

JAEGWON KIM



PHILOSOPHY OF MIND

THIRD EDITION

CHAPTER 1

Introduction

In coping with the myriad things and events that come our way at every moment of our waking life, we try to organize them into manageable chunks. We do this by sorting things into groups—categorizing them as “rocks,” “trees,” “fish,” “birds,” “bricks,” “fires,” “rains,” and countless other kinds—and describing them in terms of their properties and features as “large” or “small,” “tall” or “short,” “red” or “yellow,” “slow” or “swift,” and so on. A distinction that we almost instinctively apply to just about everything is whether it is a *living* thing. (It might be a dead bird, but still we know it is the *kind* of thing that lives, unlike a rock or a celadon vase, which couldn’t be “dead.”) There are exceptions, of course, but it is unusual for us to know what something is without at the same time knowing, or having some ideas about, whether it is a living thing. Another example: When we know a person, we almost always know whether the person is male or female.

The same is true of the distinction between things, or creatures, with a “mind” and those without a mind. This, too, is one of the most basic contrasts we use in our thoughts about things in the world. Our attitudes toward creatures that are conscious and capable of experiencing sensations like pain and pleasure are importantly different from our attitudes toward things lacking such capacities, mere chunks of matter or insentient plants, as witness the controversies about vegetarianism and scientific experiments involving live animals. And we are apt to regard ourselves as occupying a special and distinctive place in the natural world on account of our particularly highly developed mental capacities and functions, such as the capacity for abstract thoughts, self-consciousness, artistic sensibilities, complex emotions, and a capacity for rational deliberation and action. Much as we admire the miracle of the flora and fauna, we do not think that every living thing has a mind or that we need a psychological theory to understand the life cycles of elms and birches or the behavior and reproductive patterns of amoebas. Except those few of us with certain mystical inclinations, we do not think that members of the plant world are endowed with mentality, and we would exclude many members of the animal kingdom from the mental realm as well. We would not think that planarians and gnats have a mental life that is fit for serious psychological inquiry.

When we come to higher forms of animal life, such as cats, dogs, and chimpanzees, we find it entirely natural to grant them a fairly rich mental life. They are surely *conscious* in that they experience *sensations*, like pain, itch, and pleasure; they *perceive* their surroundings more or less the way we do and use the information so gained to guide their behavior. They also *remember* things—that is, store and use information about their surroundings—and *learn* from experience, and they certainly appear to have *feelings* and *emotions*, such as fear, frustration, and anxiety. We describe their psychological life using the expressions we normally use for fellow human beings: “Phoebe is feeling cramped inside the pet carrier and all that traffic noise has made her nervous. The poor thing is dying to be let out.”

But are the animals, even the more intelligent ones like horses and dolphins, capable of complex social emotions like embarrassment and shame? Are they capable of forming intentions, engaging in deliberation and making decisions, or performing logical reasoning? When we go down the ladder of animal life to, say, oysters, crabs, and earthworms, we would think that their mental life is considerably impoverished in comparison with that of, say, a domestic cat. Surely these creatures have sensations, we think, for they react in appropriate ways to noxious stimuli, and they have sense organs through which they gain information about what goes on around them and adjust and modify their behavior accordingly. But do they have minds? Are they conscious? Do they have mentality? What is it to have a mind, or mentality?

WHAT IS PHILOSOPHY OF MIND?

Philosophy of mind, like any other field of inquiry, is defined by a group of problems. As we expect, the problems that constitute this field concern mentality and mental properties. What are some of these problems? And how do they differ from the scientific problems about mentality and mental properties, those that psychologists, cognitive scientists, and neuroscientists investigate in their research?

There is, first of all, the problem of answering the question raised earlier: What is it to be a creature with a mind? Before we can fruitfully consider questions like whether inorganic electromechanical devices (for example, computers and robots) can have a mind, or whether speechless animals are capable of having thoughts, we need a reasonably clear idea about what mentality is and what having a thought consists in. What conditions must a creature or system meet if we are to attribute to it a “mind” or “mentality”? We commonly distinguish between mental phenomena, like thoughts and sensory experiences, and those that are not mental, like digestive processes or the circulation of blood through the arteries. Is there a general characteristic that distinguishes mental phenomena from nonmental, or “merely” physical, phenomena? We canvass some suggestions for answering these questions later in this chapter.

There are also problems concerning specific mental properties or kinds of mental states and their relationship to one another. Are pains only sensory events (they hurt), or must they also have a motivational component (such as aversiveness)? Can there be pains of which we are not aware? Do emotions like anger and jealousy necessarily involve felt qualities? Do they involve a cognitive component, like belief? What is a belief anyway, and how does a belief come to have the content it has (say, that it is raining outside, or that $7 + 5 = 12$)? Do beliefs and thoughts require a capacity for speech?

A third group of problems concerns the relation between minds and bodies, or between mental and physical phenomena. Collectively called “the mind-body problem,” this has been a central problem of philosophy of mind since Descartes introduced it nearly four centuries ago. It is a central problem for us in this book as well. The task here is to clarify and make intelligible the relation between our mentality and the physical nature of our being—or more generally, the relationship between mental and physical properties. But why should we think there is a philosophical problem here? Just what needs to be clarified and explained?

A simple answer might go like this: The mental seems *prima facie* so utterly different from the physical, and yet the two seem intimately related to each other. When you think of conscious experiences—such as the smell of basil, a pang of remorse, or the burning painfulness of a freshly bruised elbow—it is hard to imagine anything that could be more different from mere configurations and motions, however complex, of material particles, atoms and molecules, or mere physical changes involving cells and tissues. In spite of that, these conscious phenomena don’t come out of thin air, or from some immaterial source; rather, they arise from certain configurations of physical-biological processes of the body, including neural processes in the brain. We are at bottom physical-biological systems—complex biological structures wholly made up of bits of matter. (In case you disagree, we consider Descartes’s contrary views in chapter 2.) How can biological-physical systems come to have states like thoughts, fears, and hopes, experience feelings like guilt and pride, act for reasons, and be morally responsible? It strikes many of us that there is a fundamental, seemingly unbridgeable gulf between mental and physical phenomena and that this makes their apparently intimate relationships puzzling and mysterious.

It seems beyond doubt that phenomena of the two kinds are intimately connected. For one thing, evidence indicates that mental events occur as a result of physical-neural processes. Stepping barefoot on an upright thumbtack causes a sharp pain in your foot. It is likely that the proximate basis of the pain is some event in your brain: A bundle of neurons deep in your hypothalamus or cortex discharges, and as a result you experience a sensation of pain. Impingement of photons on your retina starts off a chain of

events, and as a result you have a certain visual experience, which in turn leads you to form the belief that there is a tree in front of you. How could a series of physical events—little particles jostling against one another, electric current rushing to and fro, and so on—blossom all of a sudden into a conscious experience, like the burning hurtfulness of a badly scalded hand, the brilliant red and purple sunset you see over the dark green ocean, or the smell of freshly mown lawn? We are told that when certain special neurons (nociceptive neurons) fire, we experience pain, and presumably there is another group of neurons that fire when we experience an itch. Why are pain and itch not switched around? That is, why is it that we feel pain, rather than itch, when just these neurons fire and we experience itch, not pain, when those other neurons fire? Why is it not the other way around? Why should any experience emerge from molecular-biological processes?

Moreover, we take it for granted that mental events have physical effects. It seems essential to our concept of ourselves as agents that our bodies are moved in appropriate ways by our wants, beliefs, and intentions. You see a McDonald's sign across the street and you decide to get something to eat, and somehow your perception and decision cause your limbs to move in such a way that you now find your body at the doors of the restaurant. Cases like this are among the familiar facts of life and are too boring to mention. But how did your perception and desire manage to move your body, all of it, across the street? You say, that's easy: Beliefs and desires first cause certain neurons in the motor cortex of my brain to discharge, these neural impulses are transmitted through the network of neural fibers all the way down to the peripheral control systems, which cause the appropriate muscles to contract, and so on. All that might be a complicated story, you say, but it is something that brain science, not philosophy, is in charge of explaining. But how do beliefs and desires manage to cause those little neurons to fire to begin with? How can this happen unless beliefs and desires are themselves just physical happenings in the brain? But is it coherent to suppose that these mental states are simply physical processes in the brain? These questions do not seem to be questions that can be answered just by doing more research in neuroscience; they seem to require philosophical reflection and analysis beyond what we can learn from science alone. This is what is called the problem of mental causation, one of the most important issues concerning the mind ever since Descartes first formulated the mind-body problem.

In this book, we are chiefly, though not exclusively, concerned with the mind-body problem. We begin, in the next chapter, with an examination of Descartes's mind-body dualism—a dualism of material things and immaterial minds. In contemporary philosophy of mind, however, the world is conceived to be fundamentally material: There are persuasive (some will say compelling) reasons to believe that the world we live in is made up wholly of material particles and their structured aggregates, all behaving strictly in accordance with physical laws. How can we accommodate minds and mentality in such an austere material world? That is our main question.

But before we set out to consider specific doctrines concerning the mind-body relationship, it will be helpful to survey some of the basic concepts, principles, and assumptions that guide the discussions to follow.

METAPHYSICAL PRELIMINARIES

For Descartes, “having a mind” had a literal meaning. On his view, minds are things of a special kind, souls or immaterial substances, and having a mind simply amounts to having a soul, something outside physical space, whose essence consists in mental activities like thinking and being conscious. (We examine this view of minds in chapter 2.) A substantival view of mentality like Descartes’s is not widely accepted today. However, to reject minds as substances or objects in their own right is not to deny that each of us “has a mind”; it is only that we need not think of “having a mind” as there being some object called a “mind” that we literally “have.” Having a mind need not be like having brown eyes or a laptop. Think of “dancing a waltz” or “taking a walk”: When we say, “Sally danced a waltz,” or “Sally took a leisurely walk along the river,” we do not mean—at least we do not need to mean—that there are *things* in this world called “waltzes” or “walks” such that Sally picked out one of them and danced it or walked it. Where are these dances and walks when no one is dancing or walking them? What could you do with a dance except dance it? Dancing a waltz is not like owning an SUV or kicking a tire. Dancing a waltz is merely a *manner* of dancing, and taking a walk is a *manner* of moving your limbs in a certain relationship to the physical surroundings. In using these expressions, we need not accept the existence of entities like waltzes and walks; all we need to admit into our ontology—the scheme of entities we accept as real—are persons who waltz and persons who walk.

Similarly, when we use expressions like “having a mind,” “losing one’s mind,” “being out of one’s mind,” and the like, there is no need to suppose there are objects in this world called “minds” that we have, lose, or are out of. Having a mind can be construed simply as having a certain group of *properties*, *features*, and *capacities* that are possessed by humans and some higher animals but absent in things like rocks and trees. To say that some creature “has a mind” is to classify it as a certain sort of being, capable of certain characteristic sorts of behaviors and functions—sensation, perception, memory, learning, reasoning, consciousness, action, and the like. It is less misleading, therefore, to speak of “mentality” than of “having a mind”; the surface grammar of the latter abets the problematic idea of a substantival mind—mind as an object of a special kind. However, this is not to preclude substantival minds at the outset; the point is only that we should not infer their existence from our use of certain forms of expression. As we will see in the chapter to follow, there are serious philosophical arguments that we must accept minds as immaterial things. Moreover, an influential contemporary view identifies minds with brains (discussed in chapter 4). Like Descartes’s substance dualism, this view gives a literal meaning to “having a mind”: It would simply mean having a brain of certain structure and capacities. The main point we should keep in mind is that all this requires philosophical considerations and arguments, as we will see in the rest of this book.

Mentality is a broad and complex category. As we just saw, there are numerous specific properties and functions through which mentality manifests itself, such as experiencing sensations, entertaining thoughts, reasoning and judging, making decisions, and feeling emotions. There are also more specific properties that fall within these categories, such as experiencing a throbbing pain in the right elbow, believing that Kabul is in Afghanistan, wanting to visit Tibet, and being annoyed at your roommate. In this book, we often talk in terms of “instantiating,” “exemplifying,” or “having” this or that property. When you shut a door on your thumb, you will likely *instantiate* or *exemplify* the property of being in pain; most of us *have*, or *instantiate*, the property of believing that snow is white; some of us have the property of wanting to visit Tibet; and so on. Admittedly this is a somewhat cumbersome, not to say stilted, way of talking, but it gives us a uniform and simple way of referring to certain entities and their relationships. Throughout this book, the expressions “mental” and “psychological” and their respective cognates are used interchangeably. In most contexts, the same goes for “physical” and “material.”

We will now set out in general terms the kind of ontological scheme that we presuppose in this book and explain how we use certain terms associated with the scheme. We suppose, first, that our scheme includes *substances*, that is, *things* or *objects* (including persons, biological organisms and their organs, molecules, computers, and such) and that they have various *properties* and stand in various *relations* to each other. (Properties and relations are together called *attributes*.) Some of these are physical, like having a certain mass or temperature, being one meter long, being longer than, and being between two other objects. Some things—in particular, persons and certain biological organisms—can also instantiate mental properties, like being in pain, fearing darkness, and disliking the smell of ammonia. We also speak of mental or physical *events*, *states*, and *processes* and sometimes of *facts*. A process can be thought of as a (causally) connected series of events and states; events differ from states in that they suggest *change*, whereas states do not. The terms “phenomenon” and “occurrence” can be used to cover both events and states. We often use one or another of these terms in a broad sense inclusive of the rest. (For example, when we say “every event has a cause,” we are not excluding states, phenomena, and the rest.) How events and states are related to objects and their properties is a question of some controversy in metaphysics. We simply assume here that when a person instantiates, at time *t*, a mental property—say, being in pain—then there is the event (or state) of that person’s being in pain at *t*, and there is also the fact that the person is in pain at *t*. Some events are psychological events, such as pains, beliefs, and onsets of anger, and these are instantiations by persons and other organisms of mental properties. Some events are physical, such as earthquakes, hiccups and sneezes, and the firing of a bundle of neurons, and these are instantiations of physical properties. Another point to note: In the context of the mind-body problem, the physical usually goes beyond the properties and phenomena studied in physics; the biological, the chemical, the geological, and so on, also count as physical.

So much for the ontological preliminaries. Sometimes clarity and precision demand attention to ontological details, but as far as possible we will try to avoid general metaphysical issues that are not germane to our concerns about the nature of mind.

MIND-BODY SUPERVENIENCE

Consider the apparatus called the “transporter” in the science-fiction television series *Star Trek*. You walk into a booth. When the transporter is activated, your body is instantly disassembled; exhaustive information concerning your bodily structure and composition, down to the last molecule, is transmitted, instantaneously, to another location, often a great distance away, where a body that is exactly like yours is reconstituted (presumably with local material). And someone who looks just like you materializes on the spot and starts doing the tasks you were assigned to do there.

Let us not worry about whether the person who is created at the destination is really you or only your replacement. In fact, we can avoid this issue by slightly changing the story: Exhaustive information about your bodily composition is obtained by a scanner that does no harm to you, and on the basis of this information, an exact physical replica of your body—a molecule-for-molecule identical duplicate—is created at another location. By assumption, you and your replica have exactly the same *physical* properties; you and your replica could not be distinguished by any *current intrinsic* physical differences. We say “current” to rule out the obvious possibility of distinguishing you from your duplicate by tracing the causal chains backward to the past. We say “intrinsic” because you and your replica have different relational, or extrinsic, properties; for example, you have a mother but your replica does not.

Given that your replica is your *physical* replica, will she also be your *psychological* replica? Will she be identical with you in all mental respects as well? Will she be as smart and witty as you are, and as prone to daydream? Will she share your likes and dislikes in food and music and behave just as you would when angry or irritable? Will she prefer blue to green and have a visual experience exactly like yours when you and she both gaze at a Van Gogh landscape of yellow wheat fields against a dark blue sky? Will her twinges, itches, and tickles feel to her just the way yours feel to you? Well, you get the idea. An unquestioned assumption of *Star Trek* and similar science-fiction fantasies seems to be that the answer is yes to each of these questions. If you are like the many *Star Trek* fans in going along with this assumption, that is because you have tacitly consented to the following “supervenience” thesis:

Mind-Body Supervenience I. The mental supervenes on the physical in that things (objects, events, organisms, persons, and so on) that are exactly alike in all physical properties cannot differ with respect to mental properties. That is, physical indiscernibility entails psychological indiscernibility.

Or as it is sometimes put: No mental difference without a physical difference. Notice that this principle does not say that things that are alike in psychological respects must be alike in physical respects. We seem to be able coherently to imagine intelligent extraterrestrial creatures whose biochemistry is different from ours (say, their physiology is not carbon-based) and yet who share the same psychology with us. As we might say, the same psychology could be realized in different physical systems. Now, that may or may not be the case. The thing to keep in mind, though, is that mind-body supervenience asserts only that creatures could not be psychologically different and yet physically identical.

There are two other important ways of explaining the idea that the mental supervenes on the physical. One is the following, known as “strong supervenience”:

Mind-Body Supervenience II. The mental supervenes on the physical in that if anything *x* has a mental property *M*, there is a physical property *P* such that *x* has *P*, and necessarily any object that has *P* has *M*.

Suppose that a creature is in pain (that is, it has the mental property of being in pain). This supervenience principle tells us that in that case there is some physical property *P* that the creature has that

“necessitates” its being in pain. That is to say, pain has a physical substrate (or “supervenience base”) such that anything that has this underlying physical property must be in pain. Thus, this formulation of mind-body supervenience captures the idea that the instantiation of a mental property in something “depends” on its instantiating an appropriate physical “base” property (that is, a neural correlate or substrate). How is this new statement of mind-body supervenience related to the earlier statement? It is pretty straightforward to show that the supervenience principle (II) entails (I); that is, if the mental supervenes on the physical according to (II), it will also supervene according to (I). Whether (I) entails (II) is more problematic.¹ For practical purposes, however, the two principles may be considered equivalent, and we make use of them in this book without worrying about their subtle differences.

There is another common way of understanding the supervenience relationship:

Mind-Body Supervenience III. The mental supervenes on the physical in that worlds that are alike in all physical respects are alike in all mental respects as well; in fact, worlds that are physically alike are exactly alike overall.²

This formulation of supervenience, called “global” supervenience, states that if there were another world that is just like our world in all physical respects, with the same particles, atoms, and molecules in the same places and the same laws governing their behavior, the two worlds could not differ in any mental respects. If God created this world, all he had to do was to put the right basic particles in the right places and fix basic physical laws, and all else, including all aspects of mentality, would just come along. Once the basic physical structure is put in place, his job is finished; he does not *also* have to create minds or mentality, any more than trees or mountains or bridges. The question whether this formulation of supervenience is equivalent to either of the earlier two is a somewhat complicated one; let it suffice to say that there are close relationships between all three. In this book, we do not have an occasion to use (III); however, it is stated here because this is the formulation some philosophers favor and you will likely come across it in the philosophy of mind literature.

To put mind-body supervenience in perspective, it might be helpful to look at supervenience theses in other areas—in ethics and aesthetics. Most moral philosophers would accept the thesis that the ethical, or normative, properties of persons, acts, and the like are supervenient on their nonmoral, descriptive properties. That is, if two persons, or two acts, are exactly alike in all nonmoral respects (say, the persons are both honest, courageous, kind, generous, and so on), they could not differ in moral respects (say, one of them is a morally good person but the other is not). Supervenience seems to apply to aesthetic qualities as well: If two pieces of sculpture are physically exactly alike (the same shape, size, color, texture, and all the rest), they cannot differ in some aesthetic respect (say, one of them is elegant, heroic, and expressive while the second has none of these properties). A world molecule-for-molecule identical with our world will contain works of art just as beautiful, noble, and mysterious as our Michelangelos, Vermeers, and Magrittes. One more example: Just as mental properties are thought to supervene on physical properties, most consider biological properties to supervene on more basic physicochemical properties. It seems natural to suppose that if two things are exactly alike in basic physical and chemical features, including, of course, their material composition and structure, it could not be the case that one of them is a living thing and the other is not, or that one of them is performing a certain biological function (say, photosynthesis) and the other is not. That is to say, physicochemically indiscernible things must be biologically indiscernible.

As noted, most philosophers accept these supervenience theses; however, whether they are true, or why they are true, are philosophically nontrivial questions. And each supervenience thesis must be evaluated and assessed on its own merit. Mind-body supervenience, of course, is our present concern. Our ready acceptance of the idea of the *Star Trek* transporter shows the strong intuitive attraction of mind-body

supervenience. But is it true? What is the evidence in its favor? Should we accept it? These are deep and complex questions. One reason is that, in spirit and substance, they amount to the following questions: Is physicalism true? Should we accept physicalism?

MATERIALISM AND PHYSICALISM

Since materialism, or physicalism, broadly understood is the basic framework in which contemporary philosophy of mind has been debated, it is useful for us to begin with some idea of what it is. Materialism is the doctrine that all things that exist in the world are bits of matter or aggregates of bits of matter. There is no thing that isn't a material thing—no transcendental beings, Hegelian “absolutes,” or immaterial minds. Physicalism is the contemporary successor to materialism. The thought is that the traditional notion of material stuff was ill-suited to what we now know about the material world from contemporary physics. For example, the concept of a “field” is widely used in physics, but it is unclear whether fields would count as material things in the traditional sense. Physicalism is the doctrine that all things that exist are entities recognized by the science of physics, or systems aggregated out of such entities.³ According to some physicalists, so-called nonreductive physicalists, these physical systems can have nonphysical properties, properties that are not recognized by physics or reducible to them. Psychological properties are among the prime candidates for such nonphysical properties possessed by physical systems.

If you are comfortable with the idea of the *Star Trek* transporter, that means you are comfortable with physicalism as a perspective on the mind-body problem. The wide and seemingly natural acceptance of the transporter idea shows how pervasively physicalism has penetrated contemporary culture, although when this is made explicit some people would no doubt recoil and proclaim themselves to be against physicalism.

What is the relationship between mind-body supervenience and physicalism? We have not so far defined what physicalism is, but the term itself suggests that it is a doctrine that affirms the primacy, or basicness, of what is physical. With this very rough idea in mind, let us see what mind-body supervenience implies for the dualist view (to be discussed in more detail in chapter 2) associated with Descartes that minds are immaterial substances with no physical properties whatever. Take two immaterial minds: Evidently, they are exactly alike in all physical respects since neither has any physical property and as a result it is impossible to distinguish them from a physical perspective. So if mind-body supervenience, in the form of (I), holds, it follows that they are alike in all mental respects. That is, under mind-body supervenience (I), all Cartesian immaterial souls are exactly alike in all mental respects, from which it follows that they are exactly alike in all possible respects. From this it seems to follow that there can be at most one immaterial soul! No serious mind-body dualist would find these consequences of mind-body supervenience tolerable. This is one way of seeing why the dualist will want to reject mind-body supervenience.

To appreciate the physicalist implication of mind-body supervenience, we must consider one aspect of supervenience that we have not so far discussed. Many philosophers regard the supervenience thesis as affirming a relation of *dependence* or *determination* between the mental and the physical; that is, the mental properties a given thing has depend on, or are determined by, the physical properties it has. Consider version (II) of mind-body supervenience: It says that for every mental property M, if anything has M, it has some physical property P that *necessitates* M—if anything has P, it *must* have M. This captures the idea that mental properties must have neural, or other physical, “substrates” from which they arise and that there can be no instantiation of a mental property that is not grounded in some physical property. So a dependence relation can naturally be read into the claim that the mental supervenes on the physical, although, strictly speaking, the supervenience theses as stated only make claims about how mental properties *covary* with physical properties. In any case, many physicalists interpret supervenience as implying mind-body dependence in something like the following sense:

Mind-Body Dependence. The mental properties a given thing has depend on, and are determined by,

the physical properties it has. That is, our psychological character is wholly determined by our physical nature.

The dependence thesis is important because it is an explicit affirmation of the *ontological primacy*, or *priority*, of the physical in relation to the mental. The thesis seems to accord well with the way we ordinarily think of the mind-body relation, as well as with scientific assumptions and practices. Few of us would think that there can be mental events and processes that float free, so to speak, of physical processes; most of us believe that what happens in our mental life, including the fact that we have a mental life at all, is dependent on what happens in our body, in particular in our nervous system. Furthermore, it is because mental states depend on what goes on in the brain that it is possible to intervene in the mental goings-on. To ease your headache, you take aspirin—the only way you can affect the headache is to alter the neural base on which it supervenes. There apparently is no other way.

For these reasons, we can think of the mind-body supervenience thesis, in one form or another, as *minimal physicalism*, in the sense that it is one commitment that all who consider themselves physicalists must accept. But is it sufficient as physicalism? That is, can we say that anyone who accepts mind-body supervenience is ipso facto a full physicalist? Opinions differ on this question. We saw earlier that supervenience does not by itself completely rule out the existence of immaterial minds, something antithetical to physicalism. But we also saw that supervenience has consequences that no serious dualist can accept. Whether supervenience itself suffices to deliver physicalism depends, by and large, on what we consider to be full and robust physicalism. As our starting options, then, let us see what varieties of physicalism are out there.

First, there is an ontological claim about what objects there are in this world:

*Substance Physicalism.*⁴ All that exists in this world are bits of matter in space-time and aggregate structures composed of bits of matter. There is nothing else in the space-time world.

This thesis, though it is disputed by Descartes and other substance dualists, is accepted by most contemporary philosophers of mind. The main point of contention concerns the *properties* of material or physical things. Certain complex physical systems, like higher organisms, are also psychological systems; they exhibit psychological properties and engage in psychological activities and functions. How are the psychological properties and physical properties of a system related to each other? Broadly speaking, an ontological physicalist has a choice between the following two options:

Property Dualism, or Nonreductive Physicalism. The psychological properties of a system are distinct from, and irreducible to, its physical properties.⁵

Reductive Physicalism, or Type Physicalism. Psychological properties (or kinds, types) are reducible to physical properties (kinds, types). That is, psychological properties and kinds are physical properties and kinds. There are only properties of one sort exemplified in this world, and they are physical properties.

Remember that for our purposes “physical” properties include chemical, biological, and neural properties, not just those properties investigated in basic physics (such as energy, mass, or charm). You could be a property dualist because you reject mind-body supervenience, but then you would not count as a physicalist since, as we argued, mind-body supervenience is a necessary element of physicalism. So the physicalist we have in mind is someone who accepts mind-body supervenience. However, it is generally supposed that mind-body supervenience is consistent with property dualism, the claim that the supervenient psychological properties are irreducible to, and not identical with, the underlying physical

base properties. In defense of this claim, some point to the fact that philosophers who accept the supervenience of moral properties on nonmoral, descriptive properties for the most part reject the reducibility of moral properties, like being good or being right, to nonmoral, purely descriptive properties. The situation seems the same with the case of aesthetic supervenience and aesthetic properties.⁶

Some philosophers who reject reductive, or type, physicalism as too ambitious and overreaching embrace “token” physicalism—the thesis that although psychological types are not identical with physical types, each and every individual psychological event, or event-token, is a physical event. So pain, as a mental kind, is not identical with, or reducible to, a kind of physical event or state, and yet each individual instance of pain—this pain here now—is usually a physical event. Token physicalism is considered a form of nonreductive physicalism. The continuing debate between nonreductive physicalists and reductive physicalists has largely shaped the contemporary debate on the mind-body problem.⁷

VARIETIES OF MENTAL PHENOMENA

It is useful at this point to look at some major categories of mental events and states. This will give us a rough idea about the kinds of phenomena we are concerned with and also remind us that the phenomena that come under the rubric “mental” or “psychological” are extremely diverse and variegated. The following list is not intended to be complete or systematic, and some categories obviously overlap others.

First, we may distinguish those mental phenomena that involve *sensations* or *sensory qualities*: pains, itches, tickles, having an afterimage, seeing a round green patch, smelling ammonia, feeling nauseous, and so on. These mental states are said to have a “phenomenal” or “qualitative” character—the way they *feel* or the way they *look* or *appear*. To use a popular term, there is *something it is like* to experience such phenomena or be in such states. Thus, pains have a special qualitative feel that is distinctive of pains—they hurt. Similarly, itches itch and tickles tickle. When you look at a green patch, there is a distinctive way the patch looks to you: It looks *green*, and your visual experience involves this green look. Each such sensation has its own distinctive feel and is characterized by a sensory quality that we seem to be able to identify directly, at least as to the general type to which it belongs (for example, pain, itch, or seeing green). These items are called “phenomenal” or “qualitative states,” or sometimes “raw feels.” However, “qualia” has now become the standard term for these sensory, qualitative states, or the sensory qualities experienced in such states. Collectively, these mental phenomena are said to constitute “phenomenal consciousness.”

Second, there are mental states that are attributed to a person by the use of embedded that-clauses: for example, President Barack Obama *hopes* that Congress will pass a health-care bill this year; Senator Harry Reid *is certain* that this will happen, and Newt Gingrich doubts that Obama will get what he wants. Such states are called “propositional attitudes.” The idea is that these states consist in a subject’s having an “attitude” (for example, hoping, being certain, doubting, and believing) toward a “proposition” (for example, that Congress will pass a health-care bill, that it will rain tomorrow). The propositions are said to constitute the “content” of the propositional attitudes, and that-clauses that specify these propositions are called “content sentences.” Thus, the content of Obama’s hope is the proposition that Congress will pass a health-care bill this year, which is also the content of Gingrich’s doubt, and this content is expressed by the sentence “Congress will pass a health-care bill this year.” These states are also called “intentional”⁸ or “content-bearing” states.

Do these mental states have a phenomenal, qualitative aspect? We do not normally associate a specific feel with beliefs, another specific feel with desires, and so on. There does not seem to be any special belief-like feel, a common sensory quality, associated with your belief that Providence is south of Boston and your belief that two is the smallest prime number. At least it seems that we can say this much: If you believe that two is the smallest prime and I do too, there does not seem to be—nor need there be—any common sensory quality that both of us experience in virtue of sharing this belief. The importance of these intentional states cannot be overstated. Much of our ordinary psychological thinking and theorizing (“commonsense” or “folk” psychology) involves propositional attitudes; we make use of them all the time to explain and predict what people will do. Why did Mary cross the street? Because she wanted some coffee and thought that she could get it at the Starbucks across the street. These states are essential to social psychology, and their analogues are found in various areas of psychology and cognitive science.

And then there are various mental states that come under the broad heading of *feelings* and *emotions*. They include anger, joy, sadness, depression, elation, pride, embarrassment, remorse, regret, shame, and many others. Notice that emotions are often attributed to persons with a that-clause. In other words, some states of emotions involve propositional attitudes: For example, you could be *embarrassed* that you had forgotten to call your mother on her birthday, and she could be *disappointed* that you did. Further, some

emotions involve belief: If you are embarrassed that you had forgotten your mother's birthday, you must believe that you did. As the word *feeling* suggests, there is often a special qualitative component we associate with many emotions, such as anger and grief, although it is not certain that all instances of emotion are accompanied by such qualitative feels, or that there is a single specific sensory feel to each kind of emotion.

There are also what some philosophers call "volitions," like intending, deciding, and willing. These states are propositional attitudes; intentions and decisions have content. For example, I may intend to take the ten o'clock train to New York tomorrow; here the content is expressed by an infinitive construction ("to take"), but it is easily spelled out in a full sentence, as in "I intend that I take the ten o'clock train to New York tomorrow." In any case, these states are closely related to actions. When I intend to raise my arm *now*, I must *now* undertake to raise my arm; when you intend, or decide, to do something, you commit yourself to doing it. You must be prepared not only to take the necessary steps toward doing it but also to initiate them at an appropriate time. This is not to say that you cannot change your mind, or that you will necessarily succeed; it is to say that you need to change your intention to be released from the commitment to action. According to some philosophers, all intentional actions must be preceded by an act of volition.

Actions typically involve motions of our bodies, but they do not seem to be mere bodily motions. My arm is going up, and so is yours. However, you are raising your arm, but I am not—my arm is being pulled up by someone else. The raising of your arm is an action; it is something you do. But the rising of my arm is not an action; it is not something that I do but something that happens to me. There appears to be something mental about your raising your arm that is absent from the mere rising of an arm; perhaps it is the involvement of your desire, or intention, to raise your arm, but exactly what distinguishes actions from "mere bodily motions" has been a matter of philosophical dispute. Or consider something like buying a loaf of bread. Evidently someone who can engage in the act of buying a loaf of bread must have appropriate beliefs and desires; she must, for example, have a desire to buy bread, or at least a desire to buy something, and knowledge of what bread is. And to do something like buying, you must have knowledge, or beliefs, about what constitutes buying rather than, say, borrowing or simply taking, about money and exchange of goods, and so on. That is to say, only creatures with beliefs and desires and an understanding of appropriate social conventions and institutions can engage in activities like buying and selling. The same goes for much of what we do as social beings; actions like promising, greeting, and apologizing presuppose a rich and complex background of beliefs, desires, and intentions, as well as an understanding of social relationships and arrangements.

There are other items that are ordinarily included under the rubric of "psychological," such as traits of character and personality (being honest, obsessive, witty, introverted), habits and propensities (being industrious, punctual), intellectual abilities, artistic talents, and the like. But we can consider them to be mental in an indirect or derivative sense: Honesty is a mental characteristic because it is a tendency, or disposition, to form desires of certain sorts (for example, the desire to tell the truth, or not to mislead others) and to act in appropriate ways (in particular, saying only what you sincerely believe).

In the chapters to follow, we focus on sensations and intentional states. They provide us with examples of mental states when we discuss the mind-body problem and other issues. We also discuss some specific philosophical problems about these two principal types of mental states. We will largely bypass detailed questions, however, such as what types of mental states there are, how they are interrelated, and the like.

But in what sense are all these variegated items "mental" or "psychological"? Is there some single property or feature, or a reasonably simple and perspicuous set of them, by virtue of which they all count as mental?

IS THERE A “MARK OF THE MENTAL”?

Various characteristics have been proposed by philosophers to serve as a “mark of the mental,” a criterion that would separate mental phenomena or properties from those that are not mental. Each has a certain degree of plausibility and can be seen to cover a range of mental phenomena, but as we will see, none seems to be adequate for all the diverse kinds of events and states, functions and capacities, that we normally classify as “mental” or “psychological.” Although we will not try to formulate our own criterion of the mental, a review of some of the prominent proposals will give us an understanding of the principal ideas traditionally associated with the concept of mentality and highlight some of the important characteristics of mental phenomena, even if, as noted, no single one of them seems capable of serving as a universal, necessary, and sufficient condition of mentality.

Epistemological Criteria

You are experiencing a sharp toothache caused by an exposed nerve in a molar. The toothache that you experience, but not the condition of your molar, is a mental occurrence. But what is the basis of this distinction? One influential answer says that the distinction consists in certain fundamental differences in the way you come to have *knowledge* of the two phenomena.

Direct or Immediate Knowledge. Your knowledge that you have a toothache, or that you are hoping for rain tomorrow, is “direct” or “immediate”; it is not based on evidence or inference. There is nothing else that you know or need to know from which you infer that you have a toothache; that is, your knowledge is not mediated by other beliefs or knowledge. This is seen in the fact that in cases like this the question “How do you know?” seems to be out of place (“How do you know you are hoping for rain and not snow?”). The only possible answer, if you take the question seriously, is that you *just* know. This shows that here the question of “evidence” is inappropriate: Your knowledge is direct and immediate, not based on evidence. Yet your knowledge of the physical condition of your tooth is based on evidence: Knowledge of this kind usually depends on the testimonial evidence provided by a third party—for example, your dentist. And your dentist’s knowledge presumably depends on the evidence of X-rays, visual inspection of your teeth, and so on. The question “How do you know that you have an exposed nerve in a molar?” makes good sense and can receive an informative answer.

But isn’t our knowledge of certain simple physical facts just as “direct” and “immediate” as knowledge of mental events like toothaches and itches? Suppose you are looking at a large red circle painted on a wall directly in front of you: Doesn’t it seem that you know, directly and without the use of any further evidence, that there is a round red patch in front of you? Don’t I know, in the same way, that here is a piece of white paper in front of me or that there is a tree just outside my window?

Privacy, or First-Person Privilege. One possible response to the foregoing challenge is to invoke the privacy of our knowledge of our own mental states, namely, the apparent fact that this direct access to a mental event is enjoyed by a single subject, the person to whom the event is occurring. In the case of the toothache, it is only you, not your dentist or anyone else, who is in this kind of specially privileged position. But this does not hold in the case of seeing the red patch. If you can know “directly” that there is a round red spot on the wall, so can I and anyone else who is suitably situated in relation to the wall. There is no single person with specially privileged access to the round red spot. In this sense, knowledge of mental events exhibits an *asymmetry* between first person and third person: It is only the first person, namely the subject who experiences a pain, who enjoys a special epistemic privilege as regards the pain. Others, that is, third persons, do not. In contrast, for knowledge of physical objects and states—say, the red round spot on the wall—there is no meaningful first-person /third-person distinction; everyone is a third person. Moreover, the first-person privilege holds only for knowledge of *current* mental occurrences, not for knowledge of *past* ones: You know that you had a toothache yesterday, a week ago, or two years ago, from the evidence of memory, an entry in your diary, your dental record, and the like.

But what about those bodily states we detect through proprioception, such as the positions and motions of our limbs (for example, knowing that your legs are crossed or that you are raising your right hand)? Our proprioceptors and associated neural machinery are in the business of keeping us *directly* informed of certain physical conditions of our bodies, and proprioception is, in general, highly reliable. Moreover, first-person privilege seems to hold for such cases: It is only I who know, through proprioception, that my right knee is bent; no third party has similar access to this fact. And yet it is knowledge of a bodily condition, not of a mental occurrence. Perhaps this example could be handled by appealing to the

following criterion.

Infallibility and Transparency. Another epistemic feature sometimes associated with mentality is the idea that in some sense your knowledge of your own current mental states is “infallible” or “incorrigible,” or that it is “selfintimating” (or that your mind is “transparent” to you). The main idea is that mental events—especially events like pains and other sensations—have the following property: You cannot be mistaken about whether you are experiencing them. That is, if you *believe* that you are in pain, then it follows that you *are* in pain, and if you believe that you are not in pain, then you are not; it is not possible to have false beliefs about your own pains. In this sense, your knowledge of your own pain is *infallible*. So-called psychosomatic pains are pains nonetheless; they can hurt just as badly. The same may hold for your knowledge of your own propositional attitudes like belief; Descartes famously said that you cannot be mistaken about the fact that you doubt, or that you think.⁹ In contrast, when your belief concerns a physical occurrence, there is no guarantee that your belief is true: Your belief that you have a decayed molar may be true, but its truth is not entailed by the mere fact that you believe it. Or so goes the claim, at any rate. Returning briefly to knowledge gained through proprioception, the reply would be that such knowledge may be reliable but not infallible; there can be incorrect beliefs about your bodily position based on proprioception.

Transparency is the converse of infallibility: A state or event *m* is said to be transparent to a person just in case, necessarily, if *m* occurs, the person is aware that *m* occurs—that is, she knows that *m* occurs.¹⁰ The claim, then, is that mental events are transparent to the subjects to whom they occur. If pains are transparent in this sense, there could not be *hidden* pains—pains that the subject is unaware of. Just as the infallibility of beliefs about your own pains implies that pains with no physiological cause are at least conceivable, the transparent character of pain implies that even if all normal physical and physiological causes of pain are present, if you are not aware of any pain, then you are not in pain. There are reports about soldiers in combat and athletes in the heat of competition that they experienced no pain in spite of severe physical injuries; if we assume pains are transparent, we would have to conclude that pain, as a mental event, is not occurring to these subjects. We may define “the doctrine of the transparency of mind” as the claim that nothing that happens in your mind escapes your awareness—that is, nothing in your mind is hidden from you. The conjunction of this doctrine and the doctrine of infallibility is often associated with the traditional conception of the mind, especially that of Descartes.

Infallibility and transparency are extremely strong properties. It would be no surprise if physical events and states did not have them; a more interesting question is whether all or even most mental events satisfy them. Evidently, not all mental events or states have these special epistemic properties. In the first place, it is now commonplace to speak of “unconscious” or “subconscious” beliefs, desires, and emotions, like repressed desires and angers—psychological states the subject is not aware of and would even vehemently deny having but that evidently shape and influence his action and behavior. Second, it is not always easy for us to determine whether an emotion that we are experiencing is, say, one of embarrassment, remorse, or regret—or one of envy, jealousy, or anger. And we are often not sure whether we “really” believe or desire something. Do I believe that globalization is a good thing? Do I believe that I am by and large a nice person? Do I want to be sociable and gregarious, or do I prefer to stay somewhat aloof and distant? If you reflect on such questions, you may not be sure what the answers are. It is not as though you have suspended judgment about them—you may not even know *that*. Epistemic uncertainties can happen with sensations as well. Does this overripe avocado smell a little like a rotten egg, or is it okay for the salad? Special epistemic access is perhaps most plausible for sensations like pains and itches, but here again, not all our beliefs about pains appear to have the special authoritative character indicated by the epistemic properties we have surveyed. Is the pain I am now experiencing more intense

than the pain I felt a moment ago in the same elbow? Just where in my elbow is the pain? Clearly there are many characteristics of pains, even introspectively identifiable ones, about which I could be mistaken and don't feel fully certain.

It is also thought that you can misclassify, or misidentify, the type of sensation you are experiencing: For example, you may report that you are itchy in the shoulder when the correct description would be that you have a ticklish sensation there. However, it is not clear just what such cases show. It might be replied, for example, that the error is a verbal one, not one of belief. Although you are using the sentence "My left shoulder is itchy" to report your belief, your belief is to the effect that your left shoulder *tickles*, and this belief is true.

Thus, exactly how the special epistemic character of mental events is to be characterized can be a complex and controversial business, and unsurprisingly there is little agreement in this area. Some philosophers, especially those who favor a scientific approach to mentality, would take pains to minimize these *prima facie* differences between mental and bodily events. But it is apparent that there are important epistemological differences between the mental and the nonmental, however the differences are to be precisely described. Especially important is the first-person epistemic authority noted earlier: We seem to have special access to our own mental states—or at least to an important subclass of them if not all of them. Such access may well fall short of infallibility or incorrigibility, and it seems beyond doubt that our minds are not wholly transparent to us. But the differences we have noted, even if they are not quite the way described, are real enough, and they may be capable of serving as a starting point for thinking about our concepts of the mental and the physical. It may be that we get our initial purchase on the concept of mentality through the core class of mental states for which some form of special first-person authority holds and that we derive the broader class of mental phenomena by extending and generalizing this core in various ways.^{[11](#)}

Mentality as Nonspatial

For Descartes, the essential nature of a mind is that it is a thinking thing (“*res cogitans*”), and the essential nature of a material thing is that it is a spatially extended thing—something with a three-dimensional bulk. A corollary of this, for Descartes, is that the mental is essentially nonspatial and the material is essentially lacking in the capacity for thinking. Most physicalists would reject this corollary even if they accept the thesis that the mental is definable as thinking; they will say that as it happens, some material things, like higher biological organisms, can think, feel, and be conscious. But there may be a way of developing the idea that the mental is nonspatial that leaves the question of physicalism open.

For example, we might try something like this: To say that M is a mental property is to say that the proposition that something has M *does not logically imply* that it is a spatially extended thing. This allows the possibility that something that has M is in fact a spatially extended thing, though it is not required to be. So it may be that *as a matter of contingent fact*, all things that have mental properties are spatially extended things, like human beings and other biological organisms.

Thus, from the proposition that something x believes that four is an even number, it does not seem to follow that x is a spatially extended thing. There may be no immaterial angels in this world, but it does not seem logically contradictory to say that there are angels or that angels have beliefs and other mental states, like desires and hopes. But it evidently is a contradiction to say that something has a physical property—say, the color red, a triangular shape, or a rough texture—and at the same time to deny that it is something with spatial extensions. What about being located at a geometric point? Or *being* a geometric point, for that matter? But no physical *thing* is a geometric point; geometric points are not physical objects, and no physical object has the property of being a point or being located wholly at a point in space.

How useful is this nonspatiality approach toward a mark of the mental? It would seem that if you take this approach seriously, you must also take the idea of immaterial mental substance seriously. For you must allow the existence of possible worlds in which mental properties are instantiated by nonphysical beings (beings without spatial extension). The reasoning leading to this conclusion is straightforward: Any mental property M is such that something can instantiate M without being spatially extended—that is, without being a physical thing. So M can be instantiated even if there is no physical thing. It follows then that there is a possible world in which mental properties, like belief and pain, are instantiated, even though no physical things exist in that world. What objects are there in such a world to serve as instantiators, or bearers, of mental properties? Since it makes no sense to think of abstract objects, like numbers, as possessors of mental properties, the only remaining possibility seems to be immaterial mental substances. It follows then that anyone who accepts the criterion of the mental as nonspatial must accept the idea of immaterial substance as a coherent one and allow the possible existence of such substances. This means that if you have qualms about the coherence of the Cartesian conception of minds as mental substances (see chapter 2), you would be well advised to stay away from the nonspatiality criterion of mentality.

Intentionality as a Criterion of the Mental

Schliemann sought the site of Troy. He was lucky; he found it. Ponce de León sought the Fountain of Youth, but he never found it. He could not have found it, since it does not exist and never did. It remains true, though, that he looked for the Fountain of Youth with great tenacity. The nonexistence of Bigfoot or the Loch Ness Monster has not prevented people from looking for them. Not only can you *look for* something that does not exist, but you can apparently also *think about, have beliefs and desires about, write about,* and even *worship* a nonexistent object. Even if God should not exist, he could be, and has been, the object of these mental acts or attitudes on the part of many people. Contrast these mental acts and states with physical ones, like cutting, kicking, and being to the left of. You cannot cut a nonexistent piece of wood, kick nonexistent tires, or be to the left of a nonexistent tree. That you kick something logically entails that the thing exists. That you are thinking of some object does not entail its existence. Or so it seems.

The Austrian philosopher Franz Brentano called this feature “the intentional inexistence” of psychological phenomena, claiming that it is this characteristic that separates the mental from the physical. In a famous passage, he wrote:

Every mental phenomenon is characterized by what the Scholastics of the Middle Ages called the intentional (or mental) inexistence of an object, and what we might call, though not wholly unambiguously, reference to a content, direction toward an object (which is not to be understood here as meaning a thing), or immanent objectivity. Every mental phenomenon includes something as object within itself, although they do not all do so in the same way. In presentation something is presented, in judgement something is affirmed or denied, in love loved, in hate hated, in desire desired and so on.¹²

This feature of the mental—namely, that mental states are *about*, or are *directed upon*, objects that may or may not exist or have contents that may or may not be true—has been called “intentionality.”

Broadly speaking, intentionality refers to the mind-world relation—specifically, the fact that our thoughts relate to, or hook up with, the things in the world, and represent how things are in the world. The idea at bottom is the thought that mentality is the capacity for representing the world around us, and that this is one of its essential functions. In short, the mind is a repository of inner representations—an inner mirror—of the outer world. The concept of intentionality may be subdivided into *referential intentionality* and *content intentionality*. Referential intentionality concerns the *aboutness* or *reference* of our thoughts, beliefs, intentions, and the like. When Ludwig Wittgenstein asked, “What makes my image of him into an image of *him*?”¹³ he was asking for an explanation of what makes it the case that a given mental state (my “image of him”) is *about*, or *refers to*, a particular object—him—rather than someone else. (That person may have an identical twin, and your image may fit his twin just as well, perhaps even better, but your image is of him, not of his twin. You may not even know that he has a twin.) Our words, too, refer to, or are directed upon, objects; “Mount Everest” refers to Mount Everest, and “horse” refers to horses.

Content intentionality concerns the fact that, as we saw, an important class of mental states—that is, propositional attitudes such as beliefs, hopes, and intentions—have contents or meanings, which are often expressed by full sentences. It is in virtue of having contents that our mental states *represent* states of affairs in the world. My perceiving that there are sunflowers in the field represents the fact, or state of affairs, of there being sunflowers in the field, and your remembering that there was a thunderstorm last night represents the state of affairs of there having been a thunderstorm last night. The capacity of our

mental states to represent things external to them—that is, the fact that they have *representational content*—is clearly a very important fact about them. Obviously, our capacity to have representations of the outside world is critical to our ability to cope with our environment and survive and prosper. In short, it is what makes it possible for us to have knowledge of the world. On a standard account, having knowledge is a matter of having mental representations with true contents—that is, representations that correctly represent.

Thus, referential intentionality and content intentionality are two related aspects of the fact that mental states have the capacity, and function, of representing things and states of affairs in the world. Brentano's thought seems to be that this representational capacity is the essence of the mind. It is the mind's essential function and *raison d'être*.

But can intentionality serve as the defining characteristic for all of mentality? Concerning the idea of representation, there is one point we must keep in mind: A representation has “satisfaction” conditions. In the case of representations with content intentionality, like the belief that snow is white, they can be evaluated in terms of truth or correctness. Pictorial or visual representations can be evaluated in terms of degrees of accuracy and fidelity. That means that a representation may fail to correctly represent—that is, it can misrepresent. In the case of referential intentionality, like “London” and “the Fountain of Youth,” we can talk about their successfully referring to the intended object—an existing object. “London” refers to the city London, whereas “the Fountain of Youth” turned out to refer to nothing.

With this preliminary out of the way, there are two issues about intentionality as a criterion of the mental we need to discuss. The first is that some mental phenomena—in particular, bodily sensations like pains and tickles and orgasms¹⁴—do not seem to exhibit either kind of intentionality. The sensation of pain does not seem to be “about,” or to refer to, anything; nor does it have a content that can be true or false, accurate or inaccurate. Doesn't the pain in my knee “mean,” or “represent,” the fact that I have strained the torn ligament again? But the sense of “meaning” involved here seems something like causal indication; the pain “means” a damaged ligament in the same sense in which your nice new suntan means that you spent the weekend on the beach. *Prima facie*, many bodily sensations don't seem to be evaluable in terms of truth or correctness. Or consider moods, like being bored, feeling low and blue, feeling upbeat, and the like. Do they represent anything? Can they be accurate or inaccurate? However, the view that all states of consciousness, including bodily sensations, are representational in nature has recently been gaining in popularity and influence, and we will revisit this issue later (in chapters 9 and 10).

Second, it may be observed that minds, or mental states, are not the only things that exhibit intentionality. Languages, in particular words and sentences, refer to things and have representational contents. The word “London” refers to London, and the sentence “London is large” refers to, or represents, the fact, or state of affairs, that London is large. A string of zeros and ones in a computer data structure can mean your name and address, and such strings are ultimately electronic states of a physical system. If these physical items and states are capable of reference and content, how can intentionality be considered an exclusive property of mentality?

The following line of reply seems open, however. As some have argued, we might distinguish between *genuine*, or *intrinsic*, intentionality, which our minds and mental states possess, and *as-if*, or *derivative*, intentionality, which we attribute to objects and states that do not have intentionality in their own right.¹⁵ When I say that my computer printer “likes” to work with Windows XP but not with Windows Vista, I am not really saying that my printer has likes and dislikes. It is at best an “as-if” or metaphorical use of language, and no one will take my statement to imply the presence of mentality in the machine. And it seems not implausible to argue that the word “London” refers to London only because language users use the word to refer to London. If we used it to refer to Paris, it would mean Paris, not London. Or if the inscription “London” were not a word in a language, it would just be meaningless scribbles with no referential function. Similarly, the sentence “London is large” represents the state of affairs it represents

only because speakers of English use this sentence to represent that state of affairs—for example, in affirming this sentence, they express the belief that London is large. The point, then, is that the intentionality of language is derived from the intentionality of language users and their mental processes. It is the latter that have intrinsic intentionality, intentionality that is not derived from, or borrowed from, anything else. Or so one could argue.¹⁶

A more direct reply would be this: To the extent that some physical systems can be said to refer to things, represent states of affairs, and deal in meanings, they should be considered as exhibiting mentality, at least one essential form of it. No doubt, as the first reply indicates, analogical or metaphorical uses of intentional idioms abound, but this fact should not blind us to the possibility that physical systems and their states might possess genuine intentionality and hence mentality. After all, it might be argued, we are complex physical systems ourselves, and the physical-biological states of our brains are capable of referring to things and states of affairs external to them and of storing their representations in memory. Of course, it may turn out not to be possible for purely physical states to have such capabilities, but that would only show that they are not capable of mentality. It remains true, the reply goes, that intentionality is at least a sufficient condition for mentality.

A Question

In surveying these candidates for “the mark of the mental,” we realize that our notion of the mental is far from unified and monolithic and that it is in fact a cluster of many ideas. Some of the ideas are fairly closely related to one another, but others appear independent of each other. (Why should there be a connection between special epistemic access and nonspatiality?) The diversity and possible lack of unity in our conception of the mental would imply that the class of things and states that we classify as mental may be a varied and heterogeneous lot. It is standardly thought that there are two broad categories of mental phenomena: first, conscious states, in particular sensory or qualitative states (those with “qualia”), like pains and sensings of colors and textures, and, second, intentional states, states with representational contents, like beliefs, desires, and intentions. The former seem to be paradigm cases of states that satisfy the epistemic criteria of the mental, such as direct access and privacy, and the latter are the prime examples of mental states that satisfy the intentionality criterion. An important question that is still open is this: In virtue of *what common property* are both sensory states and intentional states “mental”? What do pains and beliefs have in common in virtue of which they both fall under the single rubric of “mental phenomena”?

There are two approaches that might yield an answer—and a unified conception of mentality. Some have argued that consciousness is fundamental, and that it is presupposed by intentionality—in particular, that all intentional states are either conscious or in principle possible to become conscious.¹⁷ Along the same line, one might urge that only beings with consciousness are capable of having thoughts with content and intentionality. Such a view opens the possibility that all mentality is at bottom anchored in consciousness, and that consciousness is the single foundation of minds.

In direct opposition to this, there is the increasingly influential view, mentioned above, that all consciousness, including phenomenal consciousness, is representational in character. It is held that it is of the essence of conscious states that they represent things to be in a certain way, and that this is no less the case with bodily sensations, like pain, than with perceptual experiences like seeing a green vase on the table. This would mean that all conscious states have representational, or intentional, contents and are “directed upon” the objects and properties represented. Representationalism about consciousness, therefore, leads to the view that intentionality is the single mark characterizing all mentality. Thus, one potential bonus from consciousness representationalism could be a satisfying unified concept of minds and mentality.

FOR FURTHER READING

Readers interested in philosophical issues of cognitive science may explore Andy Clark, *Mindware: An Introduction to the Philosophy of Cognitive Science* ; Barbara von Eckardt, *What Is Cognitive Science?*; Robert M. Harnish, *Minds, Brains, Computers: An Historical Introduction to the Foundations of Cognitive Science*. Also useful are two anthologies: *Minds, Brains, and Computers* , edited by Denise Dellarosa Cummins and Robert Cummins; *Readings in Philosophy and Cognitive Science*, edited by Alvin Goldman.

The Oxford Handbook of Philosophy of Mind, edited by Brian McLaughlin, Ansgar Beckermann, and Sven Walter, is a comprehensive and highly useful reference work. The following general encyclopedias of philosophy feature many fine articles (some with extensive bibliographies) on topics in philosophy of mind and related fields: *Stanford Encyclopedia of Philosophy* (<http://plato.stanford.edu>); *Macmillan Encyclopedia of Philosophy*, second edition, edited by Donald Borchert; and *Routledge Encyclopedia of Philosophy*, edited by Edward Craig. The “Mind & Cognitive Science” section of *Philosophy Compass* (www.blackwell-compass.com) includes many fine up-to-date surveys of current research on a variety of topics in philosophy of mind. *The Internet Encyclopedia of Philosophy* (www.iep.utm.edu) has many helpful entries in its “Mind & Cognitive Science” section. In general, however, readers should exercise proper caution when consulting Web resources.

There are many good general anthologies on philosophy of mind. To mention a sample: *The Philosophy of Mind*, edited by Brian Beakley and Peter Ludlow; *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers; *Problems in Mind*, edited by Jack S. Crumley II; *Philosophy of Mind: A Guide and Anthology*, edited by John Heil; *Mind and Cognition: An Anthology*, third edition, edited by William G. Lycan and Jesse Prinz; *Philosophy of Mind: Contemporary Readings*, edited by Timothy O'Connor and David Robb.

NOTES

- [1](#) For details see Brian McLaughlin and Karen Bennett, “Supervenience.”
- [2](#) Sometimes this version of supervenience is formulated as follows: “Any minimal physical duplicate of this world is a duplicate *simpliciter* of this world.” See, for example, Frank Jackson, “Finding the Mind in the Natural World.” The point of the qualifier “minimal” is to exclude the following kind of situation: Consider a world that is like ours in all physical respects but in addition contains ectoplasms and immaterial spirits. (We are assuming these things do not exist in the actual world.) There is a sense in which this world and our world are physically alike, but they are clearly not alike overall. A case like this is ruled out by the qualifier “minimal” because this strange world is not a minimal physical duplicate of our world.
- [3](#) On characterizing physicalism, see Alyssa Ney, “Defining Physicalism.”
- [4](#) Also called “ontological physicalism.”
- [5](#) Nonreductive physicalism, as a form of physicalism, also includes mind-body supervenience; property dualism as such is not committed to supervenience. In fact, Cartesian substance dualism entails property dualism.
- [6](#) We should keep in mind the possibility that these philosophers who accept supervenience but reject reducibility are just mistaken.
- [7](#) For more on token and type physicalism, see Jaegwon Kim, “The Very Idea of Token Physicalism.”
- [8](#) Why they are called “intentional” states is not simple to explain or motivate; it is best taken simply as part of philosophical terminology. If you insist on an explanation, the following might help: These states, in virtue of their contents, are *representational* states; the belief that snow is white represents the world as being a certain way—more specifically, it represents the state of affairs of snow being white. Traditionally, the term “intentionality” has been used to refer to this sort of representational character of mental states. More to follow on intentionality below.
- [9](#) René Descartes, *Meditations on First Philosophy*, Meditation II.
- [10](#) Later in the book (chapter 8) you will encounter another sense of “transparency” applied to perceptual experiences.
- [11](#) It is worth noting that many psychologists and cognitive scientists take a dim view of the claim that we have specially privileged access to the contents of our minds. See, for example, Richard Nisbett and Timothy Wilson, “Telling More Than We Can Know,” and Alison Gopnik, “How We Know Our Minds: The Illusion of First-Person Knowledge of Intentionality.”
- [12](#) Franz Brentano, *Psychology from an Empirical Standpoint*, p. 88.
- [13](#) Ludwig Wittgenstein, *Philosophical Investigations*, p. 177.
- [14](#) This example is taken from Ned Block.
- [15](#) See, for example, John Searle, *Intentionality* and *The Rediscovery of the Mind*.
- [16](#) This point has been disputed. Other possible positions are these: First, one might hold that linguistic intentionality is in fact prior to mental intentionality, the latter being derivative from the former (Wilfrid Sellars); second, we might claim that the two types of intentionality are distinct but interdependent, neither being prior to the other and neither being derivable from the other (Donald Davidson); and third, some have argued that the very distinction between “intrinsic” and “derivative” intentionality is bogus and incoherent (Daniel Dennett).
- [17](#) John Searle is a well-known advocate of this claim; see his *The Rediscovery of Mind*, chapter 7. See also Galen Strawson, “Real Intentionality 3: Why Intentionality Entails Consciousness.”

CHAPTER 2

Mind as Immaterial Substance

Descartes's Dualism

What is it for something to “have a mind,” or “have mentality”? When the ancients reflected on the contrast between us and mindless creatures, they sometimes described the difference in terms of having a “soul.” For example, according to Plato, each of us has a soul that is simple, divine, and immutable, unlike our bodies, which are composite and perishable. In fact, before we were born into this world, our souls preexisted in a pure, disembodied state, and on Plato’s doctrine of recollection, what we call “learning” is merely a process of recollecting what we already knew in our prenatal existence as pure souls. Bodies are merely vehicles of our existence in this earthly world, a transitory stage in our soul’s eternal journey. The idea, then, is that because each of us has a soul, we are the kind of conscious, intelligent, and rational creatures that we are. Strictly speaking, we do not really “have” souls, since we are literally *identical with* our souls—that is, each of us *is* a soul. My soul is the thing that I am. Each of us “has a mind,” therefore, because each of us *is* a mind.

For most of us, Plato’s story is probably a bit too speculative, too fantastical, to take seriously as a real possibility. However, many of us seem to have internalized a kind of mind-body dualism according to which, although each of us has a body that is fully material, we also have a mental or spiritual dimension that no “mere” material things can have. When we see the term “material,” we are apt to think “not mental” or “not spiritual,” and when we see the term “mental,” we tend to think “not material” or “not physical.” This may not amount to a clearly delineated point of view, but it seems fair to say that some such dualism of the mental and the material is entrenched in our ordinary thinking, and that dualism is a kind of “folk” theory of our nature as creatures with minds.

But folk dualism often goes beyond a mere duality of mental and physical properties, activities, and processes. It is part of folklore in many cultures and of most established religions that, as Plato claimed, each of us has a soul, or spirit, that survives bodily death and decay, and that we are really our souls, not our bodies, in that when our bodies die we continue to exist in virtue of the fact that our souls continue to exist. Your soul defines your identity as an individual person; as long as it exists—and only so long as it exists—you exist. And it is our souls in which our mentality inheres; thoughts, consciousness, rational will, and other mental acts, functions, and capacities belong to souls, not to material bodies. Ultimately, to have a mind, or to be a creature with mentality, is to have a soul.

In this chapter, we examine a theory of mind, due to the seventeenth-century French philosopher René Descartes, which develops a view of this kind. One caveat before we begin: Our goal here is not so much a scholarly exegesis of Descartes as it is an examination of a point of view closely associated with him. As with other great philosophers, the interpretation of what Descartes “really” said, or meant to say, continues to be controversial. For this reason, the dualist view of mind we will discuss is better regarded as Cartesian rather than as the historical Descartes’s.

DESCARTES'S INTERACTIONIST SUBSTANCE DUALISM

The dualist view of persons that Descartes defended is a form of substance dualism (sometimes called substantial, or substantival, dualism). Substance dualism is the thesis that there are substances of two fundamentally distinct kinds in this world, namely, minds and bodies—or mental stuff and material stuff—and that a human person is a composite entity consisting of a mind and a body, each of which is an entity in its own right. Dualism of this form contrasts with monism, according to which all things in the world are substances of one kind. We later encounter various forms of material monism that hold that our world is fundamentally material, consisting only of bits of matter and complex structures made up of bits of matter, all behaving in accordance with physical laws. This is materialism, or physicalism. (The terms “materialism” and “physicalism” are often used interchangeably, although there are subtle differences: We can think of physicalism as a contemporary successor to materialism—materialism informed by modern physics.) There is also a mental version of monism, unhelpfully called idealism. This is the view that minds, or mental items at any rate (“ideas”), constitute the fundamental reality of the world, and that material things are mere “constructs” out of thoughts and mental experiences. This form of monism has not been very much in evidence for some time, though there are reputable philosophers who still defend it.¹ We will not be further concerned with mental monism in this book.

So substance dualism maintains that minds and bodies are two different sorts of substance. But what is a substance? Traditionally, two ideas have been closely associated with the concept of a substance. First, a substance is something in which properties “inhere”; that is, it is what has, or instantiates, properties.² Consider this celadon vase on my table. It is something that has properties, like weight, shape, color, and volume; it is also fragile and elegant. But a substance is not in turn something that other things can exemplify or instantiate; nothing can have, or instantiate, the vase as a property. Linguistically, this idea is sometimes expressed by saying that a substance is the subject of predication, something to which we can attribute predicates like “blue,” “weighs a pound,” and “fragile,” while it cannot in turn be predicated of anything else.

Second, and this is more important for us, a substance is thought to be something that has the capacity for independent existence. Descartes himself wrote, “The notion of a substance is just this—that it can exist by itself, that is without the aid of any other substance.”³ What does this mean? Consider the vase and the pencil holder to its right. Either can exist without the other existing; we can conceive the vase as existing without the pencil holder existing, and vice versa. In fact, we can, it seems, conceive of a world in which only the vase (with all its constituent parts) exists and nothing else, and a world in which only the pencil holder exists and nothing else. It is in this sense that a substance is capable of independent existence. This means that if my mind is a substance, it can exist without any body existing, or any other mind existing. Consider the vase again: There is an intuitively intelligible sense in which its color and shape cannot exist apart from the vase, whereas the vase is something that exists in its own right. (The color and shape would be “modes” that belong to the vase.) The same seems to hold when we compare the vase and its surface. Surfaces are “dependent entities,” as some would say; their existence depends on the existence of the objects of which they are surfaces, whereas an object could exist without the particular surface it happens to have at a given time. As was noted, there is a possible world of which the vase is the sole inhabitant. Compare the evidently absurd claim that there is a possible world in which the surface of the vase exists but nothing else; in fact, there is no possible world in which only surfaces exist and nothing else. For surfaces to exist they must be surfaces of some objects—existing objects.⁴

Thus, the thesis that minds are substances implies that minds are objects, or things, in their own right; in this respect, they are like material objects—it’s only that, on Descartes’s view, they are immaterial objects. They have properties and engage in activities of various sorts, like thinking, sensing, judging, and

willing. Most important, they are capable of independent existence, and this means that there is a possible world in which only minds exist and nothing else—in particular, no material bodies. So my mind, as a substance, can exist apart from my body, and so of course could your mind even if your body perished.

Let us put down the major tenets of Cartesian substance dualism:

1. There are substances of two fundamentally different kinds in the world, mental substances and material substances—or minds and bodies. The essential nature of a mind is to think, be conscious, and engage in other mental activities; the essence of a body is to have spatial extensions (a bulk) and be located in space.

2. A human person is a composite being (a “union,” as Descartes called it) of a mind and a body.

3. Minds are diverse from bodies; no mind is identical with a body.

What distinguishes Descartes’s philosophy of mind from the positions of many of his contemporaries, including Leibniz, Malebranche, and Spinoza, is his eminently commonsensical belief that minds and bodies are in causal interaction with each other. When we perform a voluntary action, the mind causes the body to move in appropriate ways, as when my desire for water causes my hand to reach for a glass of water. In perception, causation works in the opposite direction: When we see a tree, the tree causes in us a visual experience as of a tree. That is the difference between seeing a tree and merely imagining or hallucinating one. Thus, we have the following thesis of mind-body causal interaction:

4. Minds and bodies causally influence each other. Some mental phenomena are causes of physical phenomena and vice versa.

The only way we can influence the objects and events around us, as far as we know, is first to move our limbs or vocal cords in appropriate ways and thereby start a chain of events culminating in the effects we desire—like opening a window, retrieving a hat from the roof, or starting a war. But as we will see, it is this most plausible thesis of mind-body causal interaction that brought down Cartesian dualism. The question was not whether the interactionist thesis was in itself acceptable; rather, the main question was whether it was compatible with the radical dualism of minds and bodies—that is, whether minds and bodies, sundered apart by the dualist theses (1) and (3), could be brought together in causal interaction as claimed in (4).

WHY MINDS AND BODIES ARE DISTINCT: SOME ARGUMENTS

Before we consider the supposed difficulties for Descartes's interactionist dualism, let us first consider some arguments that apparently favor the dualist thesis that minds are distinct from bodies. Most of the arguments we will consider are Cartesian—some of them perhaps only vaguely so—in the sense that they can be traced one way or another to Descartes's *Second* and *Sixth Meditations* and that all are at least Cartesian in spirit. It is not claimed, however, that these are in fact the arguments that Descartes offered or that they were among the considerations that moved Descartes to advocate substance dualism. You might want to know first of all why anyone would think of minds as substances—why we should countenance minds as objects or things in addition to people and creatures with mentality. As we will see, some of the arguments do address this issue, though not directly.

At the outset of his *Second Meditation*, Descartes offers his famous “cogito” argument. As every student of philosophy knows, the argument goes “I think, therefore I exist.” This inference convinces him that he can be absolutely certain about his own existence; his existence is one perfectly indubitable bit of knowledge he has, or so he is led to think. Now that he knows he exists, he wonders what kind of thing he is, asking, “But what then am I?” Good question! Knowing that you exist is not to know very much; it has little content. So what kind of being is Descartes? He answers: “A thinking thing” (“sum res cogitans”). How does he know that? Because he has proved his existence from the premise that he thinks; it is through his knowledge of himself as a thinker that he knows that he exists. To get on with his dualist arguments we will grant him the proposition that he is a thinking thing, namely a mind. The main remaining issue for him, and for us, is the question whether the thinking thing can be his body—that is, why we should not take his body, perhaps his brain, as the thing that does the thinking.

We first consider three arguments based on epistemological considerations. The simplest—perhaps a bit simplistic—argument of this form would be something like this:

Argument 1

I am such that my existence cannot be doubted.

My body is not such that its existence cannot be doubted.

Therefore, I am not identical with my body.

Therefore, the thinking thing that I am, that is, my mind, is not identical with my body.

This argument is based on the apparent asymmetry between knowledge of one's own existence and knowledge of one's body's existence: While I cannot doubt that I exist, I can doubt that my body exists. We could also put the point this way: As the cogito argument shows, I can be absolutely certain that I exist, but my knowledge that my body exists, or that I have a body, does not enjoy the same degree of certainty. I must make observations to know that I have a body, and such observations could go astray. We leave it to the reader to evaluate this argument.

According to Descartes, I am a "thinking thing." What does this mean? He says that a thinking thing is "a thing that doubts, understands, affirms, denies, is willing, is unwilling, and also imagines and has sensory perceptions."⁵ For Descartes, then, "thinking" is a generic term, roughly meaning "mental activity," and specific mental states and activities, like believing, doubting, affirming, reasoning, sensing a color, hearing a sound, experiencing a pain, and the rest fall under the broad rubric of thinking. In Descartes's own terms, thinking is the general essence of minds, and the specific kinds of mental activities and states are its various "modes."

Our second epistemological argument exploits another related difference between our knowledge of our own minds and our knowledge of our bodies.

Argument 2

My mind is transparent to me—that is, nothing can be in my mind without my knowing that it is there. My body is not transparent to me in the same way. Therefore, my mind is not identical with my body.

As stated, the first premise is quite strong and likely not to be entirely true. Most of us would be prepared to acknowledge that at least some of our beliefs, desires, and emotions are beyond our cognitive reach—that is, that there are “unconscious” or “subconscious” mental states, like suppressed beliefs and desires, angers and resentments, of which we are unaware. This, however, doesn’t seem like a big problem: The premise can be stated in a weaker form, to claim only that my mind is transparent at least with respect to *some* of the events that occur in it. This weaker premise suffices as long as we understand the second premise as asserting that *none* of my bodily events have this transparent character. To find out any fact about my body, I must make observations and sometimes make inferences from the evidence gained through observations. Often a third party—my physician or dentist—is in a better position to know the conditions of my body.

We now consider our last epistemological argument for substance dualism:

Argument 3

Each mind is such that there is a unique subject who has direct access to its contents.

No material body has a specially privileged knower—knowledge of material things is in principle public and intersubjective.

Therefore, minds are not identical with material bodies.

We are said to know something “directly” when the knowledge is not based on evidence, or inferred from other things we know. When knowledge is direct, like my knowledge of my toothache, it makes no sense to ask, “How do you know?” The present argument exploits this difference between knowledge of minds and knowledge of bodies: For each mind, there is a unique person who is in a privileged epistemic position, whereas this is not the case with bodies. It is in this sense that knowledge of our own minds is said to be “subjective.” In contrast, knowledge of bodies is said to be “objective”—different observers can in principle have equal access to such knowledge. Thus, the present argument can be called the argument from the subjectivity of minds.

What should we think of these arguments? We will not formulate and develop specific objections and difficulties, or discuss how the dualist might respond; that is left to the reader. But one observation is in order: It is widely believed that there is a problem with using epistemic (or more broadly, “intentional”) properties to differentiate things. To show that $X \neq Y$, it is necessary and sufficient to come up with a single property P such that X has P but Y lacks it, or Y has P but X lacks it. Such a property P can be called a *differential property* for X and Y . The question, then, is whether epistemic properties, like being known with certainty (or an intentional property like being believed to be such and such), can be used as a differential property. Consider the property of being known to the police to be the hit-and-run driver. The man who sped away in a black SUV is known to the police to be the hit-and-run driver. The man who drove away in a black SUV is identical with my neighbor, and yet my neighbor is not known to the police to be the hit-and-run driver (or else the police would have him in custody already). The epistemic properties invoked in the three arguments are not the same—or exactly of the same sort—as the one just used. It is fair to say that the last of the arguments presented above, the argument from subjectivity, seems the most compelling, and anyone wishing to reject it should have good reasons.

We now turn to metaphysical arguments, which instead of appealing to epistemic differences between minds and bodies attempt to invoke real metaphysical differences between them. Throughout the *Second* and *Sixth Meditations*, there are constant references to the essence of mind as thinking and the essence of body as being extended in space. By extension in space Descartes means three-dimensional extension, that is, bulk. Surfaces or geometric lines do not count as material substances; only things that have a bulk count as such. A simple argument could be formulated in terms of essences or essential natures, like this:

Argument 4

My essential nature is to be a thinking thing.

My body's essential nature is to be an extended thing in space.

My essential nature does not include being an extended thing in space.

Therefore, I am not identical with my body. And since I am a thinking thing (namely a mind), my mind is not identical with my body.

How could the first and third premises be defended? Perhaps a Cartesian dualist could make two points in defense of the first premise. First, as the “cogito” argument shows, I know that I exist only insofar as I am a thinking thing, and this means that my existence is inseparably tied to the fact that I am a thinking thing. Second, an essential nature of something is a property without which the thing cannot exist; when something loses its essential nature, that is when it ceases to exist. Precisely in this sense, being a thinking thing is my essential nature; when I cease to be a thinking thing, that is, a being with a capacity for thought and consciousness, that is when I cease to be, and so long as I am a thinking thing, I exist. On the other hand, I can conceive of myself as existing without a body; there is no inherent incoherence, or contradiction, in the idea of my disembodied existence, whereas it seems manifestly incoherent to think of myself as existing without a capacity to think and have conscious experience. Hence, being an extended object in space is not part of my essential nature.

What should we think of this argument? Some will question how the third line of the argument might be established, pointing out that all Descartes shows is that our disembodied existence is conceivable, or imaginable. But from the fact that something is *conceivable*, however clearly and vividly, it does not follow that it is *really possible*. A body moving at a speed exceeding the speed of light is conceivable, but we know it is not possible.⁶ Or consider this: We seem to be able to conceive how Goldbach's conjecture, the proposition that every even number greater than two is the sum of two prime numbers, might turn out to be true, and also to conceive how it might turn out to be false. But Goldbach's conjecture, being a mathematical proposition, is necessarily true if true, and necessarily false if false. So it cannot be both possibly true and possibly false. (To the reader: Why?) But if conceivability entails possibility, it would have to be possibly true and possibly false. This issue about conceivability and real possibility has led to an extended series of debates too complex to enter into here.⁷ It is a live current issue in modal metaphysics and epistemology. We should note, though, that unless we use reflective and carefully scrutinized conceivability as a guide to possibility, it is difficult to know what other resources we can call on when we try to determine what is possible and what is not, what is necessarily the case and what is only contingently so, and other such modal questions.

Let us say that something is “essentially” or “necessarily” F, where “F” denotes a property, just in case whenever or wherever it exists (or in any possible world in which it exists), it is F. In this sense, we are presumably essentially persons, but not essentially students or teachers; for we cannot continue to exist while ceasing to be persons, whereas we could cease to be students, or teachers, without ceasing to exist. In the terminology of the preceding paragraph, for something to have property F essentially or necessarily is to have F as part of its essential nature. Consider, then, the following argument:

Argument 5

If anything is material, it is essentially material.

However, I am possibly immaterial—that is, there is a possible world in which I exist without a body.

Hence, I am not essentially material.

Hence, it follows (with the first premise) that I am not material.

This is an interesting argument. There seems to be a lot to be said for the first premise. Take something material, say, a bronze bust of Beethoven: This object could perhaps exist without being a bust of Beethoven—it could have been fashioned into a bust of Brahms. In fact, it could exist without being a bust of anyone; it could be melted down and made into a doorstop. If transmutation of matter were possible (surely this is not something a priori impossible), it could even exist without being bronze. But could this statue exist without being a material thing? The answer seems a clear no. If anything is a material object, being material is part of its essential nature; it cannot exist without being a material thing. So it appears that the acceptability of the argument depends crucially on the acceptability of the second premise. Is it possible that I exist without a body? That surely is conceivable, Descartes would insist. But again, is something possible just because it is conceivable? Can we say more about the possibility of our disembodied existence?

Consider the bronze bust again. There is here a piece of sculpture and a quantity of bronze. Is the sculpture the very same thing as the bronze? Many philosophers would say no: Although the two share many properties in common (such as weight, density, and location), they differ at least in one respect, namely, their persistence condition. If the bust is melted down and shaped into a cube, the bust is gone but the bronze continues to exist. According to the next dualist argument, my body and I differ in a somewhat similar way.

Argument 6

Suppose I am identical with this body of mine.

In 2001 this body did not exist.

Hence, from the first premise, it follows that I did not exist in 2001.

But I existed in 2001.

Hence, a contradiction, and the supposition must be false.

Hence, I am not identical with my body.

In 2001 this body did not exist because all the molecules making up a human body are completely cycled out every six or seven years. When all the molecular constituents of a material thing are replaced, we have a new material thing. The body that I now have shares no constituents with the body I had in 2001. The person that I am, however, persists through changes of material constituents. So even if I have to have some material body or other to exist, I do not have to have any particular body. But if I am identical with a body, I must be identical with some particular body and when this body goes, so go I. That is the argument. (This probably was not one of Descartes's actual arguments.)

An initial response to this argument could run as follows: When I say I am identical with this body of mine, I do not mean that I am identical with the "time slice"—that is, a temporal cross section—of my body at this instant. What I mean is that I am identical with the temporally elongated "worm" of a three-dimensional organism that came into existence at my birth and will cease to exist when my biological death occurs. This four-dimensional object—a three-dimensional object stretched along the temporal dimension—has different material constituents at different times, but it is a clearly delineated system with a substantial unity and integrity. It is this material structure with a history with which I claim I am identical. (To the reader: How might a Cartesian dualist reformulate the argument in answer to this objection?)

Another reply, related to the first, could go as follows: My body is not a mere assemblage or structure made up of material particles; rather, it is a biological organism, a human animal. And the persistence condition appropriate to mere material things is not necessarily appropriate for animals. In fact, animals can retain their identities even though the matter constituting them changes over time (this may well be true of all living things, including plants), just as in the case of persons. The criterion of identity over time for animals (however it is to be spelled out in detail) is the one that should be applied to human bodies.⁸ Does the substance dualist have a reply to this? I believe an answer may be implicit in the next argument we consider.

Tully is the same person as Cicero. There is one person here, not two. Can there be a time at which Tully exists but not Cicero? Obviously not—that is no more possible than for Tully to be at a place where Cicero is not. Given that Cicero = Tully in this world, is there a possible world in which Cicero is not identical with Tully? That is, given that Cicero is Tully, is it *possible* that Cicero is not Tully? Suppose there is a possible world in which Cicero \neq Tully; call it W. Since Cicero \neq Tully in W, there must be some property, F, such that, in W, Cicero has it but Tully does not. Let's say that F is the property of being tall. So in W, Cicero is tall but Tully isn't. But how is that possible? Here in this world is a single person, called both Cicero and Tully. How is it possible for this one person to be tall and at the same time not tall in world W? That surely is an impossibility, and world W is not a possible world. In fact, there is no possible world in which Cicero \neq Tully. We therefore have the following principle ("NI" for "necessity of identities"):

(NI) If $X = Y$, then necessarily $X = Y$ —that is, if $X = Y$ in this world, $X = Y$ in every possible world.

(NI) is special in that in general it is not the case that if a proposition is true, it is necessarily true. For example, I am standing; from this it does not follow that necessarily I am standing, for I could be sitting.

Given the principle (NI), we can formulate another dualist argument: [9](#)

Argument 7

Suppose I am identical with this body of mine.

Then, by (NI), I am necessarily identical with this body—that is, I am identical with it in every possible world.

But that is false, for (a) in some possible worlds I could be disembodied and have no body; or at least (b) I could have a *different* body in another possible world.

So it is false that I am identical with this body in every possible world, and this contradicts the second line.

Therefore, I am not identical with my body.

The principle (NI) is considered unexceptionable. So if there is a vulnerability in this argument, it would have to be the third line; to criticize this premise effectively, we would have to eliminate both (a) and (b) as possibilities. As we have seen, (a) is vulnerable to criticism; however, (b) may be less so. John Locke's well-known story of the prince and the cobbler can be taken as supporting (b); Locke writes:

Should the soul of a prince, carrying with it the consciousness of the prince's past life, enter and inform the body of a cobbler, as soon as deserted by his own soul, everyone sees he would be the same person with the prince, accountable only for the prince's action.... Had I the same consciousness that I saw the ark and Noah's flood, as that I saw an overflowing of the Thames last winter, or as that I write now, I could no more doubt that I who write this now, that saw the Thames overflowed last winter, that viewed the flood at the general deluge, was the same *self* ... than that I who write this am the same *myself* now whilst I write ... that I was yesterday.¹⁰

For Locke, then, consciousness, not body, defines a person, a self; the continuity of my consciousness determines my persistence as a person. What body I have, or whether I have a body at all, is immaterial. To defeat this dualist argument, therefore, we must show that Locke's story of the prince and the cobbler is an impossibility—it isn't something that could happen. This will require some ingenuity and creative thinking.

The leading idea driving all of these metaphysical arguments is the thought that although I may be a composite being consisting of a mind and a body, my relation to my mind is more intimate and essential than my relation to my body and that I am “really” my mind and could not exist apart from it, while it is a contingent fact that I have the body that I happen to have. Descartes's interest in defending minds as immaterial substances was apparently motivated in part by his desire to allow for the possibility of survival after bodily death.¹¹ Most established religions have a story to tell about the afterlife, and the conceptions of an afterlife in some of them seem to require, or at least allow, the possibility of our existence without a body. But all that is a wish list; it does not make the possibility of our disembodied existence a real one (Descartes was under no such illusion). The arguments we have looked at must earn their plausibility on their own merits, not from the allure of their conclusions.

We will now consider our final metaphysical argument for substance dualism. As we will see, this argument is rather difficult to articulate clearly, but it enjoys the allegiance of some well-known and well-respected philosophers, so it is worth a serious look. The skeletal structure of the argument can be set out like this:

Argument 8

Thoughts and consciousness exist.

Hence, there must be objects, or substances, to which thoughts and consciousness occur—that is, things that think and are conscious.

Thoughts and consciousness cannot occur to material things—they cannot be states of material objects, like the brain.

Hence, thoughts and consciousness must occur to immaterial things, like Cartesian mental substances.

Hence, mental substances exist and they are the things that think and are conscious, and bear other mental properties.

Some would question the move from the first to the second line—the assumption that thoughts and consciousness, and, more generally, states and properties, require “bearers,” things to which they occur, or in which they inhere; this, however, is a general metaphysical issue and it will be tedious and out of place to pursue it here. Moreover, the crucial premise is staring us in the face—it is the third line, the proposition that material things, like the human brain, are unfit to serve as bearers of thoughts and consciousness. Think about numbers, like three and fifteen: Numbers aren’t the sort of thing that can have colors like blue or red, or occupy a location in space, or be transparent or opaque. Or think about events, like earthquakes or wildfires. They can be sudden, severe, and destructive; but events aren’t the sort of thing that can be soluble in water, divisible by four, or weigh ten tons. The claim then is that there is an essential incongruity between mental states, like thoughts and consciousness, on one hand and material things on the other, so that the former cannot inhere in, or occur to, the latter, just as weight and color cannot inhere in numbers. If our thoughts and consciousness cannot occur to anything material, including our brains, then they must occur to immaterial things, or Cartesian minds. Only immaterial things can be conscious and have thoughts. Since we are conscious and have thoughts, we must be immaterial minds.

But why can’t consciousness, thoughts, and other mental states occur to material things? It is often thought that Leibniz was first to give an argument, or at least hint at one, why that must be so:

It must be confessed, moreover, that *perception*, and that which depends on it, *are inexplicable by mechanical causes*, that is by figures and motions. Supposing there were a machine so constructed as to think, feel and have perception, we could conceive of it as enlarged and yet preserving the same proportions, so that we might enter it as into a mill. And this granted, we should only find, on visiting it, pieces which push one against another, but never anything by which to explain a perception. This must be sought for, therefore, in the simple substance and not in the composite or in the machine.¹²

Leibniz appears to be saying that a material thing is at bottom a mechanical system in which the parts causally interact with one another (“pieces pushing one against another”), and it is not possible to see anything in this picture that would account for the presence of thought or consciousness. This is not altered when a more sophisticated modern picture of what goes on in a complex biological system, like a human brain, replaces Leibniz’s mill: What we have is still a large assemblage of microscopic material things, molecules and atoms and particles, interacting with one another in accordance with laws of physics and chemistry, producing further scenes of such interactions. Nowhere in this picture do we see a thought or perception or consciousness; molecules jostling and bumping against one another is all the action that is taking place. Again, if this picture looks unsophisticated, replace it with the most sophisticated scientific picture you know, and see if that invalidates Leibniz’s point.

Is this all one can say in defense of the Leibnizian proposition that material systems are just the wrong

kind of thing to bear thoughts and other mental states? It might be helpful to consider what some philosophers have said to defend this proposition. Alvin Plantinga, referring to the Leibniz paragraph above, writes:

Leibniz's claim is that thinking can't arise by virtue of physical interaction among objects or parts of objects. According to current science, electrons and quarks are simple, without parts. Presumably neither can think—neither can adopt propositional attitudes; neither can believe, doubt, hope, want, or fear. But then a proton composed of quarks won't be able to think either, at least by way of physical relations between its component quarks, and the same will go for an atom composed of protons and electrons, and a molecule composed of atoms, a cell composed of molecules, and an organ (e.g., a brain) composed of cells. If electrons and quarks can't think, we won't find anything composed of them that *can* think by way of the physical interaction of its parts.¹³

Does this reading of Leibniz shed new light on his argument and make it seem more plausible? It is something to ponder. Some, for example the emergentists, will argue that thoughts and consciousness arise in material systems when they reach higher levels of organizational complexity, and that from the fact that the constituent parts of a system lack a certain property it does not follow that the system itself must lack that property.

Another philosopher, John Foster, who holds the view that subjects of mentality must be “wholly nonphysical,” argues:

If something is just an ordinary material object, whose essential nature is purely physical, there seems to be no way of understanding how it could be [the subject] of mentality.... If something is merely a material object, any understanding of how it is equipped to be a mental subject will presumably have to be achieved by focusing on its physical nature. But focusing on an object's physical nature will only reveal how it is equipped to be in states or engage in activities which are directly to do with its possession of that nature—with its condition as a physical thing.... Focusing on the physical nature of an object simply offers no clue as to how it can be the basic subject of the kinds of mentality which the dualist postulates.¹⁴

Perhaps some readers will find these quotations helpful and clarifying; others may not. In any case, one question we should ask at this point is this: Is it any easier to understand how thoughts and consciousness can arise in an immaterial substance, especially if, as Leibniz and many other dualists urge, such a substance is an absolute “simple” with no constituent parts? How could immaterial minds, without structure and outside physical space, possess beliefs and desires directed at things in the physical world? How could our rich and complex mental life inhere in something that has no parts and hence no structure? Isn't the proposal recommended by Leibniz, and by Plantinga and Foster, merely a solution by stipulation? What do we know about mental substances that can help us understand how they could be the bearers of consciousness and perception and thought? Understanding how mentality can arise in something immaterial may be no easier than understanding how it could arise in a material system; in fact, it might turn out to be more difficult.

As was mentioned above, it is not easy to make clear the thoughts that lie behind Argument 8, in particular its crucial third line. However, this is an intriguing and influential line of dualist thinking, and readers are urged to reflect on it.¹⁵

PRINCESS ELISABETH AGAINST DESCARTES

As will be recalled, the fourth component of Descartes's dualism is the thesis that minds and bodies causally influence each other. In voluntary action, the mind's volition causes our limbs to move; in perception, physical stimuli impinging on sensory receptors cause perceptual experiences in the mind. This view is not only commonsensical but also absolutely essential to our conception of ourselves as agents and cognizers: Unless our minds, in virtue of having certain desires, beliefs, and intentions, are able to cause our bodies to move in appropriate ways, how could human agency be possible? How could we be agents who act and take responsibility for our actions? If objects and events in the physical world do not cause us to have perceptual experiences and beliefs, how could we have any knowledge of what is happening around us? How could we know that we are holding a tomato in our hand, that we are coming up on a stop sign, or that a large bear is approaching from our left?

Descartes has something to say about how mental causation works. In the *Sixth Meditation*, he writes:

The mind is not immediately affected by all parts of the body, but only by the brain, or perhaps just by one small part of the brain.... Every time this part of the brain is in a given state, it presents the same signals to the mind, even though the other parts of the body maybe in a different condition at the time.... For example, when the nerves in the foot are set in motion in a violent and unusual manner, this motion, by way of the spinal cord, reaches the inner parts of the brain, and there gives the mind its signal for having a certain sensation, namely the sensation of a pain as occurring in the foot. This stimulates the mind to do its best to get rid of the cause of the pain, which it takes to be harmful to the foot.¹⁶

In *The Passions of the Soul*, Descartes identifies the pineal gland as the “seat of the soul,” the locus of direct mind-body interaction. This gland, Descartes maintains, can be moved directly by the soul, thereby moving the “animal spirits” (bodily fluids in the nerves), which then transmit causal influence to appropriate parts of the body:

And the activity of the soul consists entirely in the fact that simply by willing something it brings it about that the little gland to which it is closely joined moves in the manner required to produce the effect corresponding to this desire.¹⁷

In the case of physical-to-mental causation, this process is reversed: Disturbances in the animal spirits surrounding the pineal gland make the gland move, which in turn causes the mind to experience appropriate sensations and perceptions. For Descartes, then, each of us as an embodied human person is a “union” or “intermingling” of a mind and a body in direct causal interaction.

In what must be one of the most celebrated letters in the history of philosophy, Princess Elisabeth of Bohemia, an immensely astute pupil of Descartes's, wrote to him in May 1643, challenging him to explain

how the mind of a human being, being only a thinking substance, can determine the bodily spirits in producing bodily actions. For it appears that all determination of movement is produced by the pushing of the thing being moved, by the manner in which it is pushed by that which moves it, or else by the qualification and figure of the surface of the latter. Contact is required for the first two conditions, and extension for the third. [But] you entirely exclude the latter from the notion you have of the soul, and the former seems incompatible with an immaterial thing.¹⁸

(For “determine,” read “cause”; for “bodily spirits,” read “fluids in the nerves and muscles.”) Elisabeth’s demand is clearly understandable. First, see what Descartes has said about bodies and their motion in the *Second Meditation*:

By a body I understand whatever has determinate shape and a definable location and can occupy a space in such a way as to exclude any other body; it can be perceived by touch, sight, hearing, taste or smell, and can be moved in various ways, not by itself but by whatever else comes into contact with it.¹⁹

For Descartes, minds are immaterial; that is, minds have no spatial extension and are not located in physical space. If bodies can be moved only by contact, how could an unextended mind, which is not even in space, come into contact with an extended material thing, even the finest and lightest particles in animal spirits, thereby causing it to move? This seems like a perfectly reasonable question.

In modern terminology we can put Elisabeth’s challenge as follows: For anything to cause a physical object to move, or cause any change in one, there must be a flow of energy, or transfer of momentum, from the cause to the physical object. But how could there be an energy flow from an immaterial mind to a material thing? What kind of energy could it be? How could anything “flow” from something *outside space* to something *in space*? If an object is going to impart momentum to another, it must have mass and velocity. But how could an unextended mind outside physical space have either mass or velocity? The question does not concern the intrinsic plausibility of Descartes’s thesis of mind-body interaction; the question is whether this commonsensical interactionist thesis is tenable within Descartes’s dualist ontology of nonspatial immaterial minds and material things in the space-time world.

Descartes responded to Elisabeth in a letter written in the same month:

I observe that there are in us certain primitive notions which are, as it were the originals on the pattern of which we form all of other thoughts, ... as regards the mind and body together, we have only the primitive notion of their union, on which depends our notion of the mind’s power to move the body, and the body’s power to act on the mind and cause sensations and passions.²⁰

Descartes is defending the position that the idea of mind-body union is a “primitive” notion—a fundamental notion that is intelligible in its own right and cannot be explained in terms of other more basic notions—and that the idea of mind-body causation depends on that of mind-body union. What does this mean? Although on Descartes’s view, minds and bodies seem on an equal footing causally, there is an important asymmetry between them: My mind can exercise its causal powers—on other minds as well as on bodies around me—only by first causally influencing my own body, and nothing can causally affect my mind except through its causal influence on my body. But my body is different: It can causally interact with other bodies quite independently of my mind. My body—or my pineal gland—is the necessary causal conduit between my mind and the rest of the world; in a sense, my mind is causally isolated from the world by being united with my body. To put it another way, my body is the enabler of my mind’s causal powers; it is by being united with my body that my mind can exercise its causal powers in the world—on other minds as well as on other bodies. Looked at this way, the idea of mind-body union does seem essential to understanding the mind’s causal powers.

Elisabeth is not satisfied. She immediately fires back:

And I admit that it would be easier for me to concede matter and extension to the mind than it would be for me to concede the capacity to move a body and be moved by one to an immaterial thing.²¹

This is a remarkable statement; it may well be the first appearance of the causal argument for materialism (see chapter 4). For she is in effect saying that to allow for the possibility of mental causation, she would rather accept materialism concerning the mind (“it would be easier to concede matter and extension to the mind”) than accept what she regards as an implausible dualist account offered by her mentor.

Why should anyone find Descartes’s story so implausible? A couple of paragraphs back, it was pointed out that my mind’s forming a “union” with my body amounts to the fact that my body serves as a necessary and omnipresent proximate cause and effect of changes in my mind and that my body is what makes it possible for my mind to have a causal influence on the outside world. Descartes, however, would reject this characterization of a mind-body union, for the simple reason that it would beg the question as far as the possibility of mind-body causation is concerned. That is presumably why Descartes claimed that the notion of mind-body union is a “primitive”—one that is intelligible per se but is neither further explainable nor in need of an explanation. Should this answer have satisfied Elisabeth, or anyone else? A plausible case can be made for a negative answer. For when we ask what makes this body my body, not someone else’s, a causal answer seems the most natural one and the only correct one. This is my body because it is the only body that I, or my desires and volitions, can directly move—that is, without moving or causally influencing anything else, whereas I can move other bodies, like this pen on my desk or the door to the hallway, only by moving my body first. Moreover, to cause any changes in my mind—or my mental states—you must first bring about appropriate changes in my body (presumably in my brain). What could be a more natural account of how my mind and my body form a “union”? But this explanation of mind-body union presupposes the possibility of mind-body causation, and it would be circular to turn around and say that an understanding of mind-body causation “depends” on the idea of mind-body union. Descartes’s declaration that the idea of a union is a “primitive” and hence not in need of an explanation is unlikely to impress someone seeking an understanding of mental causation; it is liable to strike his critics simply as a dodge—a refusal to acknowledge a deep difficulty confronting his approach.

THE “PAIRING PROBLEM”: ANOTHER CAUSAL ARGUMENT

We will develop another causal argument against Cartesian substance dualism. If this argument works, it will show not only that immaterial minds cannot causally interact with material things situated in space but also that they are not able to enter into causal relations with anything else, including other immaterial minds. Immaterial objects would be causally impotent and hence explanatorily useless; positing them would be philosophically unmotivated.

Here is the argument.²² To set up an analogy and a point of reference, let us begin with an example of physical causation. A gun, call it *A*, is fired, and this causes the death of a person, *X*. Another gun, *B*, is fired at the same time (say, in *A*'s vicinity, but this is unimportant), and this results in the death of another person, *Y*. What makes it the case that the firing of *A* caused *X*'s death and the firing of *B* caused *Y*'s death, and not the other way around? That is, why did *A*'s firing not cause *Y*'s death and *B*'s firing not cause *X*'s death? What principle governs the “pairing” of the right cause with the right effect? There must be a relation *R* that grounds and explains the cause-effect pairings, a relation that holds between *A*'s firing and *X*'s death and also between *B*'s firing and *Y*'s death, but not between *A*'s firing and *Y*'s death or between *B*'s firing and *X*'s death. What is this *R*, the “pairing relation,” as we might call it? We are not necessarily supposing that there is a single such *R* for all cases of physical causation, only that some relation must ground the fact that a given cause is a cause of the particular effect that is caused by it.

Two ideas come to mind. First, there is the idea of a *causal chain*: There is a continuous causal chain connecting *A*'s firing with *X*'s death, as there is one connecting *B*'s firing with *Y*'s death, whereas no such chains exist between *A*'s firing and *Y*'s death or between *B*'s firing and *X*'s death. Indeed, with a highspeed video camera, we could trace the bullet's flight from each gun to its impact point on the target. The second idea is the thought that each gun when it fired was at a certain distance and in appropriate orientation in relation to the person it hit, but not to the other person. That is, *spatial relations* do the job of pairing causes with their effects.

A moment's reflection shows that the causal chain idea does not work as an independent solution to the problem. A causal chain, after all, is a series of events related as cause to effect, and interpolating more cause-effect pairs does not solve the pairing problem. For obviously it begs the question: What pairing relations ground these interpolated cause-effect pairs? It seems plausible that ultimately spatial relations—and more broadly, spatiotemporal relations—are the only way of generating pairing relations. Space appears to have nice causal properties; for example, as distance increases, causal influence diminishes, and it is often possible to set up barriers at intermediate positions to block or impede the propagation of causal influence. In any case, the following proposition seems highly plausible:

(M) It is metaphysically possible for there to be two distinct physical objects, *a* and *b*, with the same intrinsic properties and hence the same causal potential or powers; one of these, say, *a*, causes a third object, *c*, to change in a certain way, but object *b* has no causal influence on *c*.

The fact that *a* but not *b* causes *c* to change must be grounded in some fact about *a*, *b*, and *c*. Since *a* and *b* have the same intrinsic properties, it must be their *relational properties* with respect to *c* that provide the desired explanation of their different causal roles. What relational properties or relations can do this job? It is plausible to think that when *a*, *b*, and *c* are physical objects, it is the spatial relation between *a* and *c* and that between *b* and *c* that are responsible for the causal difference between *a* and *b* vis-à-vis *c*. (The object *a* was in the right spatial relation to *c*; *b* was “too far away” to exert any influence on it.) At least, there seems no other obvious candidate that comes to mind. Later we give an explanation of what it is about spatial relations that enables them to do the job.

Consider the possibility of immaterial souls, outside physical space, causally interacting with material objects in space. The following companion principle to (M) seems equally plausible, and if an interactionist substance dualist wishes to reject it, she should give a principled explanation why.

(M*) It is metaphysically possible for there to be two souls, A and B, with the same intrinsic properties²³ such that they both act in a certain way at the same time and as a result a material object, C, undergoes a change. Moreover, it is the action of A, not that of B, that is the cause of the physical change in C.

What makes it the case that this is so? What pairing relation pairs the first soul, but not the second soul, with the material object? Since souls, as immaterial substances, are outside physical space and cannot bear spatial relations to anything, it is not possible to invoke spatial relations to ground the pairing. What possible relations could provide causal pairings across the two domains, one of spatially located material things and the other of immaterial minds outside space?

Consider a variation on the foregoing example: There are two physical objects, P_1 and P_2 , with the same intrinsic properties, and an action of an immaterial soul causally affects one of them, say, P_1 , but not P_2 . How can we explain this? Since P_1 and P_2 have identical intrinsic properties, they must have the same causal capacity (“passive” causal powers as well as “active” causal powers), and it would seem that the only way to make them discernible in a causal context is their relations to other things. Doesn’t that mean that any pairing relation that can do the job must be a spatial relation? If so, the pairing problem for this case is unsolvable since the soul is not in space and bears no spatial relation to anything. The soul cannot be any “nearer” to, or “more properly oriented” toward, one physical object than another. Nor could we say that there was a causal barrier “between” the soul and one of the physical objects but not the other, for what could “between” mean as applied to something in space and something outside it? It is a total mystery what nonspatial relations there could be that might help distinguish, from the point of view of an immaterial soul, between two intrinsically indiscernible physical objects.

Could there be causal interactions among immaterial substances? Ruling out mind-body causal interaction does not in itself rule out the possibility of a causally autonomous domain of immaterial minds in which minds are in causal commerce with other minds. Perhaps that is the picture of a purely spiritual afterlife envisioned in some religions and theologies. Is that a possibility? The pairing problem makes such an idea a dubious proposition. Again, any substance dualist who wants causation in the immaterial realm must allow the possibility of there being three mental substances, M_1 , M_2 , and M_3 , such that M_1 and M_2 have the same intrinsic properties, and hence the same causal powers, and yet an action by M_1 , but not the same action by M_2 at the same time, is causally responsible for a change in M_3 . If such is a metaphysically possible situation, what pairing relation could connect M_1 with M_3 but not M_2 with M_3 ? If causation is to be possible within the mental domain, there must be an intelligible and motivated answer to this question. But what mental relations could serve this purpose? It is difficult to think of any.

Consider what space does for physical causation. In the kind of picture envisaged, where a physical thing or event causally acts on only one of the two objects with identical intrinsic properties, what distinguishes these two objects has to be their spatial locations with respect to the cause. Space provides a “principle of individuation” for material objects. Pure qualities and causal powers do not. And what enables space to serve this role is the fact that physical objects occupying exactly the same location in space at the same time are one and the same object.²⁴ This is in effect the venerable principle of “impenetrability of matter,” which can usefully be understood as a sort of “exclusion” principle for space: Material things compete for, and exclude one another from, spatial locations. From this it follows that if physical objects a and b bear the same spatial relations to a third object c , a and b are one and the same

object. This principle is what enables space to individuate material things with identical intrinsic properties. The same goes for causation in the mental domain. What is needed to solve the pairing problem for immaterial minds is a kind of mental coordinate system, a “mental space,” in which these minds are each given a unique “location” at a time. Further, a principle of “impenetrability of minds” must hold in this mental coordinate system; that is, minds that occupy the same “location” in this space must be one and the same. It seems fair to say that we do not have any idea how a mental space of this kind could be constructed. Moreover, even if we could develop such a space for immaterial minds, that still would fall short of a complete solution to the pairing problem; to solve it for causal relations across the mental and physical domains, we need to somehow coordinate or fuse the two spaces, the mental and the physical, to yield unitary pairing relations across the domains. It is not clear that we have any idea where to begin.

If there are Cartesian minds, therefore, they are threatened with total causal isolation—from each other as well as from the material world. The considerations presented do not show that causal relations cannot hold within a single mental substance (even Leibniz, famous for disallowing causation between monads, allowed it within a single monad). However, what has been shown seems to raise serious challenges for substance dualism. If this is right, we have a causal argument for a physicalist ontology. Causality requires a spacelike structure, and as far as we know, the physical domain is the only domain with a structure of that kind.

IMMATERIAL MINDS IN SPACE?

All these difficulties with the pairing problem arise because of the radically nonspatial nature of minds in traditional substance dualism. According to Descartes, not only do minds lack spatial extension but also they are not in space at all. So why not bring minds into space, enabling them to have spatial locations and thereby solve the pairing problem? Most popular notions of minds as immaterial spirits do not seem to conceive them as wholly nonspatial. For example, when a person dies, her soul is thought to “rise” from the body, or otherwise “leave” it, implying that before the death the soul was inside the body and that the soul is capable of moving in space and changing its locations. Sometimes the departed souls of our loved ones are thought to be able to make their presence known to us in various ways, including in a visible form (think about Hamlet’s ghostly father). It is probably impossible to make coherent sense of these popular ideas, but is there anything in principle wrong with locating immaterial minds in physical space and thereby making it possible for them to participate in the causal transactions of the world?

As we will see, the proposal to bring immaterial minds into space is fraught with complications and difficulties and probably not worth considering as an option. First there is the question of just where in space to put them. Is there a principled and motivated way of assigning a location to each soul? We might suggest that I locate my soul in my body, you locate your soul in your body, and so on. That may sound like a natural and reasonable suggestion, but it faces a number of difficulties. First, what about disembodied souls, souls that are not “united” with a body? Since souls are supposed to be substances in their own right, such souls are metaphysically possible. Second, if your soul is located in your body, exactly where in your body is it located? In the brain, we might reply. But exactly where in the brain? It could not be spread all over the brain because minds are not supposed to be extended in space. If it has a location, the location has to be a geometric point. Is it coherent to think that there is a geometric point somewhere in your brain at which your mind is located? Descartes called the pineal gland the “seat of the soul,” presumably because the pineal gland is where mind-body causal interaction was supposed to take place, although of course his official doctrine was that the soul is not in space at all.

Following Descartes’s strategy here, however, does not seem to make much sense. For one thing, there is no evidence that there is any single place in the brain—a dimensionless point at that—at which mind-body interaction takes place. As far as we know, various mental states and activities are distributed over the entire brain and nervous system, and it does not make scientific sense to think, as Descartes did in regard to the pineal gland, that there is a single identifiable organ responsible for all mind-body causal interaction. Second, how could an entity occupying a single geometric point cause all the physical changes in the brain that are involved in mind-body causation? By what mechanism could this happen? How is energy transmitted from this geometric point to the neural fibers making up the brain? And there is this further question: What keeps the soul at that particular location? When I stand up from my chair in the study and go downstairs to the living room, somehow my soul tags along and moves exactly on the same trajectory as my body. When I board an airplane and the airplane accelerates on the runway and takes off, somehow my pointlike immaterial mind manages to gain speed exactly at the same rate and begins to cruise at the speed of 560 miles an hour! It seems that the soul is somehow firmly glued to some part of my brain and moves as my brain moves, and when I die it miraculously unglues itself from my body and migrates to a better (or perhaps worse) place in the afterlife. Does any of this make sense? Descartes was wise, we must conclude, to keep immaterial minds wholly outside physical space.

In any case, giving locations to immaterial minds will not in itself solve the pairing problem. As we saw, spatial locations of physical objects help solve the pairing problem in virtue of the principle that physical objects can be individuated in terms of their locations. As was noted, this is the principle of impenetrability of matter: Distinct objects exclude one another from spatial regions. That is how the causal roles of two intrinsically indiscernible physical objects could be differentiated. For the spatial

locations of immaterial minds to help, therefore, we need a similar principle of spatial exclusion for immaterial minds—or the principle of impenetrability of mental substance—to the effect that distinct minds cannot occupy exactly the same point in space. What reason is there to think such a principle holds? Why cannot a single point be occupied by all the souls that exist, like the thousand angels dancing on the head of a pin? Such a principle is needed if we are to make sense of causation for spatially located pointlike souls. But this does not mean that the principle is available; we must be able to produce independently plausible evidence or give a credible argument to show that the principle holds.

When we see all the difficulties and puzzles to which the idea of an immaterial mind, or soul, appears to lead, it is understandable why Descartes declared the notion of mind-body union to be primitive and not further explainable in terms of more fundamental ideas. Even a contemporary writer has invoked God and theology to make sense of how a particular mind (say, your mind) gets united to a particular body (your brain).²⁵ The reader is urged to think about whether such an appeal to theology gives us real help with the problems the dualist faces.

SUBSTANCE DUALISM AND PROPERTY DUALISM

It has seemed to most contemporary philosophers that the concept of mind as a mental substance is fraught with too many difficulties and puzzles without compensating explanatory gains. In addition, the idea of an immaterial and immortal soul usually carries with it various, often conflicting, religious and theological associations and aspirations that many of us would rather avoid in philosophical contexts. For example, the traditional conception of the soul involves a sharp and unbridgeable gap between humans and the rest of animal life. Even if our own mentality could be explained as consisting in the possession of a soul, what might explain the mentality of nonhuman animals? It is not surprising that substance dualism has not been a prominent alternative in contemporary philosophy of mind. But there is no call to exclude it *a priori*, without serious discussion; some highly reputable and respected philosophers continue to defend it as a realistic—perhaps the only—option (see “For Further Reading”).

To reject the substantial view of mentality is not to deny that each of us “has a mind”; it is only that we should not think of “having a mind” literally—that is, as there being some object or substance called a “mind” that we literally possess. As discussed earlier (in chapter 1), having a mind is not like—at least, it need not be like—having brown eyes or a good throwing arm. To have brown eyes, there must be brown eyes that you have. To “be out of your mind” or to “keep something in mind,” you do not have to *have* some object—namely, a mind—which you are out of, or in which you keep something. If you have set aside substance dualism, at least for now, you can take having a mind simply as having a certain special set of *properties*, *capacities*, and *characteristics*, something that humans and some higher animals possess but sticks and stones do not. To say that something “has a mind” is to classify it as a certain sort of thing—as a thing with capacities for certain characteristic sorts of behavior and functions, such as sensation, perception, memory, learning, consciousness, and goal-directed action. For this reason, it is less misleading to speak of “having mentality” than “having a mind.” (As you will recall, this is what the last dualist argument we considered above, “Leibniz’s mill,” challenges; the point of the argument is precisely that no material system can have mentality.)

In any case, substance dualism has played a small role in contemporary philosophy of mind. Philosophical attention has focused instead on mental activities and functions—or mental events, states, and processes—and the mind-body problem has turned into the problem of understanding how these mental events, states, and processes are related to physical and biological events, states, and processes, or how our mental or psychological capacities and functions are related to the nature of our physical structure and capacities. In regard to this question, there are two principal positions: *property dualism* and *reductive physicalism* (also called *type physicalism*). Dualism is no longer a dualism of two sorts of substances; it is now a dualism of two sorts of properties, mental and physical. “Property” is used here in a broad sense: Mental properties comprise mental functions, capacities, events, states, and the like, and similarly for physical properties. It is a catchall term referring to events, activities, states, and the rest. So property dualism is the view that mental properties are diverse from and irreducible to physical properties. In contrast, reductive physicalism defends the position that mental properties are reducible to, and therefore can be identified with, physical properties. As we will see, there are various forms of both property dualism and reductionist physicalism. However, they all share one thing in common: the rejection of immaterial minds. Contemporary property dualism and reductive physicalism acknowledge only objects of one kind in the world—bits of matter and increasingly complex structures aggregated out of bits of matter. (This anti-Cartesian position is called substance physicalism.) Some of these physical systems exhibit complex behaviors and activities, like perceiving, sensing, reasoning, and consciousness. But these are only properties of material structures. The main point of dispute concerns the nature of the relationship between these mental features and activities on one hand and the structures’ physical

characteristics on the other. This is the central question for the remainder of this book.

FOR FURTHER READING

The primary source of Descartes's dualism is his *Meditations on First Philosophy*, first published in 1641. See especially Meditations II and VI. There are numerous English editions; a good version (including *Objections and Replies*) can be found in *The Philosophical Writings of Descartes*, vol. 2, translated and edited by John Cottingham, Robert Stoothoff, and Dugald Murdoch. Helpful historical and interpretive literature on Descartes's philosophy of mind includes: Daniel Garber, *Descartes Embodied* (especially chapter 8, "Understanding Causal Interaction: What Descartes Should Have Told Elisabeth"); Marleen Rozemond, *Descartes's Dualism*, chapter 1; and Lilli Alanen, *Descartes's Concept of Mind*, chapter 2.

On the pairing problem, see Kim, *Physicalism, or Something Near Enough*, chapter 3. For dualist responses, John Foster, "A Defense of Dualism"; Andrew Baily, Joshua Rasmussen, and Luke Van Horn, "No Pairing Problem."

For some contemporary defenses of substance dualism, see John Foster, *The Immaterial Self*; W. D. Hart, *The Engines of the Soul*; William Hasker, *The Emergent Self*; E. J. Lowe, "Non-Cartesian Substance Dualism and the Problem of Mental Causation," and "Dualism"; Alvin Plantinga, "Against Materialism"; Dean Zimmerman, "Material People"; and Richard Swinburne, *The Evolution of the Soul*.

Also recommended are Noa Latham, "Substance Physicalism," and Tim Crane, "Mental Substances."

NOTES

- [1](#) See, for example, John Foster, *The Case for Idealism*.
- [2](#) Descartes writes: "Substance: this term applies to every thing in which whatever we perceive immediately resides, as in a subject.... By 'whatever we perceive' is meant any property, quality or attribute of which we have a real idea." See "Author's Replies to the Second Set of Objections," p. 114.
- [3](#) René Descartes, "Author's Replies to the Fourth Set of Objections," p. 159.
- [4](#) Many philosophers in Descartes's time, including Descartes himself, held that, strictly speaking, God is the only being capable of independent existence and therefore that the only true substance is God, all others being "secondary" or "derivative" substances.
- [5](#) René Descartes, *Meditations on First Philosophy*, Meditation II, p. 19.
- [6](#) One might say that this is only a case of physical possibility and necessity, not possibility and necessity *tout court*. A more standard example would be the proposition that water = H₂O. It is widely accepted that this is a necessary truth (though a posteriori) but that its falsehood is conceivable.
- [7](#) See some of the essays in *Conceivability and Possibility*, edited by Tamar Szabo Gendler and John Hawthorne. Gendler and Hawthorne's introduction is a good starting point.
- [8](#) This approach, called "animalism," has recently been receiving much attention. See, for example, Eric T. Olson, *The Human Animal: Personal Identity Without Psychology*.
- [9](#) Strictly, (NI) holds only when X and Y are "rigid designators." A name is said to be "rigid" just in case it names the same thing in every possible world in which it exists. In this sense, "Cicero" and "Tully," along with most proper names, are rigid. For details, see Saul Kripke, *Naming and Necessity*.
- [10](#) John Locke, *An Essay Concerning Human Understanding*, Book II, chapter 27, secs. 15, 16.
- [11](#) As noted by Marleen Rozemond in her *Descartes's Dualism*, p. 3.
- [12](#) Gottfried Leibniz, *Monadology*, 17.
- [13](#) Alvin Plantinga, "Against Materialism," p. 13.
- [14](#) John Foster, "A Brief Defense of the Cartesian View," pp. 25-26. "The kinds of mentality which the dualist postulates" refers to mentality conceived as irreducible to physical processes. Foster of course believes that mentality cannot be physically reduced; the point is that if mental states are reduced to, say, neural states of an organism, there would be no special problem about how material things can have mentality.
- [15](#) Functionalism (chapters 5, 6) can be seen as providing a story that explains how physical systems can have beliefs, desires, emotions, and so on. As we will see, functionalism construes mental states as "functional states," that is, states defined in terms of the causal work they perform. Such states are "realized" by states in physical systems and it is claimed that these physical realizers do the causal work required for intentional states. Thus, a physical system has a certain belief when one of its physical states realizes the belief. See also chapter 10 on David Chalmers on the "hard" and "easy" problems of consciousness. Dualists like Plantinga will reject the claim that mental states are functional states.
- [16](#) René Descartes, *Meditations on First Philosophy*, Meditation VI, pp. 59-60.
- [17](#) René Descartes, *The Passions of the Soul*, I, 41, p. 343.
- [18](#) Daniel Garber, "Understanding Interaction: What Descartes Should Have Told Elisabeth," p. 172. This and other quotations from the correspondence between Elisabeth and Descartes are taken from this chapter of Garber's book, *Descartes Embodied*.
- [19](#) René Descartes, *Meditations on First Philosophy*, Meditation II, p. 17.
- [20](#) Descartes to Princess Elisabeth, May 21, 1643, in Garber, *Descartes Embodied*, p. 173.
- [21](#) Princess Elisabeth to Descartes, June 1643, in Garber, *Descartes Embodied*, p. 172.

[22](#) For a fuller presentation of this argument, see Kim, *Physicalism, or Something Near Enough*, chapter 2. For some dualist responses, see the “For Further Reading” section.

[23](#) If you are inclined to invoke the identity of intrinsic indiscernibles for souls to dissipate the issue, the next situation we consider involves only one soul and this remedy does not apply. Moreover, the pairing problem can be generated without assuming that there can be distinct intrinsic indiscernibles. This assumption, however, helps to present the problem in a simple and compelling way.

[24](#) There is the familiar problem of the statue and the lump of clay of which it is composed (the problem of coincident objects). Some claim that although these occupy the same region of space and coincide in many of their properties (for example, weight, shape, size), they are distinct objects because their persistence conditions are different (for example, if the clay is molded into a cube, the clay, but not the statue, continues to exist). We must set this problem aside, but it does not affect our argument. Note that the statue and the lump of clay share the same causal powers and suffer the same causal fate (except perhaps coming into being and going out of existence).

[25](#) John Foster, “A Brief Defense of the Cartesian View.”