

# PHILOSOPHY OF MIND

THIRD EDITION

JAEGWON KIM



A Member of the Perseus Books Group

Westview Press was founded in 1975 in Boulder, Colorado, by notable publisher and intellectual Fred Praeger. Westview Press continues to publish scholarly titles and high-quality undergraduate- and graduate-level textbooks in core social science disciplines. With books developed, written, and edited with the needs of serious nonfiction readers, professors, and students in mind, Westview Press honors its long history of publishing books that matter.

Copyright © 2011 by Westview Press

Published by Westview Press,  
A Member of the Perseus Books Group

All rights reserved. Printed in the United States of America. No part of this book may be reproduced in any manner whatsoever without written permission except in the case of brief quotations embodied in critical articles and reviews. For information, address Westview Press, 2465 Central Avenue, Boulder, CO 80301.

Find us on the World Wide Web at [www.westviewpress.com](http://www.westviewpress.com).

Every effort has been made to secure required permissions for all text, images, maps, and other art reprinted in this volume.

Westview Press books are available at special discounts for bulk purchases in the United States by corporations, institutions, and other organizations. For more information, please contact the Special Markets Department at the Perseus Books Group, 2300 Chestnut Street, Suite 200, Philadelphia, PA 19103, or call (800) 810-4145, ext. 5000, or e-mail [special.markets@perseusbooks.com](mailto:special.markets@perseusbooks.com).

Designed by Trish Wilkinson  
Set in 10.5 point Minion Pro

Library of Congress Cataloging-in-Publication Data

Kim, Jaegwon.

Philosophy of mind / Jaegwon Kim.—3rd ed.

p. cm.

ISBN 978-0-8133-4458-4 (alk. paper)

1. Philosophy of mind. I. Title.

BD418.3.K54 2011

128'.2—dc22

E-book ISBN 978-0-8133-4520-8

2010040944

10 9 8 7 6 5 4 3 2 1

# 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.<sup>1</sup> 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.<sup>2</sup> 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.”<sup>3</sup> 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.<sup>4</sup>

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 “*co-gito*” 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."<sup>5</sup> 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 “*co-gito*” 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 mani-

festly 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.<sup>6</sup> 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.<sup>7</sup> 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 substantival 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.<sup>8</sup> 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:

### *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.<sup>10</sup>

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.<sup>11</sup> 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.<sup>12</sup>

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.<sup>13</sup>

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.<sup>14</sup>

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.<sup>15</sup>

## 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 may be 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.<sup>16</sup>

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.<sup>17</sup>

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.<sup>18</sup>

(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.<sup>19</sup>

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 non-spatial 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.<sup>20</sup>

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.<sup>21</sup>

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.<sup>22</sup> 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 high-speed 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 properties<sup>23</sup> 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.<sup>24</sup> 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).<sup>25</sup> 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 substantialist 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 ex-

istence 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 = H<sub>2</sub>O. 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.”