates of the physical framework. Among dualistic theories, dual-attribute theories claim that the very same individual can have both mental and physical properties. Substance dualistic theories deny this. The very same individual cannot have both mental properties and physical properties, they claim. According to substance dualists, mental beings such as you and I have no physical properties at all, and physical beings such as human bodies have no mental

properties. This implies that there are not only two fundamentally distinct kinds of properties, but also two fundamentally distinct kinds of individuals as well: those with exclusively mental properties, and those with exclusively physical properties.

In addition to the foregoing theories, there are three others that fall outside the main classification of monistic and dualistic theories. Instrumentalism falls outside the classification because it denies a realist understanding of psychological discourse. Mind–body pessimism falls outside the classification because it denies the possibility of giving a completely satisfactory account of how mental and physical phenomena are related, and hylomorphism falls outside the standard classification because it denies that human behavior can be described accurately in terms of a mental–physical distinction.

Mind–body problems have two features in common: the distinction between mental phenomena and physical phenomena, and premises that make it difficult to understand how mental and physical phenomena are related. The problem of other minds is an example. It makes it difficult to understand how it is possible for us to know what other people are thinking or feeling based on our knowledge of their bodily behavior. The problem of psychophysical emergence, on the other hand, makes it difficult to see how it is possible for mental phenomena to exist at all if the world is fundamentally physical, and the problem of mental causation makes it difficult to see how mental and physical phenomena can interact in the ways they appear to.

1.1 Mind and Brain

The surgeon removed the section of skull and cut through the dura mater revealing the brain beneath. It pulsed gently in sync with the patient's heartbeat. He was 12-year-old R. W. (name concealed for privacy). He'd had a difficult birth but had otherwise developed normally until the seizures began. Three years of failed treatments and months of tortured deliberation had brought him and his parents to this point. Doctors were going to remove part of his brain – in theory, the part responsible for his seizures. The difficulty was identifying exactly what part that was and removing it without damaging the surrounding tissue and with it R. W.'s ability to speak or laugh, to recognize faces or remember facts, to play the piano or smell cookies baking in the oven.

R. W. received a local anesthetic as they cut through his scalp, and was mildly sedated now, but was otherwise awake and alert. The lead surgeon began touching one section of brain tissue after another with two metal probes that carried an electric current. Based on R. W.'s symptoms he guessed this was where the seizures were originating. They always began the same way: an experience of colored

triangles – like the afterimages of bright lights only clearer. R. W. would then become confused about his surroundings, and see men moving toward him with guns. Those who saw R. W. during one of these episodes could hear the terror in his voice, and see it on his face as his eyes and head moved from right to left, following, it seemed, the movements of the men across the room.

As the surgeon now touched sections of R. W.'s brain he observed R. W.'s behavior carefully, and asked that R. W. describe any changes he experienced. After stimulating one area in particular R. W. said with astonishment, "Oh, gee, gosh, robbers are coming at me with guns!" A few moments later the stimulation was repeated, "Yes, the robbers, they are coming after me ... Oh gosh! There they are, my brother is there. He is aiming an air rifle at me." R. W.'s eyes moved slowly to the left ...¹

The foregoing story describes a real operation performed by the neurosurgeon Wilder Penfield (1891–1976). Penfield did pioneering work mapping functional areas of the brain using electrical stimulation in an effort to treat patients like R. W. He kept detailed records of his observations. Another of Penfield's cases involved a 32-year-old woman, A. Bra., who began having seizures a year earlier. Penfield's notes report the effects of stimulating various numbered areas of her right temporal lobe:

15. "I hear singing."
15. Repeated. "Yes, it is White Christmas." When asked if anyone was singing, she said, "Yes, a choir." When asked if she remembered it being sung with a choir, she said she thought so.
16. "That is different, a voice – talking – a man."
17. "Yes, I have heard it before. A man's voice – talking."
18. "There is the sound again – like a radio program – a man talking." She said it was like a play, the same voice as before.
19. "The play again!" Then she began to hum. When asked what she was humming, she said she did not know, it was what she heard.
19. Repeated. Patient began to hum. She continued at the ordinary pace of a song. "I know it but I don't know the name – I have heard it before. I hear it, it is an instrument – just one." She thought it was a violin.
26. Patient said, "It hurts." Stimulation was stopped. She said, "I see a picture." She added, "It was a face which comes from a picture."
27. "The same thing. The play and they are banging on something like a drum."
28. "I see people walking."²

The effects Penfield produced are familiar to students of neuroscience. Electrical stimulation of the cortex can cause patients to move their limbs, to sense numbness or tingling on the skin, to experience flashes of light or buzzing sensations, to feel fear, experience *déjà vu*, or have a sense that they are in a dream.³ It can also inhibit functioning: in a dramatic demonstration reported on the front page

of the *New York Times*, for instance, the neuroscientist Jose Delgado stopped a charging bull in its tracks with the push of a button.⁴

Penfield's observations are interesting for many reasons, not least for the philosophical questions they raise. What, for instance, is the relationship between mental phenomena and physical phenomena? What was the relationship between, say, the activation of cells in R. W.'s temporal lobe and his visual experience of seeing robbers? The cells were tiny components that operated according to simple mechanical principles; they were located inside R. W.'s skull; they had mass, volume, and all the other properties physical things have. R. W.'s experience, on the other hand, did not appear to have these properties. It did not appear to be tiny, and the figures he saw seemed as large as ordinary people. Likewise, the experience did not appear to be a mechanical process located inside his skull, but a qualitative awareness in the surrounding room. Nor is it evident that the experience had mass or volume. How after all could we have weighed or measured it? We know how to weigh and measure the brain cells, and we would know how to weigh and measure the robbers if they had existed, but how could we have weighed or measured the experience itself? The experience and the brain cells seem very different, and yet they were obviously intimately related. But how?

Philosophers, scientists, and others disagree about the answer. Some, for instance, would claim that R. W.'s experience was identical to the activity of his brain cells – that the experience and the brain activity were the very same thing described using two different vocabularies. When using an informed, scientific vocabulary, we would call the event in R. W.'s skull 'temporal lobe activity', but when using an ordinary, prescientific vocabulary, we would call it 'seeing robbers approaching with guns'. Other philosophers would deny that R. W.'s experience was identical to the activity of his brain cells. Experiences are not the same thing as brain activity, they would say; experiences are not physical events at all but nonphysical events caused by brain activity. Yet others would take this type of answer a step further: not only are experiences nonphysical, people are too. You, R. W., and I are nonphysical entities that are intimately connected to human bodies. (Incidentally, this is the kind of answer Penfield himself preferred.) Still other philosophers would claim that the question is ill-posed, that it is a mistake to ask how mental and physical phenomena are related since it is a mistake to describe human experience in dichotomous mental and physical terms to begin with.

These answers represent different mind–body theories. Mind–body theories and the problems they try to solve form the subject-matter of philosophy of mind. We will consider them in detail in the chapters that follow: what they claim, why people believe them, what implications they have for our understanding of human life, and most importantly, what reasons we have for thinking they are true or false. We will begin with a brief overview of the main options and some of the problems they attempt to solve.

1.2 Mind–Body Theories

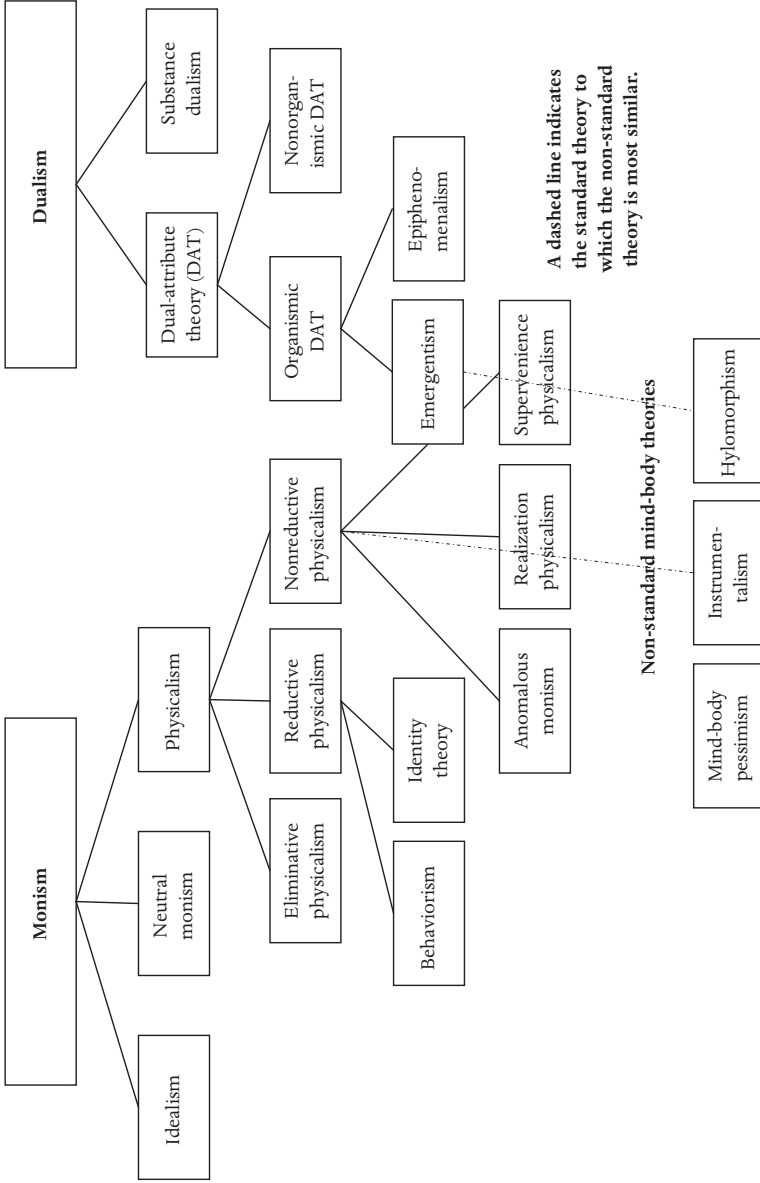
Mind–body theories offer different ways of understanding how mental and physical phenomena are related. There are two broad categories of mind–body theories: monistic and dualistic. Monistic theories claim there is fundamentally one kind of thing; dualistic theories, that there are fundamentally two kinds of things. The divisions among monistic and dualistic theories of mind are represented in Figures 1.1 and 1.2. The first figure shows the divisions among mind–body theories; the second depicts those differences in an intuitive way.

Monistic mind–body theories are of three broad types. **Physical monism** or **physicalism** claims that everything is physical; everything can be exhaustively described and explained by physics. **Mental monism**, on the other hand, which is commonly called **idealism**, claims that everything is mental; everything can be exhaustively described and explained using mentalistic concepts such as *belief*, *desire*, and *feeling*. Finally, **neutral monism** claims that everything is neutral; everything can be exhaustively described and explained using a conceptual framework that is neither mentalistic nor physicalistic but neutral. All monistic theories, then, are committed to the claim that there is fundamentally one kind of thing; they differ over what that one kind of thing is: mental, physical, or neutral.

Dualistic theories, by contrast, deny that there is only one kind of thing. Fundamentally, there are two kinds of things, they claim: mental and physical. Dualistic theories are further subdivided into two broad categories. All of them claim there are two distinct kinds of **properties** or characteristics things can have, mental properties and physical properties, but they differ over the kinds of individuals that have them. **Dual-attribute theories** claim that the very same individual can have both mental and physical properties – that mental and physical properties can coincide in a single individual. **Substance dualistic theories** deny this. The same individual, they say, cannot have both mental and physical properties. According to substance dualists, then, there are not just two distinct kinds of properties; there are also two distinct kinds of individuals: individuals that have only mental properties, and individuals that have only physical properties.

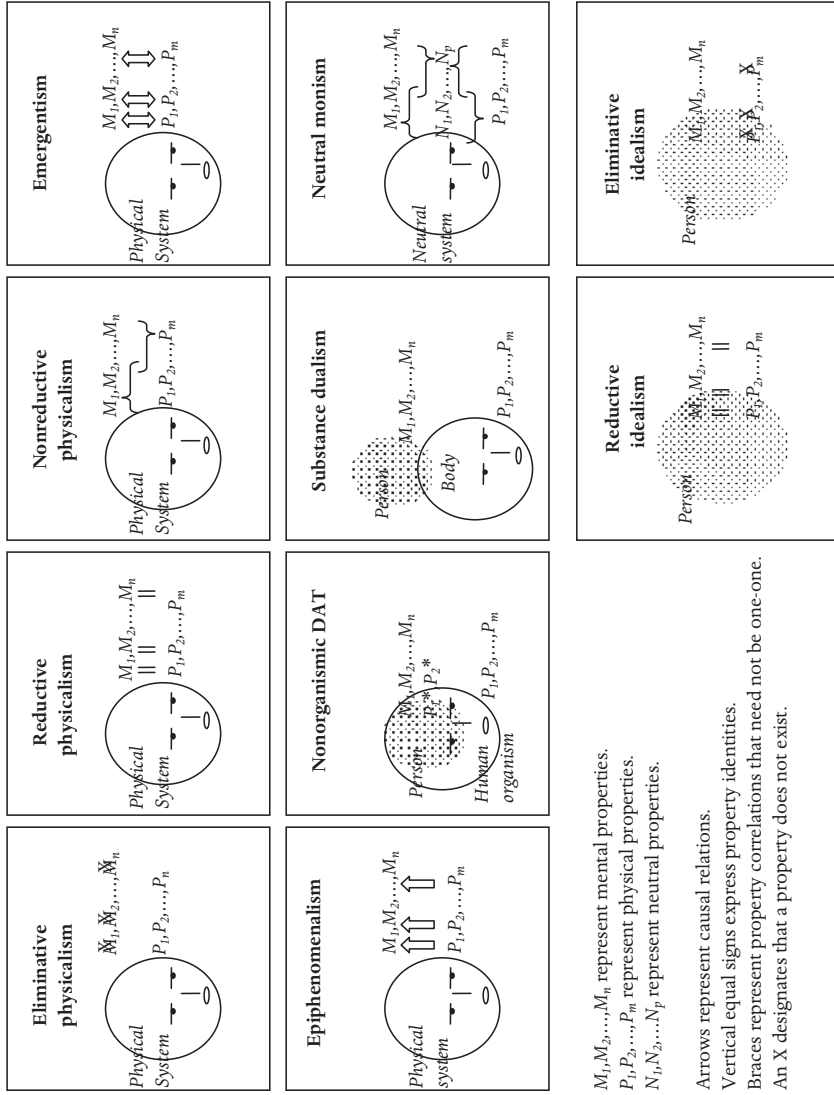
Most of the mind–body theories depicted in Figures 1.1 and 1.2 start with the same picture of the physical universe – a picture of a vast undifferentiated sea of matter and energy, an ocean of fundamental physical particles or materials governed by laws that are described by or will be described by our best physics. Physicalism claims that this is a complete picture of everything; there is nothing but this vast physical sea. Physics, they say, gives us the exhaustive description and explanation of everything that exists: of all the individuals, all their properties, and all the principles governing their behavior. According to most physicalists, however, we can describe these individuals, their properties, and behavior in many different ways. Instead of describing individual electrons or quarks or other fundamental physical particles, for instance, we can describe collections of those

Standard mind–body theories



A dashed line indicates the standard theory to which the non-standard theory is most similar.

Figure 1.1 Standard and non-standard mind–body theories



M_1, M_2, \dots, M_n represent mental properties.
 P_1, P_2, \dots, P_m represent physical properties.
 N_1, N_2, \dots, N_p represent neutral properties.

Arrows represent causal relations.

Vertical equal signs express property identities.

Braces represent property correlations that need not be one-one.

An X designates that a property does not exist.

Figure 1.2 Standard mind-body theories

particles such as tables and chairs or plants and animals. When we use terms such as ‘table’ or ‘human’, however, we are not describing entities in addition to those postulated by physics. The table or the human is not an entity in addition to the fundamental physical particles that compose it. Terms such as ‘table’ and ‘human’ are terms we use to refer to collections of particles – they are analogous to terms such as ‘team’. A team is not an entity in addition to the individual members of the team; the term ‘team’ is rather a way of referring to the team’s individual members; it is a term of collective reference. According to many physicalists, the proper names of humans such as ‘Alexander’ are analogous to the proper names of teams such as ‘The New Jersey Devils’: they refer to collections of particles if they refer to anything at all. Moreover, predicates such as ‘is alive’, ‘is in pain’, ‘believes that there are eight planets in our solar system’ do not express properties in addition to those described by physics; they instead express very complex relations among collections of fundamental particles, just as the predicates ‘is solid’ and ‘is liquid’ do. Hence, when we speak of Alexander’s being alive or experiencing pain or having a belief, we are really expressing a very complex relation among a large collection of fundamental physical particles.

Dual-attribute theories start with the same picture of the physical universe, but they disagree with physicalists about the descriptive and explanatory scope of physics. Physics does not provide an exhaustive description of all individuals and properties, they say, nor is it able to provide an exhaustive explanation of every individual’s behavior. Some individuals – people, for instance – have properties in addition to those physics describes and explains. These properties can only be described and explained using different conceptual resources such as those of psychological discourse. Psychological predicates such as ‘is in pain’ or ‘believes that there are eight planets in our solar system’ express these properties – non-physical properties, ones different from those described by physics. To say that Alexander is in pain or has a belief is not to describe a very complex relation among fundamental physical particles, say dual-attribute theorists; it is instead to express Alexander’s possession of a unique kind of property distinct from any of those possessed by fundamental physical particles. Dual-attribute theories claim, in other words, that there are two distinct kinds of properties or attributes (hence the name ‘dual-attribute theory’), and because there are two distinct kinds of properties or attributes, we need to describe things using both a mental vocabulary and a physical vocabulary.

The most popular dual-attribute theories in recent years have been forms of **epiphenomenalism** and **emergentism**. Both claim that mental properties are produced or caused by physical occurrences. Fundamental physical interactions of the sort described by physics cause or give rise to nonphysical properties – including mental properties such as belief, desire, and pain. Epiphenomenalists and emergentists differ over the causal power of emergent properties – whether those properties are able to exert a causal influence on the physical interactions from which they emerge. Epiphenomenalists say they cannot. According to them,

emergent properties, including mental properties, are causally inert; they have no causal power in their own right; they can influence nothing that happens in the universe. They do exist, say epiphenomenalists, and for that reason a complete account of the universe must include a description of them using a vocabulary suited to the task – a mental vocabulary, for instance. But emergent properties are merely causal byproducts of physical processes; they themselves cause or produce nothing. Emergentists disagree with epiphenomenalists about the causal status of mental properties. Mental properties are not causally inert, they say; rather, mental properties have causal powers distinct from those described by physics, and they make a distinctive causal contribution to the flow of physical events.

Dual-attribute theorists and physicalists all claim that any individuals having mental properties also have physical properties. Their views are all compatible, for instance, with the claim that you and I are human organisms – physical beings of a particular sort. But substance dualists disagree. According to them, persons – mental beings – are completely nonphysical. Whereas physicalism implies that physics can describe *all* of my properties, and dual-attribute theory implies that physics can describe *some* of my properties, substance dualism implies that physics can describe *none* of my properties. According to substance dualists, you and I are completely nonphysical entities; we have no physical properties at all. The only properties we have are mental ones: beliefs, desires, hopes, joys, fears, loves, and so forth. According to substance dualists, therefore, you and I are not human organisms. We are not, for instance, the humans we see in the mirror when we shave, or fix our hair, or put on makeup. We are not physical entities of any sort. We might be connected in some way to human organisms, and because of that connection you might take special interest in the appearance or the reproductive destiny of a particular human organism, but you are not that organism. In addition to the universe described by physics, then, substance dualists claim that there is another nonphysical universe described by psychological discourse. We use mental predicates and terms to describe and explain what happens in that nonphysical domain. Nonorganismic dual-attribute theories are similar in spirit to substance dualism. Like substance dualists, nonorganismic dual-attribute theorists deny that we are organisms; they nevertheless claim that we have some physical properties such as spatial location even if we do not have all the physical properties organisms do.

Physicalism, dual-attribute theory, and substance dualism are the most popular mind–body theories, but there are others as well. Idealism is like a reverse image of physicalism: just as physicalism claims that everything is physical, idealism claims that everything is mental. According to most idealists, when we take ourselves to refer to physical entities and properties we are really referring to experiences. When I talk about this table, for instance, I am not really describing an entity that is distinct from my experiences – an entity that could exist in the absence of me or someone else perceiving it. I am instead describing an expanse

of color in my visual field, together with a sense of three-dimensional depth, a feeling of solidity, a sensation of texture, and so on. Likewise, when I describe what I take to be a physical property such as solidity or mass, I am not describing a feature that something might possess independent of an experience; I am rather describing one of my experiences using a nonmental vocabulary. When I say that this table is solid, I am really saying that I am having an experience of my hand not passing through it the way it passes through the air. Likewise, when I say the table is heavy, I am saying that I have an experience of anticipating that I would experience effort or exertion if I had the experience of trying to lift it – the experience of trying to lift it, in other words, would be accompanied by an experience of meeting resistance. According to most idealists, then, calling something solid or heavy is really just a way of talking about my experiences, and the same is true of all physical things: our mental and physical vocabularies are just different ways of describing the same phenomena – phenomena that are all mental; there is no mind-independent domain described by physics.

By contrast with the foregoing theories, neutral monism claims that the universe is fundamentally neither mental nor physical. The universe consists, rather, of individuals, properties, and events that can be exhaustively described and explained using a conceptual framework that is neither mental nor physical but neutral. Our mental and physical vocabularies are just ways of expressing properties or events that are neutral.

In addition to the foregoing theories, three require special consideration: **instrumentalism**, **hylomorphism**, and **mind–body pessimism**. These theories are not included among the standard theories depicted in the top portion of Figure 1.1. The reason is that the classification of standard theories is based on three assumptions, and instrumentalism, hylomorphism, and mind–body pessimism each rejects one of them.

First, the standard theories in Figure 1.1 are committed to a **realist** understanding of psychological discourse; they all claim that the predicates of psychological discourse are supposed to express real properties. When we say that Eleanor enjoys the taste of sushi, for instance, realists claim that we are trying to express the possession of a real property by a real individual. Even eliminative physicalists, who deny that mental properties exist, are realists in this sense. They claim that predicates like ‘enjoys the taste of sushi’ fail to express real properties because there are no such properties. They nevertheless agree that when we use psychological predicates we are at least *trying* to express real properties. Yet this is what instrumentalists deny. Psychological discourse does not aim at expressing real properties, instrumentalists say. Psychological discourse is a mere tool or instrument we use for predicting human behavior. When we describe and explain human behavior using psychological predicates and terms we are concerned not with getting an accurate picture of reality, but simply with making useful predictions of what people will do. Consequently, although there are beliefs, desires, and other mental states, according to instrumentalists, this claim does not carry an

ontological commitment as strong as realists suppose. To claim that there are beliefs and desires is simply to claim that it is useful to describe and explain people's behavior using the predicates 'believes' and 'desires'.

Second, the standard theories represented in Figure 1.1 are all committed to the mental–physical distinction. They disagree about whether mental phenomena are identical to physical phenomena: dualists claim that psychological language and physical language express two different kinds of phenomena, and monists deny this. But monists and dualists alike agree that there are two vocabularies or conceptual frameworks for describing and explaining human behavior: a mental vocabulary and a physical one. Hylomorphists reject this claim. Human behavior can only be adequately described and explained, they say, using a unique vocabulary that is neither mental nor physical but that shares features of both. Neutral monists say something similar, but, unlike neutral monists, hylomorphists reject monism. They deny that there is a single conceptual framework – mental, physical, or neutral – that is sufficient to describe and explain everything that exists. The distinctive vocabulary we use to describe and explain human behavior, for instance, can be used to describe the behavior of other living things only by drawing analogies with human behavior. Other living things have their own distinctive structures and patterns of behavior, however, and because of that we have to use descriptive and explanatory resources that are suited to them if we want to give fully accurate descriptions and explanations of their behavior. Because hylomorphists reject the mental–physical distinction, hylomorphism does not fit into the standard classification of mind–body theories.

Finally, exponents of the standard theories in Figure 1.1 are all committed to the idea that it is possible to give a satisfactory account of how mental and physical phenomena are related. Mind–body pessimists reject this assumption. They claim that it is impossible to give a satisfactory account of mind–body relations because there are inbuilt limitations on human cognitive capacities that will prevent us from ever understanding how mental and physical phenomena are related. There may be a coherent account of mental–physical relations; there may even be entities in the universe whose minds are powerful enough to grasp what those relations are, but our minds are not. Our cognitive powers are limited in such a way that we will never be able to solve **mind–body problems**.

Now that we have surveyed some mind–body theories, let us consider some mind–body problems.

1.3 Mind–Body Problems

Mind–body problems arise when we try to understand how thought, feeling, perception, action, and other mental phenomena are related to events in the human nervous system. In our day-to-day dealings we take ourselves to be free

beings who act as we do because we have beliefs, desires, hopes, and fears. We describe ourselves as beings who experience joy and sadness, love and hate, pain and pleasure; beings who act to get what we want, who make choices, and who can be held accountable for the choices we make, whose actions, habits, and character traits can be evaluated as good or bad, right or wrong. In our scientific dealings, however, we see the universe as a vast sea of matter and energy that at a fundamental level has none of the features we recognize in our day-to-day lives. At the level of fundamental physics there are no differences among humans, rocks, trees, and other living things. All of them are made of the same basic materials, and at the level of those materials there is nothing that distinguishes you from a rock or a dog. The subatomic particles we find in you are the same as the subatomic particles we find in them, and those particles behave in you and in them in exactly the same ways. From the standpoint of fundamental physics, then, the familiar objects of ordinary experience are just so many collections of the same kinds of microscopic particles – particles that have none of the features we take to distinguish people from other things. Electrons and quarks do not have beliefs or desires, hopes or fears; they do not want things or deliberate about how to get them. They are not free; they do not choose to act. Their behavior is not subject to moral praise or blame, nor do they develop personalities, form character traits or habits, or experience love or hate, sadness or joy.

We thus confront two images of human life: the everyday, prescientific image of ourselves as free, rational beings with mental and moral lives, and the scientific image of ourselves as complex biochemical systems. Understanding how there can be free, mental, moral beings in a universe that at a fundamental physical level has none of these features is one of the principal issues that philosophers, scientists, theologians, and others have struggled with for the past 350 years. It has been the basis of some of the major problems of modern philosophy including the **problem of free will and determinism**, the problems generated by the fact–value dichotomy, and mind–body problems. All of these problems originate in the disparity between the way we describe the world scientifically and the way we describe it in our everyday dealings. The astronomer and physicist Sir Arthur Eddington (1882–1944) once illustrated the disparity in a memorable way:

I have settled down to the task of writing these lectures and have drawn up my chairs to my two tables. Two tables! Yes ... One of them has been familiar to me from earliest years. It is a commonplace object of that environment which I call the world ... It has extension; it is comparatively permanent; it is colored ... Table No. 2 is my scientific table ... [T]here is a vast difference between [it] ... and the table of everyday conception ... It does not belong to the world previously mentioned – that world which spontaneously appears around me when I open my eyes ... My scientific table is mostly emptiness. Sparsely scattered in that emptiness are numerous electric charges rushing about with great speed ... [M]odern physics has by delicate test and remorseless logic assured me that my second scientific table is

the only one which is really there ... On the other hand ... modern physics will never succeed in exorcising that first table ... which lies visible to my eyes and tangible to my grasp.⁵

Eddington describes a tension between two descriptions of the world: a scientific description, and a commonsense, prescientific one. Only one, it seems, can lay claim to describing the real table. There is, in other words, a single descriptive role to be filled, the role labeled 'real', and only one description can occupy it. Consequently, if we accept the description offered by science, we must reject the description offered by common sense, and if we accept the description offered by common sense, we must reject the description offered by science. The problem is that we do not want to reject either description, and we have good reason to think both are true. As a result, we have difficulty understanding how the scientific description is related to the prescientific one.

Mind–body problems have a similar structure. We have two frameworks for describing and explaining human behavior: a scientific framework and a commonsense, prescientific one. Each has conceptual resources that seem fully adequate for describing and explaining human thought, feeling, and action. Consider an example: an action such as raising your arm over your head. Go ahead and try it. Notice that we can explain your action in two different ways. We can explain it scientifically by appeal to the contractions of muscles in your shoulder and the activity of neurons in a particular region of your brain. Yet we can also explain your action mentally by appeal to your desires and beliefs – a desire, for instance, to understand mind–body problems and a belief that raising your arm might help. What is true of action is also true of perception. We can describe your current visual experience in terms of its qualitative features – a series of small black shapes on a white background, say. But we can also describe it in terms of the states of the neurons in your eyes and brain, and the atomic structure of the paper and ink that reflect light to them. Science and common sense each provide resources that seem fully adequate for describing and explaining our actions and experiences; each purports to satisfy our requests for information; each purports to reveal the reasons why people act and perceive as they do. But if there can be only one such reason – the *real* reason, as we might refer to it – then science and common sense look like rivals competing to occupy a single explanatory role, and we face the vexing task of judging between them; we face a mind–body problem. Consider another example: the **problem of psychophysical emergence**.

1.4 The Problem of Psychophysical Emergence

Life and mind did not always exist in the universe. Early in the universe's history there were not even atoms since energy levels were too high to allow protons and

electrons to form stable pairs. Life and consciousness are thus relative newcomers on the cosmic scene. To many people, this suggests that the physical conditions that existed before their emergence had to be responsible in some way for bringing them about. Scientists are becoming increasingly clear about the physical conditions that were responsible for the emergence of life, but consciousness is a different story.

We are conscious beings: we have experiences. Yet we are composed entirely of nonconscious parts: molecules, atoms, and other microscopic entities described by physics. These microscopic entities are not conscious the way we are. How, then, do our conscious experiences emerge out of these nonconscious physical interactions? It is difficult to see how they could. Consider the difference between the rich, colorful, quality-laden features of conscious experiences and basic properties of matter – the difference, for instance, between R. W.’s visual experiences and the cells in his brain, or between your current visual, auditory, and other experiences, and the mass of an electron. How could collisions among a number of nonconscious subatomic particles combine to produce something as rich and varied as your current awareness of the various colors, sounds, smells, and textures that surround you? If the movements of some number N of fundamental physical particles do not constitute a conscious experience, it is implausible to suppose $N+1$ particles could. What difference, after all, could one particle make to whether or not something is conscious? Surely an individual particle does not have the power magically to produce consciousness. Consequently, if N particles are insufficient to produce consciousness, it looks like $N+1$ particles will be insufficient as well, and, in that case, it looks like no number of subatomic collisions will be sufficient to produce conscious experiences. Why? Well, we can agree that the movement of just one subatomic particle is insufficient to produce consciousness; after all, if the movement of just one subatomic particle were sufficient to produce consciousness, then consciousness would probably not have emerged as late in the universe’s history as it did, for subatomic particles existed almost from the beginning. We also just agreed, however, that if the movement of one subatomic particle is insufficient to produce consciousness, then the movements of two subatomic particles will be insufficient as well: if N particles are insufficient, then $N+1$ particles are insufficient too. So if one particle is insufficient, so are two. The same is true, moreover, of two subatomic particles: if their movements are insufficient to produce consciousness, then so are the movements of three, and if the movements of three particles are insufficient, then so are the movements of four, and also the movements of five, and of six, and seven, and so on for any number N . It appears, then, that no number of subatomic collisions will be sufficient to produce consciousness. How, then, did consciousness manage to emerge in the course of the universe’s history? And how, for that matter, does consciousness manage to emerge in you and me right now? This is the problem of psychophysical emergence.

Notice that I have formulated the problem by advancing some considerations that lead to a puzzling question. Another way of formulating the problem is to present a set of jointly inconsistent claims such as the following:

- 1 We are conscious beings.
- 2 We are composed entirely of nonconscious parts.
- 3 No number of nonconscious parts could combine to produce a conscious whole.
- 4 The properties of a whole are determined by the properties of its parts.

Claim (1) says that I am conscious, and claim (4) implies that my consciousness should be determined by my parts. My mass, for instance, is determined by the masses of the microscopic particles composing me; it thus seems plausible that if I am conscious, my consciousness is determined by the properties of my parts just as my mass is. But now we face a problem: According to claim (2), none of my parts is conscious, and according to (3) these parts cannot produce consciousness either. So claims (1) and (4) imply that my consciousness must be produced by my parts, but claims (2) and (3) imply that it cannot. Given reasonable assumptions, then, claims (1)–(4) are inconsistent with one another; they cannot all be true, and yet it is not clear which is false since there are good reasons for endorsing each.

One benefit of formulating philosophical problems this way is that the formulation makes it clear what any solution must accomplish: A solution must show either that one of the claims is false, or else that, despite appearances, the claims are not really inconsistent. A solution to the problem of psychophysical emergence, for instance, must either show that (1), or (2), or (3), or (4) is false, or else show that despite appearances these claims are not really inconsistent.

Another benefit of formulating philosophical problems this way is that it helps us evaluate solutions various philosophers have to offer (Figure 1.3). Consider some proposed solutions to the problem of psychophysical emergence. Eliminative physicalists reject claim (1). They deny there is any such thing as consciousness, and hence deny that we are conscious beings. Some dual-attribute theorists – panpsychists and panprotopsychists in particular – reject claim (2). They claim that the entities that compose us, including fundamental physical particles, have conscious or protoconscious states. Substance dualists, idealists, and nonorganismic dual-attribute theorists also reject (2) but for different reasons. They deny that we are composed of physical materials at all. Many emergentists and epiphenomenalists, by contrast, reject claim (3); they claim that consciousness emerges from nonconscious materials by virtue of brute psychophysical laws. Reductive physicalists (whose position we will discuss in detail in Chapter 5) also reject (3). They look to identify conscious states with complex relations among physical particles. Consequently, if enough physical particles stand in the right relations, there are

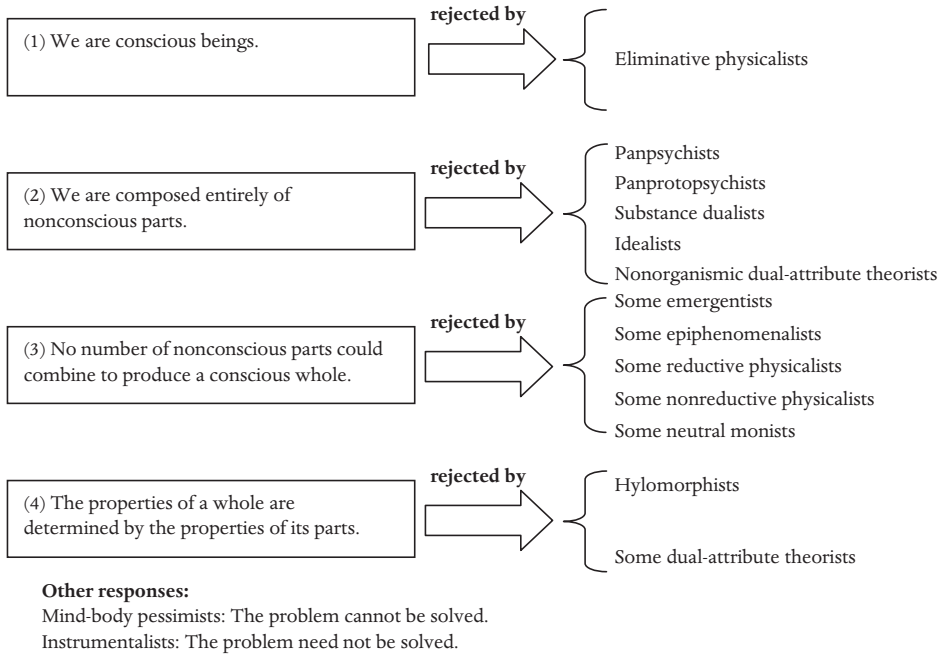


Figure 1.3 Solutions to the problem of psychophysical emergence

conscious states. Neutral monists and nonreductive physicalists (whose position we will discuss in Chapter 6) say something similar. The relations between conscious states and physical particles are not as straightforward as reductive physicalists suppose, they say, but whenever we talk about conscious states we are still talking about complex relations among physical or neutral particles. Hylomorphists, on the other hand, reject claim (4). Living wholes, they say, are structured in ways that confer on them capacities not had by their parts – including the ability to be conscious. Something’s structure, however, is not produced or determined by the materials it configures; structure is instead a basic ontological and explanatory principle in addition to those that govern materials in their own right. Finally, some mind–body pessimists claim that the problem is insoluble, that there are absolute limits on our ability to know and understand the world, and these limits manifest themselves in philosophical problems that cannot be solved like this one. Instrumentalists, on the other hand, do not insist that the problem cannot be solved but emphasize that it need not be solved. Psychological discourse is a useful tool, they say, and we need not solve mind–body problems to continue using it.

Knowing how different theories solve a given mind–body problem is not the end of a philosopher’s task, however, but only the beginning, for in order to evaluate the proposed solutions, a philosopher must evaluate the theories themselves, and this is a complicated task. Proponents of each theory advance reasons for

thinking their theory is true, and opponents of each theory advance reasons for thinking it is false. Evaluating the arguments for and against mind–body theories will be our primary concern in the chapters that follow. But first let us consider some more mind–body problems.

1.5 The Problem of Other Minds

The problem of other minds arises from a tension between our objective, third-person knowledge of human behavior, and our apparently subjective, first-person knowledge of our own conscious states. One of our basic starting points for understanding who and what we are is the idea that we are social beings. We know, for instance, that there are other people in the world, and that we often know through our ordinary interactions with them what they think and how they feel. Yet it is difficult to see how this kind of knowledge is possible. Mental states seem to be private, subjective phenomena. You do not have direct access to my mental states, nor do I have direct access to yours. You can hide your thoughts and feelings from me, and I can hide my thoughts and feelings from you. Thoughts and feelings seem to belong to a private, inner domain of subjective experiences – a domain distinct from the public, outer domain of bodily behavior. But if thoughts and feelings are private, if I alone have access to my mental states, then other people cannot know what my mental states are; in fact, they cannot even know whether I have any mental states since a human body seems capable of operating in just the way it does without having any conscious states at all. Even if I had no conscious experiences, it would still be possible for my nervous system to produce the kinds of bodily movements we associate with intelligent action. Conversely, because I cannot access other people’s mental states, I cannot really know what their mental states are, or even whether they have any mental states. For all I know, the public, objectively observable human bodies I see around me might be completely devoid of consciousness – they might simply be automata that act in every way as if they have conscious experiences like mine, while yet having no inner mental life at all. How, then, is it possible to have the knowledge of people we ordinarily take ourselves to have?

We can formulate the problem of other minds in terms of the following jointly inconsistent claims:

- 1 We often know what other people think and how they feel.
- 2 What other people think and how they feel belong to a private, subjective domain.
- 3 If what other people think and how they feel belong to a private, subjective domain, then we cannot know what other people think and how they feel as often as we suppose.

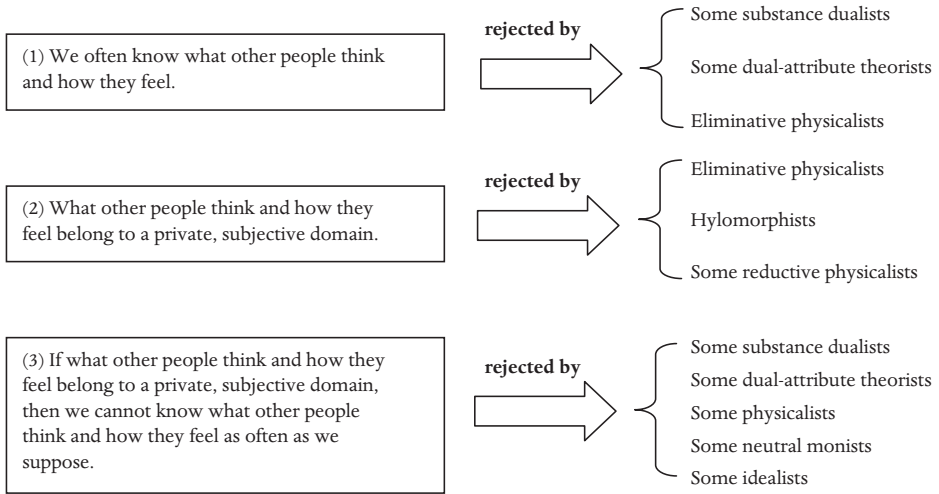


Figure 1.4 Solutions to the problem of other minds

There are several ways of solving the problem. Consider some examples (Figure 1.4). Some substance dualists and dual-attribute theorists reject claim (1). Our reasons for believing that mental states are subjective occurrences are stronger, they say, than our reasons for believing we know other people's mental states. Eliminative physicalists reject (1) as well and also (2): if there are no thoughts or feelings, then there are no thoughts or feelings that belong to a subjective domain. Moreover, say eliminativists, since we cannot know about something that doesn't exist, we cannot know about other people's mental states. Hylomorphists and some physicalists also reject claim (2) but for different reasons. Thoughts and feelings are not private subjective occurrences, they say, but patterns of social and environmental interaction that are just as objectively observable as patterns on a chessboard. Finally, many philosophers reject claim (3). Dual-attribute theorists, some substance dualists, as well as many physicalists and neutral monists claim that we are able to know other people's mental states by making inferences from the objectively observable behavior of human bodies. In addition, some substance dualists and idealists claim that we can know other people's thoughts and feelings directly through a special mind-reading faculty that we possess.

1.6 The Problem of Mental Causation

Consider one more mind–body problem: the **problem of mental causation**. It arises from a tension between our commonsense understanding of people's reasons for performing actions and our scientific understanding of the physical

mechanisms involved in their performance. We take it for granted that our beliefs, desires, and other mental states are able to influence the physical universe. When I order dinner at a restaurant, I take it for granted that my body's speech mechanisms are triggered by my desire to order or my intention to speak. Likewise, when I step into an automobile, I assume that it is my beliefs and desires about where to go and how to get there that will govern how my body steers the wheel, presses the accelerator, or applies the brakes. In general, I assume that my mental states are responsible for producing my actions. In fact, the very existence of actions seems to presuppose that mental states can influence physical behavior. Actions, at least the sort involving bodily movements, appear to be physical events with mental causes. If I accidentally trip on the rug we do not call it an action: if, however, I am clowning around and *intend* to trip on the rug, we do call it an action. Why does the one case count as an action while the other does not? The difference is not a physical one. Both cases might involve exactly the same physical occurrences: my foot catching the rug, my frame catapulting forward, my hands striking the floor, and so on. The difference, it seems, is mental. The second case qualifies as an action because it has a mental cause: my intention to trip on the rug. One of our basic assumptions about the world, then, seems to be that beliefs, desires, and other mental states can cause changes in the physical universe – an assumption as basic as our belief that there are such things as actions.

Physical events, however, can be triggered by other physical events. If we stimulate your nervous system in the right way, we can trigger exactly the same bodily movements that are involved in your actions. The neuroscientist Jose Delgado was a pioneer of the technique of neural manipulation. He once implanted electrodes in a bull's midbrain that were activated by remote control – a device he called a 'stimoceiver'. Using the device, he was able to make the bull halt in mid-charge with the push of a button – a dramatic demonstration of neural manipulation reported on the front page of the *New York Times*. Delgado performed experiments on other animals as well – cats, monkeys, chimpanzees, and also humans. He altered the behavior of over 20 human subjects – their feelings, moods, and movements – by electrically stimulating regions of their brains. The human body is after all a vast collection of fundamental physical particles governed by fundamental physical laws. We can manipulate its movements and states in the same ways we can manipulate the movements and states of any other physical system. Usually, however, we leave it to our nervous systems to manage their own affairs. The physical movements involved in your actions are typically caused by other physical events in your nervous system not by external devices, but the principles involved in producing those movements are the same in both cases; they are the kinds of principles described by physics.

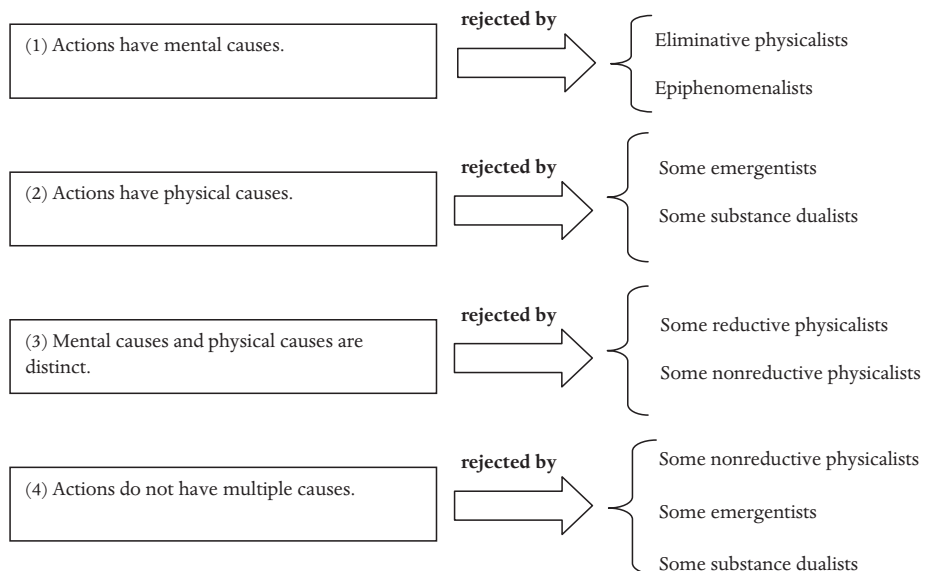
With these points in mind, consider a simple action – reaching for an object near at hand, say. This action cannot occur without the contraction of muscles in your arm and shoulder. These contractions are caused by events in your nervous system, the firings of neurons. These neuronal firings are caused in turn by other

neuronal firings, and those by yet other physical events such as the impact on your nervous system by light, sound, pressure, airborne chemicals, and other environmental influences. Recall, however, that in order for your reaching to count as an action it must have a mental cause as well – a desire to grasp an object, for instance. But now we face a problem: your action has a physical cause, an event or series of events in your nervous system, and it also has a mental cause, your desire to reach. How are the mental cause of your action and its physical cause related? We can formulate this problem in terms of the following jointly inconsistent claims:

- 1 Actions have mental causes.
- 2 Actions have physical causes.
- 3 Mental causes and physical causes are distinct.
- 4 An action does not have more than one cause.

Claims (1) and (2) imply that any given action has a mental cause and also a physical cause. According to claim (3), an action's mental cause and its physical cause are distinct. The action must therefore have at least two causes, yet (4) rules this out. It says that an action does not have more than one cause. Consequently, claims (1)–(4) are inconsistent. Claims (1)–(3) imply that actions have multiple causes while claim (4) implies that they do not.

There are several ways of solving the problem of mental causation. Here are some examples (Figure 1.5). Eliminativists reject claim (1): since there are



Other responses:

Hylomorphists: The inconsistency of (1)–(4) is merely apparent since there is an equivocation on the term 'cause'. There are many different kinds of causes.

Figure 1.5 Solutions to the problem of mental causation

no mental events, there are no mental events that cause actions. Epiphenomenalists also reject claim (1) but for different reasons: there are mental events, they say, but those events do not causally contribute to anything. Some emergentists and substance dualists reject (2). They claim that physical laws are periodically violated. Whenever we perform an action such as your act of reaching, the action's physical antecedents cease to be causally efficacious – the events in your nervous system, for instance, no longer have the effects they normally would. Your reaching has only a mental cause, your desire; events in your nervous system are not responsible for bringing about your bodily movements whenever you act. Some reductive physicalists – in particular identity theorists (whose position we discuss in Chapter 5) – and some nonreductive physicalists reject (3). Your reaching has only one cause, they say; the neural firings just *are* your desire. The word 'desire', in other words, is just another way of referring to certain events in your nervous system, the way 'water' is just another way of referring to H₂O molecules. Some dual-attribute theories and substance dualists reject (4). Your reaching has two independent causes: The desire and the neuronal firings both bring about the action, the way you and a friend might both pull a single lever even though each of you is capable of doing it on your own. Your action is thus *overdetermined*, it has more than one independent, fully sufficient cause. Hylomorphists, on the other hand, argue that claims (1)–(4) are not really inconsistent. They claim that there is an equivocation on the term 'cause'. By analogy, there is an equivocation on the term 'law' in the following sentences:

The Supreme Court can overturn any law.

The Supreme Court cannot overturn the law of gravity.

Because the term 'law' is used in different ways in these two sentences, neither implies the falsity of the other. Something analogous is true, say hylomorphists, of the term 'cause' in (1)–(4). Because that term is used in different ways in the four claims, they are not really inconsistent.

We now have a preliminary understanding of mind–body theories and mind–body problems. It is time to consider them in greater detail. We begin in the next chapter by considering the distinctive characteristics of mental and physical phenomena – what philosophers take themselves to mean when they use the words 'mental' and 'physical'.

Further Reading

For more on cortical mapping and the effects of stimulating the brain electrically see Kolb and Whishaw (2003). Jose Delgado's work in neural stimulation is

discussed by John Horgan (2005). Substance dualism is discussed in detail in Chapter 3. The general physicalist worldview is discussed in Chapter 4. Reductive and nonreductive physicalism are discussed in chapters 5 and 6, respectively. Eliminativism, instrumentalism, and anomalous monism are discussed in Chapter 7. Dual-attribute theories are the subject of Chapter 8. Neutral monism, idealism, and mind–body pessimism are discussed in Chapter 9, and hylomorphism is discussed in Chapters 10 and 11.

Notes

- 1 Wilder Penfield and Phanor Perot, 1963, “The Brain’s Record of Auditory and Visual Experience: A Final Summary and Discussion,” *Brain* 86: 595–696, 615–17.
- 2 Penfield and Perot, 618.
- 3 Bryan Kolb and Ian Q. Whishaw, 2003, *Fundamentals of Human Neuropsychology*, 5th edn, New York: Worth Publishers, Chapter 11.
- 4 John A. Osmundsen, 1965, “‘Matador’ with a Radio Stops Wired Bull,” *New York Times*, May 17.
- 5 Arthur Stanley Eddington, 1928, *The Nature of the Physical World*, New York: Macmillan Company, ix–xii.