

ment. One might also raise the question why Chaos has properties that fit it for being a generator of random cosmoi. (It is, after all, to be understood by analogy to the quantum field, and the quantum field is a physical object with a very specific set of properties, properties defined in many cases by apparently arbitrary numbers.) This is a question that pertains to the teleological argument. If there were a large number of "possible chaoses" and only a very few of them had the right properties to be random cosmos-generators, we should have made no progress by appealing to Chaos in our attempt to deal with the teleological argument.

Since we know next to nothing relevant to this question, let us simply stipulate that the properties of Chaos are determined by "the laws of hyperphysics" and that there is only one possible set of laws of hyperphysics. (In one sense, the laws of hyperphysics would be the real laws of physics. What we normally think of as the laws of physics would not really be laws, since they would not apply to the whole of physical reality. They would be mere descriptions of local regularities.) It is important to remember in this connection that we are not attempting to construct an explanation for the apparent design of the cosmos that is known to be correct, but only an explanation at least as good as the hypothesis that the cosmos was designed by a rational being. And an explanation will be at least as good as that explanation if it contains no element known on independent grounds to be false or improbable—for that (together with the fact that it does explain the observed phenomenon of the fine-tuning of the cosmos) is really all that can be said in favor of the hypothesis of rational design.

The final part of our alternative explanation of apparent cosmic design takes the form of an appeal to an observational selection effect: we cannot observe the cosmoi that are unsuitable for life. We cannot observe them because something restricts the scope of our observations to our own cosmos—the space-time curvature of our cosmos, or the simple fact that everything not a part of our cosmos is just too far away for us to see, or some other factor.<sup>5</sup> And any rational beings there may be anywhere in the macrocosm (so to call the aggregate of all the cosmoi) will be in the same position as we. The members of every rational species anywhere in the macrocosm will be able to observe only the "insides" of their own cosmos. It will look to them as if they inhabited a cosmos carefully designed to be an abode for life (it will have to be suitable for life, since they are in it) and unique (since they can't see any of the others). But this will be appearance rather than reality. It will be the result of the interplay between chance and an observational selection effect.

We can gain some insight into the way this reply to the teleological argument works if we look again at the case in which you would have been killed if you had not drawn the shortest straw from a bundle of 1,048,576 straws. You were (you remember) astonished to find that, lo and behold, you had drawn the shortest straw and were still there to be astonished. I said the most reasonable thing for you to conclude would be that some unknown benefactor had rigged the drawing in your favor. And this is true—*provided* you know your situation is unique. But suppose you knew your situation was *not* unique. Suppose you knew you were only one among many millions of people who had been placed in a situation in which they had to draw the shortest straw from a bundle of 1,048,576 or else be instantly killed. In that case the most reasonable hypothesis would not be that an unknown benefactor had rigged the drawing in your favor. The most reasonable hypothesis would be that you were just lucky.

If someone tosses a coin exactly twenty times, and if there is no other occasion on which anyone tosses a coin twenty or more times, it is most improbable that anyone will ever toss a coin "heads" twenty times in a row: the chances are 1 in 1,048,576. If millions of people toss a coin twenty times, it is likely that someone will toss "heads" twenty times in a row. If millions of millions of people toss a coin twenty times, it is a virtual certainty that someone will toss "heads" twenty times in a row. The odds are, of course, the same with respect to drawing the shortest straw from a bundle of 1,048,576 straws. If your life depends on your drawing the shortest straw from a bundle of that size, and if you *do* draw the shortest straw in the bundle, and if you know that millions of millions of other people have been drawing straws from a bundle of that size, you should reason as follows: "It was all but inevitable that a fair number of people would draw the shortest straw. Luckily for me, I happened to be one of that number." But if you draw the shortest straw and know you were the only one engaged in such a drawing, you cannot reason this way, for you will know that it was *not* all but inevitable that a fair number of people would draw the shortest straw; on the contrary, you will know that it was all but inevitable that *no one* would draw the shortest straw. For just this reason, you will have to turn to the alternative hypothesis that someone has rigged the drawing in your favor.

Now given our everyday knowledge of the world, it would never be reasonable for one to believe that millions of millions of other people were in circumstances almost exactly duplicating one's own, and this would be true with a vengeance if one's circumstances were as bizarre as those laid out in the "straw-drawing" story. But suppose the conditions of human life were very different from what they are. Suppose you were in the circumstances of the hero of the straw-drawing story and *for all you knew* there were millions of millions of people in identical circumstances. Suppose nothing you knew gave you any reason to prefer either of the following two hypotheses to the other:

- This is the only drawing, and someone has rigged it in my favor.
- This is only one among millions of millions of such drawings (in some of which—a statistical certainty—the shortest straw is drawn).

What could you say then? Only this: that you didn't know whether you had an unobserved benefactor or whether you were "surrounded" by millions of millions of such drawings.

This, I believe, is exactly analogous to our situation with respect to the fine-tuning of the cosmos. As far as our present knowledge goes (aside from any divine revelations various individuals or groups may be privy to), we have no reason to prefer either of the following two hypotheses to the other:

- This is the only cosmos, and some rational being has (or rational beings have) fine-tuned it in such a way that it is a suitable abode for life.
- This is only one among a vast number of cosmoi (some of which are—a statistical certainty—suitable abodes for life).

We do not know whether the apparently purposive fine-tuning of the cosmos is reality or mere appearance, a product of chance and an observational selection effect.

The fine-tuning of the cosmos that has been pretty well established by modern physics and cosmology does, however, make one thing clear: the cosmos does not exist on its own. *Something* accounts for the fact that we observe a cosmos that has very specific features (those required for the existence of life) and, presumably, it accounts for the existence of that cosmos as well. Let us call this "something" the *Arche*. (This word—which is pronounced ar-KAY—was used by the Greeks for that which is the beginning of all things, or that which is the foundation on which the existence of all things rests.) But we—that is, we metaphysicians, speaking as metaphysicians—can say little about the *Arche* beyond the bare assertion of its existence.

The most important of the questions about the *Arche* that we should like to be able to answer but cannot may be phrased as follows: Is the *Arche* a Chaos or a Logos? We have already met the word 'chaos'. The difficult Greek word 'logos' is usually translated as 'word'—not in the sense of the basic unit of speech but in the sense of that which goes forth from a speaker, as 'Word then came to the king that . . . —or 'reason'. (I have chosen to use 'logos' in this context because 'arche' and 'logos' are associated with each other in the opening words of the Gospel according to Saint John, which are themselves a deliberate echo of the opening words of Genesis. John writes: "In the beginning [*arche*] was the Word [*logos*], and . . . the Word was God.") This question is the most important question that can be asked about the *Arche* because the question whether there is any meaning or purpose in our existence depends upon its answer: if the *Arche* is a Chaos, our existence has no meaning; if the *Arche* is a Logos, our existence has a meaning. If the *Arche* is a Chaos, we are just one of those things that happen from time to time. If the *Arche* is a Logos, we exist for a purpose, and that is the only sense in which the existence of anything can have a "meaning." (It is worth repeating that it does not follow from the assumption that we exist for a purpose that we can discover that purpose or are capable of understanding it or that we should like it if we knew what it was.)

There is perhaps nothing more the metaphysician can say about this question (whether the *Arche* is a Chaos or a Logos) except to point out that the answers "Chaos" and "Logos" can each be emotionally attractive to certain people. I suppose no one would deny that there are people who find the idea that our existence has a meaning or a purpose emotionally attractive. But it is equally worthy of remark, and less often noticed, that the thesis that our existence is without meaning or purpose is immensely attractive to certain kinds of persons. If my life has no purpose, if I and everyone else are the results of a series of accidents that are not part of the purpose of anyone or anything, I am free to live my life according to my own desires—or at any rate the only obstacles to my doing this will be those often inconvenient features of everyday life (people whose plans are incompatible with mine, the laws of physics, death—that sort of thing) with which everyone is familiar.<sup>7</sup> If, however, there is a rational being who has designed the universe for a purpose, who can say whether that being's plans and mine are compatible or what the consequences might be if they are not? Chaos, having brought us into existence, is unlikely to make any demands on our time or attention; who can tell what a Logos might have in mind for us? A special case of this sort of motivation is nicely expressed in a little verse by Arthur Hugh Clough:

"There is no God," the wicked saith,  
"And truly it's a blessing,

For what he might have done with us  
It's better only guessing."

But this motivation can be generalized beyond the case of a God who has moral concerns and punishes the wicked. If, moreover, "life, the universe, and everything" are the result of mere chance, one may well be one of the more important and impressive beings in existence; if the *Arche* is a Logos, this is certainly not the case. The desire to think of oneself as one of the most important beings in existence is transparently present in the following quotation from the Nobel Prize-winning physicist Steven Weinberg (it occurs immediately following a passage in which he discusses the emotional difficulties some people face in believing our existence to be meaningless, a position he believes is supported by modern science):

The more the universe seems comprehensible, the more it also seems pointless.

But if there is no solace in the fruits of our research, there is at least some consolation in the research itself. Men and women are not content to comfort themselves with tales of gods and giants, or to confine their thoughts to the daily affairs of life; they also build telescopes and satellites and accelerators, and sit at their desks for endless hours working out the meaning of the data they gather. The effort to understand the universe is one of the very few things that lifts human life a little above the level of farce, and gives it some of the grace of tragedy.<sup>8</sup>

It is pretty clear that these are the thoughts of someone who inhabits a mental world in which scientists are just about the most important and impressive human beings there are. And, as we have seen, human beings are no doubt the only rational beings in (at any rate) this part of the cosmos. It is not hard to guess where a Nobel Prize-winning physicist stands in the cosmic pecking order, as it is perceived by Steven Weinberg. But no human being occupies any such impressive place in the cosmos if it is the work of a Logos.

Whatever may be the accuracy of these speculations about the reasons underlying the emotional attractiveness of the thesis that our existence is without meaning or purpose, it is clear that this is an emotionally attractive position for many people. There are a great many people who by no means *reluctantly* embrace the thesis that our existence is without meaning or purpose. There are a great many people who would be quite taken aback by an invitation to move to a world in which human beings *had* been created to serve the purposes of a Logos.<sup>9</sup> In Thomas Nagel's book *The Last Word*, there is a remarkable passage—remarkable in the self-knowledge it displays—that bears witness to this contention:

[A fear of religion] has large and often pernicious consequences for modern intellectual life.

In speaking of the fear of religion, I don't mean to refer to the entirely reasonable hostility toward certain established religions and religious institutions, in virtue of their objectionable moral doctrines, social policies, and political influence. Nor am I referring to the association of many religious beliefs with superstition and the acceptance of evident empirical falsehoods. I am talking about something much deeper—namely, the fear of religion itself. I speak from experience, being strongly subject to this fear myself: I want atheism to be true and am made uneasy by the fact that some of the most intelligent and well-informed people I

know are religious believers. It isn't just that I don't believe in God and, naturally, hope that I'm right in my belief. It's that I hope there is no God! I don't want there to be a God; I don't want the universe to be like that.<sup>10</sup>

### Suggestions for Further Reading

Burrill's *The Cosmological Arguments* contains (owing to the editor's conviction that the teleological argument is a version of the cosmological argument) a useful set of readings on the teleological argument.

The greatest single work devoted to the teleological argument is Hume's *Dialogues Concerning Natural Religion* (published in 1779, three years after Hume's death). This book is one of the great masterpieces of destructive philosophical criticism. Although the book is largely devoted to the teleological argument (the version of the argument that Hume considers emphasizes the analogical aspect of the argument, the drawing of an analogy between the cosmos and a machine or other human artifact), parts of it are devoted to other topics. The reader interested primarily in the teleological argument is directed to Parts II through VIII, and particularly to Parts II and V. (Many of the points made in Chapter 9 about the weakness of the conclusion of the teleological argument are taken from the *Dialogues*.) For a powerful and sophisticated defense of the argument against Hume's criticisms, see Swinburne's "The Argument from Design." The greatest of the pre-Darwinian expositions of the teleological argument is Paley's *Natural Theology* (1802).

There is a vast literature on Darwinism and "design in nature." My conviction is that the best way to approach a vast literature (if you do not want to devote your life to it) is to look at a selection of works that take opposite and extreme positions. Here are six: Dawkins, *The Blind Watchmaker*; Monod, *Chance and Necessity*; Dennett, *Darwin's Dangerous Idea*; Denton, *Evolution: A Theory in Crisis*; Johnson, *Darwin on Trial*; Behe, *Darwin's Black Box*. The first three are "extreme" in that they not only maintain that Darwin has shown the teleological argument to be invalid, but that he has shown its conclusion to be false, has shown there is no design or purpose in nature. The authors of the latter three books accept the geologists' findings as regards the age of the earth and the paleontologists' description of the fossil record, but contend that Darwin's theory does not account for the observed facts of the history of life. Johnson and Behe maintain that only intelligent design can account for these facts, and Denton implies that some new and as yet unknown scientific theory is needed.

For Nietzsche's best expression of his "joyous" acceptance of atheism (or perhaps it would be better to say, of the obsolescence of the idea of God) see *Thus Spake Zarathustra*.

### Notes

1. One school of twentieth-century philosophers, the existentialists, held that it was contrary to the nature of a rational being—but existentialists would not use the term 'rational being'—that a purpose be imposed on its existence "from outside"; even an all-powerful Creator, the existentialists held, would be unable to confer on a rational being a purpose that existed independently of that being's free choices, for it is by freely choosing between alternatives that one creates one's own purposes. This is an interesting point of view, but we shall be unable to devote

to it the attention it deserves. I am simply going to assume that it is in principle possible for a creator to endow its creations with a purpose, even in the case—admittedly a very special case—in which those creations are rational beings.

2. A sailor notes a strange cloud formation that has suddenly appeared on the horizon. A spectroscopist regards a puzzling group of lines in the spectrogram of a certain star. A detective considers the fact that the murder victim, though otherwise fully dressed, was wearing no underwear. An archaeologist reading an ancient inscription comes upon a hitherto unrecorded hieroglyph. Each says, "I wonder what *that* means?" In each of their utterances, the word 'mean' has a different sense.

3. The importance of this principle in discussions of the implications of the fine-tuning of the cosmos has been pointed out by John Leslie in *Universes* and other publications. Leslie calls the principle the "Merchant's Thumb" principle, on the basis of the following story. A merchant, displaying an expensive silk robe to a potential buyer, consistently keeps a hole in the robe covered with his left thumb. When he is accused of dishonestly concealing the hole from the buyer, his defenders point out that everyone's left thumb has to be *somewhere*. And so it does; but one can think of a very good explanation of its having been over the hole, and it is therefore wrong to proceed on the assumption that its having been over the hole requires no more explanation than its having been at any other particular place. (The statement of the principle in the text is my own; Leslie should not be held responsible for it.)

4. It is no doubt possible to construct theories that allow only one possible cosmos-design. Theories of this sort would be the theoretical counterparts of cosmos-producing machines with no dials on them, machines into each of which a particular cosmos-design was "hard-wired." But if any such theory is possible, an infinite number of them are possible, and only an infinitesimal subset of them will be theories that fit a cosmos like this one. Most of them will fit only cosmos in which life is not possible. This point is taken from a forthcoming book by an elementary-particle physicist, Professor Steven Barr's *Modern Physics and Ancient Faith*.

5. Or so it seemed reasonable to say when the first edition of this book was written. See, however, Lee Smolin, *The Life of the Cosmos* (New York: Oxford University Press, 1997).

6. One possibility would be that, although there are many cosmos, they exist "one at a time": a cosmos has a finite "lifetime" (the heavens will wear out like a garment, as the Psalmist says), and, when it has passed away, another will somehow be born from its remains. And each cosmic "rebirth" has the effect of resetting the dials.

7. That joy is the appropriate reaction to the supposed discovery that human beings inhabit a world without meaning or purpose has been an important theme in the writings of many nineteenth- and twentieth-century thinkers. The works of the nineteenth-century German writer Friedrich Nietzsche are far more worthy of serious attention than anything else written to the same end.

8. *The First Three Minutes: A Modern View of the Origin of the Universe* (London: Andre Deutsch, 1977), p. 155.

9. The first draft of the first edition of this book did not contain the two sentences that precede this note but rather the sentences "Very few atheists *reluctantly* embrace the thesis that our existence is without meaning or purpose. Almost all atheists would be quite taken aback by an invitation to move to a world in which human beings *had* been created to serve the purposes of a Logos." A friend of mine who read the draft took exception to these two statements. He claimed that although he *believed* there was no God and that life was meaningless, he *wished* there were a God and that life had a meaning. Out of deference to his apparently very strong feelings on the subject—but with some misgivings, because I couldn't help believing he was fooling himself—I changed the words of my draft to the milder words to which this note is appended. My misgivings were in the event justified, although not for quite the reason I thought they were: a couple of months later, my friend became a theist.

physics. Or at least this is true when these questions are considered in their entirety. But there is a question that could be thought of as a part of these questions (an answer to it would be a part of the answers to them) that is pretty clearly within the domain of metaphysics. We shall devote this chapter and the following chapter to this question.

The question we shall be addressing is rather hard to state if we want to state it in a way that does not favor one answer to it over other possible answers. We might try this: What *kind* of thing are we human beings? But this formulation is too abstract to convey much. It often happens in philosophy that philosophers pose a question and suggest various answers to it and that the answers are clearer than the question. The present case is one of them. One way to deal with such a difficulty is to let the answers define the question: it is the question to which those statements are possible answers. Let us try that strategy.

The possible answers to the question we are trying to understand (at least the possible answers that are taken at all seriously today) are all forms of either *dualism* or *physicalism*. The first step in trying to understand our question is to understand these terms.

Suppose that by a "physical thing" we mean an individual thing made entirely of those things whose nature physics investigates. If current physics is correct, all the objects of our sensory experience—pieces of chalk, beetles, stars, and everything else we can touch or see—are made entirely of three kinds of elementary particles: up-quarks, down-quarks, and electrons (plus a few kinds of particles, such as photons, whose exchange by quarks and electrons enables the quarks and electrons to interact). It is an interesting technical question what we mean by 'made entirely of', but let us suppose we have an adequate intuitive understanding of this phrase. (Here is an example to aid our intuitions: A sand castle is made entirely of grains of sand—provided the child who built it did not incorporate into its structure a twig or lollipop stick or anything else not made of sand.) Thus, by the terms of our definition, all the objects of our sensory experience are physical things.

If an individual thing is neither itself a physical thing nor has any physical things as parts, we shall call it a "non-physical thing." We should note that this definition does not rule out the possibility of individual things that are neither physical things nor non-physical things. An object that had both physical things and non-physical things as parts would be neither a physical thing nor a non-physical thing. We could call such an object an "amalgam." I shall have nothing to say about amalgams, apart from a brief remark in note 4. When I talk of things that are "not physical," my remarks are meant to apply only to non-physical things and not to amalgams, even though amalgams are, strictly speaking, not physical things. (And my remarks apply only to individual things. Universals are not non-physical things in the sense I am giving the term, despite the fact that universals are not physical things.)

In addition to the concept of a physical thing, it will occasionally be useful to have the concept of a physical *property*: we shall understand a physical property to be a property that can be possessed by a physical thing and *only* by a physical thing.

Since we can see and touch human beings, and since we are human beings, it might be thought to follow from our definition of a physical thing that we are physical things. But let us make some distinctions. Let us say that a *human organism* is that

# 10 THE NATURE OF RATIONAL BEINGS: DUALISM AND PHYSICALISM

Since we know of no rational beings besides ourselves, we shall be able to discuss the problem of the nature of rational beings only in relation to ourselves. We have already said something about the nature of rational beings in one sense of 'nature': we have set out the defining characteristics of rationality. Our question will be this: What is it about human beings that enables them to be rational? Perhaps we can best understand what is meant by this question by drawing an analogy with a question about an everyday physical concept like liquidity. We may know that a "liquid" is a stuff that changes its shape to fit the shape of the container in which it is placed but retains a particular volume throughout all changes of shape. But this does not tell us what it is about water (that is, the chemical compound whose molecules are formed from two hydrogen atoms and an oxygen atom) that accounts for the fact that it is a liquid at temperatures and pressures at which table salt is a solid and carbon dioxide a gas. Explanations of this fact are available. (They appeal principally to the forces that operate between H<sub>2</sub>O molecules and the way in which these forces are determined by the properties of hydrogen and oxygen atoms and their arrangement in the H<sub>2</sub>O molecule.) We want to find an analogous explanation of the way in which rationality is "realized" in human beings (analogous, that is, to the way in which liquidity is realized in water): we want to know what "underlying" features of human beings enable them to have the properties listed in the abstract definition of rationality.

The short answer to the question, What is it about human beings that enables them to be rational? is, No one knows. The rationality that is, as far as we know, unique to human beings is a mystery, as is the conscious experience human beings share with many other animals. The two questions 'How is rationality realized?' and 'How is conscious experience realized?' are generally viewed by philosophers as belonging more to the part of philosophy called "the philosophy of mind" than to meta-

which a biologist would classify as a member of the species *Homo sapiens*. And let us say that a *human person* is that which we refer to when we use the first-person-singular pronoun ('I', 'me', 'moi', 'ego', 'ich', . . .). When I have used the words 'human being' in this and earlier chapters, I have been assuming that human persons and human organisms are one and the same. To call *x* a human being is to call *x* a human person, but with the understanding or implication that *x* is a human organism, a rational animal. (Or this, at least, is what I take 'human being' to mean. Perhaps there are those who would dispute this definition.) But the thesis that human persons and human organisms are one and the same is controversial.

If human persons and human organisms are one and the same, then, since human organisms are obviously physical things, it follows that human persons are physical things. The thesis that human persons are physical things is called *physicalism*. (This word is also used as a name for the stronger thesis that *all* individual things are physical things. And the stronger and weaker senses of the word tend not to be carefully distinguished, owing to the fact that most philosophers who believe that human persons are physical things also believe that all individual things are physical things. I shall use 'physicalism' only for the thesis that human persons are physical things.<sup>1</sup>)

The thesis that human persons are non-physical things is called *dualism*. (More exactly, the thesis that there are both physical and non-physical things and that human persons are among the non-physical things is called dualism. Some idealists perhaps hold that there are only non-physical things, persons among them; such idealists are not dualists.) This word comes from the Latin word for 'two'. The dualist believes that human persons have a "dual" nature. The person is, strictly speaking, a non-physical thing, but it is very intimately associated with a certain physical thing, a human organism, which is called the person's *body*. The body, not the person, is the thing a biologist would classify as a member of the species *Homo sapiens*. The dualist will concede that we frequently make assertions by which we appear to ascribe physical properties to human persons, assertions like, "John weighs 46 kilograms" or "Alice is 165 centimeters tall." But, according to the dualist, it is not strictly true that John weighs 46 kilograms or has any other weight; and it is not strictly true that Alice is 165 centimeters tall or has any other height. John and Alice, rather, possess such properties only vicariously: it is, strictly speaking, not they but their bodies that have weights and heights. This does not mean that there is anything wrong with saying "John weighs 46 kilograms" in ordinary contexts; this statement is to be understood as a kind of shorthand expression of the assertion that John's body weighs 46 kilograms, just as Alice's statement "I'm carrying 1,400 tons of pig iron" is a shorthand expression of the assertion that the ship of which she is the cargo officer is carrying 1,400 tons of pig iron. A "dualistic" analysis of the ordinary statement "John weighs more than he likes" well illustrates what is meant by saying that, according to the dualist, human persons have a "dual nature." Nothing, according to the dualist, could literally weigh more than it liked. Rather, the dualist holds, it is John, the non-physical person, who does the disliking, and it is his body, the physical organism, that has the weight that is the object of the dislike.

What is the "intimate association" that holds between the person and the person's body? Dualists have answered this question in more than one way. The most obvious answer, and the one that commands the widest allegiance among dualists, is contained

in a theory called "dualistic interactionism." In order to set out the content of this theory, let us look at a typical human person and see what dualistic interactionism says about the relations that have to hold between a person and an organism for that organism to be that person's body. Let us consider one Jane Tyler, the author of the well-regarded novel *The Sineus of Thy Heart*, whom we may suppose to be a typical human person. And let us consider the following words and phrases:

- Jane Tyler
- 'the author of *The Sineus of Thy Heart*'
- 'I' (spoken by Jane Tyler)
- 'you' (spoken by someone addressing Jane Tyler)
- 'she' (spoken by someone relating an anecdote about Jane Tyler)
- 'that woman over there' (spoken by someone calling someone's attention to Jane Tyler)
- 'Jane Tyler's mind'
- 'Jane Tyler's soul'

According to the dualist, when these phrases are spoken in the indicated contexts, they denote or name or stand for or refer to the same thing, a non-physical thing, a thing not composed of elementary particles and not observable by the senses, a thing without weight or mass (gravity and inertia are concepts that apply only to physical things), and having no position in space—at least it is hard to see how a non-physical thing could have a position in space, although Saint Thomas Aquinas believed that angels were non-physical things that had positions in space. (The dualist will probably also want to say that this thing has no parts: as metaphysicians say, it is a *simple*. But, in principle, one could be a dualist and hold that a human person had parts, provided they were all non-physical parts.)

In addition to Jane Tyler there is Jane Tyler's body, a physical thing, a living human organism. Our question is: What is it that makes one particular human organism *Jane Tyler's* body and not some other person's body—or no one's body at all? Dualistic interactionism tells us that this particular organism is Jane Tyler's body because of a certain two-way causal connection that holds between Jane—let us get on familiar terms with her—and that organism. A certain organism is Jane's body because she affects it and it affects her. But we must be more specific than this, because cause-and-effect relations can hold between any human person and any human organism.

There is, interactionists maintain, a very special way in which Jane can affect the one particular human organism that is her body: she can cause changes in it without causing changes in any other organism (other than its own parts; multicellular organisms have cells, which are themselves organisms, as parts). And there is a very special way in which one particular organism can affect her: it can cause changes in her without causing changes in any organism besides itself (and its own parts).

Let us look at an example. Suppose Jane begins to whistle. In doing this she causes changes in a certain organism (electrical currents flow along very specific neural pathways in the organism, its lips assume a specific configuration, and many other changes occur in it). And it may be that in beginning to whistle, she causes changes in no organism but this one and some of its constituent cells. Now *I* can also do things that

will cause changes in that organism; I can, for example, open a window on a freezing day and cause it to begin to shiver. But I can do this only by causing changes in another, wholly distinct, organism, *my* body.

Now let us look at an example of the special way in which changes in the organism that is Jane's body can cause changes in Jane the person. Suppose Jane steps on a tack. The resulting puncture wound in her foot will cause *her* to be in pain. (Being in pain would seem clearly to be a property of Jane the person. Being in pain—having the sensation we call "pain"—is a property of an organism only if the organism, or some part of it, *is* a person.) It is true that changes in other organisms than Jane's body can cause changes in Jane. If I step on a tack, the resulting puncture wound in my foot may cause her to feel concern (and feeling concern is a property of the person). But a change in my body can cause a change in Jane only by causing a change in another organism, *her* body, that is not a part of my body.

Dualistic interactionism, then, consists of two theses: dualism, the thesis that there are human persons and human organisms and that no human person is a human organism, and interactionism, the thesis that each human person (at any rate, each living human person) has a body, a unique human organism to which it is bound "directly" by mutual causal interaction. The two most important dualists in the history of metaphysics, Plato and Descartes, were interactionists. Other dualists, however, have rejected interactionism, generally because of the physical or metaphysical difficulties raised by the thesis that a non-physical thing (a thing having no physical properties like mass or electrical charge) could affect a physical thing. Descartes's follower Nicholas Malebranche, for example, held that when a person "wills" or "tries" or "sets out" to whistle, God effects appropriate changes in a certain human organism. Similarly, he held that when a human organism is punctured by a tack, God causes a certain person to experience appropriate sensations of pain. This theory is called "occasionalism," since it holds that changes in the person are never the *causes* of changes in an organism but are only the "occasions" of changes in an organism; in the same way, changes in an organism are never causes of, but only occasions of, changes in a person.

A second dualistic alternative to interactionism is "epiphenomenalism" (from a Greek word meaning 'by-product'). According to this theory, changes in a human person can be caused "directly" by changes in a particular human organism, but changes in the person never cause changes in that organism. Each change in the organism is caused by prior changes in the organism or in its immediate physical environment, and these physical events also sometimes cause changes in the person—but there is no "feedback" from the person to the organism: the non-physical events that are changes in the person never have physical effects. Persons are thus related to their bodies as billows of smoke are to the fires from which they issue: persons exist and are non-physical things, but they are mere by-products of the physical activity going on in certain organisms. (Or this is one way to understand epiphenomenalism. Epiphenomenalists have not generally expressed themselves very clearly. It is possible that at least some epiphenomenalists want to say that the person *is* the organism and that it is people's *sensations and thoughts* that are the by-products of the events going on in the organism. Other epiphenomenalists write in such a way as to suggest that persons are not individual things at all but are mere collections of the thoughts and sensations generated by "their" organisms. I can make nothing of either of these ideas.) It is a

consequence of this theory that our belief that we can influence the motions of our bodies is an illusion. The illusion is itself, according to epiphenomenalism, a by-product of the physical activity of the body.

There are several other dualistic theories of the nature of the person-body relation, but we shall not discuss them. Nor shall we further discuss occasionalism and epiphenomenalism.

We should take note of one other point about dualistic interactionism: it does not obviously follow from dualistic interactionism that the non-physical human person can exist without being in interaction with a human body. Some argument would be required to establish that a dualistic interactionist should believe a human person could exist without a body. Plato believed that the soul—that is, the person—would "automatically" continue to exist when the body it was associated with died. And he did have an argument for this thesis: that the soul is a metaphysical simple, and that a thing can cease to exist only by "coming apart," by being resolved into its elements; a simple, a thing without parts, must therefore be imperishable. This argument, however, is not particularly convincing. For example, the premise that a thing can cease to exist only by coming apart deserves further discussion. One might cite the fact that current physics treats electrons and various other particles as having no parts; yet an electron can be "annihilated" by a collision with a positron. But we shall not pursue this subject. We shall not try to discover whether Plato's argument is ultimately defensible or whether there might be other interesting arguments for the same conclusion.

The physicalist, who holds that the human person just *is* the human organism (or some part of it), does not face the problem of explaining the relation between person and organism.<sup>2</sup> Since, for the physicalist, the person and the organism (or a part of the organism) are identical, a change in the person is a change in the organism. And since the organism is a physical thing, and a physical thing is made entirely of quarks and electrons, it would seem that any change in a human person must be a change in the physical properties of the person: a change in the properties of the quarks and electrons that make the person up, or else a change in the way the quarks and electrons that make the person up are related to one another. Such a change—one that involves only a change in the physical properties of a thing—we may call a purely physical change; examples of purely physical changes would be *receiving a puncture wound in the foot and undergoing a sudden rise in body temperature and having a brain in which electrical currents suddenly begin to flow in such-and-such a way*. If a human person is a physical thing, any change whatever in a human person must be a purely physical change. If, for example, Tim becomes elated because of some news contained in a letter he has just received, this change in Tim, his becoming elated, must be the very same thing (or perhaps we should say the very same event) as some purely physical change.<sup>3</sup>

If it is indeed true that Tim's becoming elated is the very same thing as some purely physical change, then, given what we know about human physiology, it is presumably the same event as some event involving some of the particles that make up Tim's brain—no doubt a change in the way in which electrical currents flow in Tim's brain. Thus, if physicalism is correct about the nature of persons, all those changes in a person we should unreflectively call "mental" or "psychological"—whatever, exactly, these terms may mean—are physical changes in the person (and presumably changes in the

person's cerebral cortex, the part of the brain associated with conscious mental activity). The thesis that mental changes (in human persons at least) just *are* certain physical changes is called the "identity theory." The identity theory is not quite the same thing as physicalism. Physicalism (the theory that human persons are physical things) entails the identity theory (that mental changes in human persons are identical with certain purely physical changes) only on the assumption that mental changes in human persons really exist. And there are philosophers and psychologists who deny the existence of the mental (mental changes and mental states) altogether. We shall not discuss the views of these philosophers and psychologists, who subscribe to theories with names like "behaviorism" and "eliminative physicalism." We shall take the reality of the mental for granted, as do most philosophers and psychologists and, indeed, most physicalists. (Because most physicalists take the reality of the mental for granted, it is safe to say that most physicalists subscribe to the identity theory.)

The two most important theories about the nature of the only rational beings whose existence is uncontroversial (ourselves) are, therefore, dualistic interactionism and physicalism. What can be said for and against each of these theories? Can either be shown to be superior to the other?<sup>24</sup>

We shall begin our attempt to answer these questions by examining some arguments for dualism. (We shall not concern ourselves with defending dualistic interactionism; we shall take it for granted that interactionism is the most plausible form of dualism and shall investigate the question, What can be said in defense of dualism?) Arguments for dualism have this general form: you and I and other human persons are not human organisms or any other physical things because we have properties that could not belong to a physical thing. (It is obviously a valid general principle of reasoning that a thing  $x$  and a thing  $y$  cannot be identical, cannot be one and the same thing, if  $x$  has a property or feature or characteristic that  $y$  lacks.) There are many such arguments. We shall consider five of them. The first argument we shall examine is commonly ascribed to Descartes. (Some commentators find this argument in his *Meditations on First Philosophy*, others in his *Principles of Philosophy*. The passages in both books in which the argument can supposedly be found are, it must be confessed, rather obscure. But the argument is an interesting argument whether or not it is Descartes'. Without pretending to have settled any textual point, I will, simply as a matter of literary convenience, ascribe the argument to Descartes.)

This is Descartes's argument: I can conceive of my body's not existing—indeed, I can conceive of there being no physical world at all—but I cannot conceive of *my* not existing; I am therefore not my body.

When Descartes says I can conceive of my body's not existing, he is not advancing the thesis that I can form a conception of the way things would have been if my body had not existed (no doubt I can, but that I can is not his thesis); he is advancing the stronger thesis that it is possible for me to conceive of the following: things being *just as they seem to me to be* and yet there being no such thing as my body. To conceive of this, I could imagine that there exists some powerful spirit (the "evil genius" we met in Chapter 3) who has decided to deceive me about the existence of a world of physical things: there are no physical things, but the spirit deceitfully "feeds" me a series of sense impressions like the series of sense impressions I should be experiencing if I were perceiving a world of physical things.

And when Descartes says that I cannot conceive of *my* not existing, he is not saying that I cannot form a conception of the way things would have been if I had not existed (that would be false; I can conceive of that); he is saying rather that I cannot conceive of the following: *things being just as they seem to me to be* and yet there being no such thing as myself. In other words, Descartes holds that, however absurd it may seem, the hypothesis that I exist and do not have a body is a hypothesis it is *possible* for me to entertain; but the hypothesis that I do not exist is not simply a hypothesis that seems absurd to me: it is a hypothesis that is literally impossible for me to entertain. It is remotely possible that my conviction that there are physical things, including my own body, is an illusion. It is not even *remotely* possible that it is an illusion of mine that I exist. Not an illusion of mine: if I am "there" to have the illusion, I must exist.

The argument, then, is that my body has the following property:

can be conceived by me not to exist,

as does every other physical thing. But I do not have that property. Therefore, I am not identical with my body—nor am I identical with any other physical thing.

The trouble with this argument is that it proves too much. I can obviously make *some* statements of the form 'I am (identical with) . . .' (where the blank is to be filled by something other than 'I' or 'me' or 'myself') and thereby say something true, but an argument having the same form as Descartes's argument can be used to refute any such statement. Let us look at an example. The statement

I am the author of *An Essay on Free Will*

is true; that is, if I were to speak these words, I should say something true, for there is a book of that title, and I am its sole author. But suppose I were to reason as follows:

I can conceive of there being no such thing as the author of *An Essay on Free Will*. That is, I can conceive of things being just as they seem to me to be and there being no such thing as the author of *An Essay on Free Will*. The easiest way would be for me to suppose that there is no such book: my apparent memories of having written and published such a book are fantasies. But I cannot conceive of there being no such thing as myself. Therefore, the author of *An Essay on Free Will* has the property "can be conceived by me not to exist" and I do not have that property. Therefore, I am not the author of *An Essay on Free Will*.

Since this argument starts from true premises and yet has a false conclusion, it must contain some error of logic. Most philosophers would agree that the error is this: the words 'can be conceived by me not to exist' do not name or express a property, but the argument treats them as if they did. If these words did name or express a property, we ought to be able to take a sentence like 'The author of *An Essay on Free Will* can be conceived by me not to exist' and substitute for 'the author of *An Essay on Free Will* any word or phrase that denotes (designates, refers to, is a name for) the same thing and get a sentence that is true if the original sentence is true.

But this is not what in fact happens. The word 'I' denotes (when I use it) the same thing as 'the author of *An Essay on Free Will*'; but 'The author of *An Essay on Free Will* can be conceived by me not to exist' is true, and 'I can be conceived by me not to exist' is false. Let us compare 'can be conceived by me not to exist' with some phrase that re-

ally does name a property—say, 'was born during the Second World War'. The author of *An Essay on Free Will* was born during the Second World War (take my word for it). The word 'I', when I speak it, and the words 'the author of *An Essay on Free Will*' are two names for the same thing. The appropriate substitution produces the sentence 'I was born during the Second World War'. Is it true that I was born during the Second World War? Well, of course it is. It has to be, given that the author of *An Essay on Free Will* was born during the Second World War and that I am the author of *An Essay on Free Will*.

If a phrase that looks as if it named a property (like 'can be conceived by me not to exist') does not obey this simple substitution rule, then, contrary to appearance, it does not name a property. Therefore, 'can be conceived by me not to exist' does not name a property. And, therefore, Descartes's attempt to prove that persons are not physical things contains an error. There is nothing wrong with the principle of reasoning 'If  $x$  has a property  $y$  lacks, then  $x$  is not identical with  $y$ ', but Descartes misapplied this valid principle as a result of his treating 'can be conceived by me not to exist' as a name of a property.

We now turn to our second argument for dualism, a very popular one:

Physical things are incapable of thought and sensation. But human persons are capable of thought and sensation. Therefore, human persons are not physical things.

But why should we believe that physical things are incapable of thought and sensation? I am willing to grant that if we try seriously and in detail to imagine a physical thing having thoughts and sensations, we can find this notion—the notion of a physical thing having thoughts and sensations—very puzzling. There is a famous passage in Leibniz's *Monadology* that very clearly brings out the puzzling aspects of this notion:

Furthermore, we must admit that *perception*, and whatever depends on it, *cannot be explained on mechanical principles*, i.e. by shapes and movements. If we pretend that there is a machine whose structure makes it think, sense and have perception, then we can conceive it enlarged, but keeping to the same proportions, so that we might go inside it as into a mill. Suppose that we do: then if we inspect the interior we shall find there nothing but parts which push one another, and never anything which could explain a perception. Thus, perception must be sought in simple substance, not in what is composite or in machines.<sup>5</sup>

To take a more modern example, suppose someone were to claim to have programmed a computer so that it could think (in a sense that implies conscious experience and self-awareness) or to have constructed a thinking robot. If the computer or robot were enlarged so that people could walk about inside it, a party of tourists being led through the vast machine would see nothing but physical things interacting physically. And this would be no illusion. It's not as if the thought and conscious experience were hidden away in some part of the machine off limits to visitors.

But then where *are* the thoughts and the experience? Where *could* they be? How could the mere physical interaction of bits of metal and plastic and silicon "add up to" thoughts and experience? It is important to realize that this point has nothing to do with the specific kinds of physical material a computer or robot would be likely to be

made of. The point has to do only with the fact that the materials are *physical*. The point would be unchanged if we imagined a party of tourists being conducted through *our selves* (or our bodies), as in Isaac Asimov's interesting science-fiction novel *Fantastic Voyage* (or the unspeakably silly movie of the same title). If we could be greatly reduced in size and go inside a functioning human brain and have a look round, we should see no thoughts or experience, not even if we saw everything there was to see. If God looks inside a human brain, even He sees nothing but unthinking physical things like neurons and Nissl granules and amino-acid molecules and electrons in continuous mutual physical interaction. Where, then, are the thoughts? Where are the sudden feelings of elation or despair? Where are the sensations of heat and pain and pressure and color? The answer is, obviously, that they are elsewhere. And that "elsewhere" must be a place that is receptive to the presence of such things, a place where they *could* exist. They must exist in a non-physical thing. (If we like, we can say that they must exist in a non-physical thing that is *mental*: a mind or a soul. But, unless we can say something useful about what we mean by 'mental thing' or 'mind' or 'soul', to say this would be to say no more than that they must exist in a non-physical thing.)

Various physicalists—who must of course believe that physical things are capable of thought and sensation—will reply to this argument in various ways. What follows is my own reply. Some physicalists would reject some parts of it.

Let us begin with the question, *Where* are the thoughts and sensations? The answer is that since these things are changes in the cerebral cortex, they are all around you (you who have in imagination been reduced in size and are physically inside someone's brain). It does not follow from this that you *see* them, since they may involve the whole cerebral cortex or the whole brain or widely scattered parts of the brain: it may be that you cannot see them for the same reason you cannot see the event called 'the election' on election day. But let us suppose for the sake of argument that these events are sufficiently localized that you can see them. (Or some aspects of them: a human being cannot see every aspect of any event. You can see the street lamps come on in your neighborhood, but you cannot see the flow of electrons that is an indispensable component of this event.) Of course, these events do not *look* to you like mental events, but then what would you expect a mental event to look like? ("Well, something like the way mental changes in *myself* look to *me*, as when I experience a sharp pain in my left shoulder or a thrill of fear or an intellectual insight." But that's what it's like to experience *having* or *being the subject of* a mental change. That's what a mental change *in you* "looks like" to you. What would you expect mental changes in someone *else* to look like to you?) And, anyway, a change may be of a certain type without its being evident that it is of that type. Suppose a computer has been programmed to compute the orbit of a certain satellite. Suppose the computer were greatly enlarged and that you went inside it, "as into a mill." You would not see any orbital computations going on—or at least you would not see anything that "looked like" orbital computations. (What would you expect orbital computations to look like?) The Leibnizian thought-experiment, therefore, should cause the physicalist no unease. Things inside the brain look just the way they would look if physicalism were correct.

Many physicalists would think that this was a sufficient reply to the charge that the notion of a physical thing that thinks is mysterious. I cannot agree with them. I do



not deny that everything said in the preceding paragraph is correct, as far as it goes. Nevertheless, it seems to me that the notion of a physical thing that thinks is a mysterious notion and that Leibniz's thought-experiment brings out this mystery very effectively. We must remember, however, that our present question is not whether the physicalist is faced with a mystery; our question is whether dualism is to be preferred to physicalism. If thinking is a mystery for the physicalist, this fact will be relevant to our question only if it can be shown that the dualist is not confronted with the same mystery or with some corresponding mystery.

And, I believe, the dualist is. For it is thinking itself that is the source of the mystery of a thinking physical thing. The notion of a non-physical thing that thinks is, I would argue, equally mysterious. How any sort of thing could think is a mystery. It is just that it is a bit easier to see that thinking is a mystery when we suppose that the thing that does the thinking is physical, for we can form mental images of the operations of a physical thing and we can see that the physical interactions represented in these images—the only interactions that *can* be represented in these images—have no connection with thought or sensation, or none we are able to imagine, conceive, or articulate. The only reason we do not readily find the notion of a non-physical thing that thinks equally mysterious is that we have no clear procedure for forming mental images of non-physical things. Still, we are not wholly without resources for constructing mental images of non-physical things. (No doubt most of us associate some sort of mental image with the doctrine of dualistic interactionism: perhaps a human body with a vague "something" inside or above its head.) Let us see what we can do.

Leibniz, in the passage we have quoted, contends that a thinking thing must be a simple, a thing without parts. Well, let us represent, in our thought, a simple non-physical thing by a dot, and a composite non-physical thing by a bunch of dots, perhaps a bunch that is in constant internal motion like a swarm of bees. Might a *composite* non-physical thing "think, sense, and have perception"? It is hard to see how. Consider our proposed mental picture of a composite non-physical thing. If the simples that make up a composite non-physical thing do not think individually, where is the thinking in our picture? How can a bunch of things that do not individually think or sense or have perception add up to something that does think or sense or have perception? How could their causal interaction produce such properties? Note that these questions are exactly parallel to the questions Leibniz's thought-experiment raises about thought and composite *physical* things. The only real difference between the two cases is that a mental image of a composite physical thing will have reasonably "sharp" constituents drawn from our experience of actual physical things—images of gears and wheels, say—, whereas (an attempt at) a mental image of a composite non-physical thing will be vague and arbitrary (arbitrary because non-physical things necessarily lack visual characteristics; we chose dots because dots come as close to having no characteristics as anything we can picture).

Leibniz would no doubt agree that these reflections show that a composite non-physical thing cannot think. After all, his position is that a thinking thing has to be a simple.<sup>6</sup> But let us look at our proposed mental picture of a (non-physical) simple. It is just a dot. How can we cause it to change in our imagination in such a way that this change will represent its having a series of thoughts and sensations? Change of position (relative to other imagined dots) will be of no help, because that is a relational

change, and thought and sensation are supposed to be intrinsic features of thinkings, sensing things. Even a dot must have a shape, but when we use dots to represent non-physical simples we do our best not to attend to their shapes, for insofar as we think of a dot as having a shape, we think of it as being composed of smaller regions, and thus as composite.

We might think of the dot as changing color, I suppose. Let's try that. Imagine a dot continuously changing its color in some very complex way. Are you imagining something thinking or having sensation? Where are the thought and the sensation in the picture your imagination has created? My point in asking these unanswerable rhetorical questions is not to suggest that a non-physical simple cannot think. (Although I believe that human persons are physical things made of smaller physical things, I believe that God is a non-physical simple, so I should hardly want to suggest that a non-physical simple cannot think.) My point is that nothing could possibly count as a mental image of a thinking thing. Or, at least, nothing could count as a mental image that *shows* or *displays* a thing as thinking (except by convention, as, for example, "thought-balloons" in comic strips do; or via the familiar outward and visible signs of human thought, like those displayed by Rodin's *The Thinker*). And, I am suggesting, we need to keep this fact in mind when we consider Leibniz's thought-experiment. It is only the difficulty of conducting a similar thought-experiment for non-physical things that keeps us from seeing that his thought-experiment does not favor dualism over physicalism. Consider this analogy. We are amazed to see a human figure hurtling through the sky like Superman. "It's a woman!", someone shouts. "Why a woman?", we ask. "Well, it's either a man or a woman, and it's impossible for a man to fly." This argument is valid, and there are certainly good reasons for thinking that it's impossible for a man to fly. But there are equally good reasons (the same ones) for thinking that it's impossible for a woman to fly. Therefore, the argument gives us no reason to prefer the hypothesis that the human figure we saw in the sky was a woman to the hypothesis that it was a man. And this is exactly parallel to what one should say in response to Leibniz's thought-experiment: Since we are unable to imagine a non-physical thing in a way that displays it as thinking, the fact that we are unable to imagine a physical thing in a way that displays it as thinking does not give us a reason to prefer the hypothesis that we human thinkers are non-physical things to the hypothesis that we are physical things.

These points about mental images can be generalized so as to apply to any type of representation. Mental images are representations of how things are or might be, but there are representations of many other kinds, such as schematic diagrams on paper, three-dimensional cardboard models, computer models, and scientific theories. In general, to attempt to explain how an underlying reality generates some phenomenon is to construct a representation of the working of that underlying reality, a representation that in some sense "shows how" the underlying reality generates the phenomenon. (The best scientists seem to be able to "translate" their verbally and mathematically formulated representations of the workings of things into images, which they are able to manipulate mentally in fruitful ways.) Essentially the same considerations as those that show that we are unable to form a mental image that displays the generation of thought and sensation by the workings of some underlying reality (whether the underlying reality involves one thing or many, and whether the things it involves

are physical or non-physical) show that we are unable to form *any* sort of representation that displays the generation of thought and sensation by the workings of an underlying reality. Thought and sensation are therefore a mystery—although not necessarily an insoluble one. But since the mystery, soluble or insoluble, is entirely independent of whether the elements in the representation are supposed to represent physical or non-physical things, the mystery of thought and sensation does not favor dualism over physicalism.

Has the dualist any way to respond to this counter-argument? The answer to this question depends, I believe, on what the dualist can tell us about the positive nature of the non-physical thinking things whose existence dualism asserts. If the dualist can say no more about them than that they are non-physical things, dualism gains no advantage over physicalism and perhaps gains the disadvantages of postulating the existence of things of a kind physicalism does not postulate, and of having to account for the interaction between these things and physical things. Let us (once more) consider an analogy. Suppose Sir Aaron Oldham, the well-known imaginary seventeenth-century scientist, set out to explain the observed phenomenon of magnetism. Sir Aaron believed that all physical interaction was transmitted by contact between physical things, by "pushes and bumps," and he was therefore unable to believe that magnetism was a wholly physical phenomenon, since it could act across empty space and could act "through" a physical object like a sheet of glass or paper without affecting the intermediate object in any way. He therefore postulated that associated with each lump of lodestone (the only magnets he knew about) there was a non-physical thing that had the power to cause nearby iron objects to move toward the lodestone. "Should a Lodestone be enlarged," he wrote, "to such a degree that a Man were enabled to pass among the corpuscles composing it, as an Earthworm might pass among the particles of Soil comprised in my Garden, he would observe naught but corpuscles, whether at rest or in motion, a certain quantity of Motion being on frequent occasion translated from one to another of the same corpuscles by Collision. He would see therein no Action by which the motion of a distant Pin or Nail toward those corpuscles might be effected."

We may imagine—let us shift to the historical present—that one of Sir Aaron's scientific rivals puts forward an alternative theory of magnetism: that there are unknown physical interactions, interactions other than pushes and bumps, that cause pins and nails to move toward lumps of lodestone. It would seem that unless Sir Aaron can say something about the positive nature of the non-physical entities he has postulated—unless he can say something more about them than that they are non-physical—his theory enjoys no advantage over that of his rival. (Unless Sir Aaron and his rival tell us more than they have so far, this is how things stand: each theory ascribes an observed phenomenon to an unknown cause and tells us nothing about that cause that explains how it produces the phenomenon.) And it might be argued that Sir Aaron's theory is burdened by a disadvantage his rival's is free of: it postulates the existence of non-physical things in addition to physical things, and it faces the problem of explaining how the non-physical can interact with the physical.

Can the dualist tell us anything positive about the nature of human persons? Can the dualist say anything more about human persons than that they are *not* physical things? Many dualists think they can. In this they follow Descartes, who held that the

essence of a human person was thinking. This would appear to mean that the *only* intrinsic properties a human person has or could have are "mental" properties—that is, properties that relate to thought and sensation (and that the human person is essentially such: no human person could possibly have any intrinsic properties but mental properties). Thus, if Descartes is right, human persons have such properties as *being in pain* and *feeling depressed* and *wondering how to spend Saturday afternoon*; human persons do not and could not have such properties as *being 165 centimeters tall* or *weighing 46 kilograms* or any other intrinsic non-mental property.<sup>7</sup>

A typical physicalist believes that human persons have both mental and non-mental properties. A dualist might believe this also, although the dualist, unlike the physicalist, would have to say that the non-mental properties of the human person were not physical properties, either—that they were, perhaps, the members of some utterly unknowable class of properties. A dualist of this sort might even hold that our mental properties were related to these "other" properties in the way in which the typical physicalist holds that our mental properties are related to our physical properties: as the typical physicalist thinks that physical properties underlie and determine our mental properties, so the dualist might hold that the "other" properties underlie and determine our mental properties. A dualist *could* hold this, but few if any dualists do, and Descartes certainly does not. Descartes's position is that we are mental "all the way through."

Dualists therefore have available to them an account of the positive nature of the non-physical human person: the human person is a mental thing—loosely speaking, a thing having only mental properties. (At least the dualists have such an account available to them if they can solve the rather difficult technical problems raised in note 7. In the sequel, I shall assume that they have somehow solved these problems.) And most if not all dualists accept this account of the positive nature of human persons. They have, therefore, an answer to the charge that they have accounted for the phenomenon of thought and sensation simply by postulating a cause for this phenomenon whose positive nature is entirely unknown.

Does their ability to offer this positive account of the nature of human persons provide a reason for preferring dualism to physicalism? It is, I think, plausible to argue that in offering this positive account they have done essentially what Sir Aaron Oldham would have done if he had attempted to give an account of the positive nature of the non-physical things associated with lumps of lodestone by saying that these things had "magnetic" properties and no others. That would not really be an "account" at all, because the words 'magnetic property' could mean nothing but 'power to produce the observed phenomenon of magnetism'. We should have no "hold" on what a magnetic property was except through its observed effects, the very things we want to explain. The dualist who holds that we are things that have only mental properties is simply asserting the existence of things that manifest the phenomenon to be explained (thought and feeling) and which have no properties besides that of manifesting the phenomenon. It is important to stress that this argument does not have the least tendency to show that dualism is wrong. For all we have said so far (note 7 aside), there might well be things that had only mental properties. The argument is not designed to show that dualism is wrong, but only that dualism enjoys no advantage over physicalism as regards the mystery of thought and sensation.

The dualist who asserts that thoughts and sensations occur as changes in a thing all of whose properties are mental has done no more to address the mystery of thought and sensation than has the physicalist who asserts that thoughts and sensations occur as changes in a physical thing. It is true that no one has any account of how thoughts and sensations could be features of physical organisms. In fact, no one can say what an account of this would look like, even in broadest outline. But then no one has any account of how there could be a thing that had only mental properties, and no one can say what an account of this would look like, even in broadest outline.

We now turn to a third argument for the conclusion that one is not the same thing as one's body. (That is, for the conclusion that one is not the same thing as the human organism one can bring about changes in without bringing about changes in any other multicellular organism.) This argument proceeds from the observation that we do not seem to ourselves to occupy the same regions of space as our respective bodies. The twentieth-century English philosopher G. E. Moore formulated this observation in a strikingly simple phrase: "I am closer to my hands than I am to my feet." (Think about it. Look at your hands and your feet at the same time. Your feet are farther away, aren't they?) But my body is obviously not closer to my hands than to my feet—to say it was would be like saying Europe was closer to Belgium than to Italy.

The first thing to note about this argument is that, unlike the two arguments we have so far examined, it does not even claim to prove (in my case) that I am not a physical thing. It claims to prove only that I am not a *certain* physical thing: my body. Even if the conclusion of the argument is true, I might be my brain or my left cerebral hemisphere or my cerebral cortex, for those things are all closer to my hands than to my feet. And, of course, the argument has the same limitation when it is applied to you or to any other human person. One might even maintain that it is inconsistent with dualism to suppose that I am closer to my hands than to my feet. I can be closer to my hands than to my feet only if I have a position in space, and as we have remarked it is hard to see how a non-physical thing could have a position in space.

The argument is, however, unconvincing even as an argument for the conclusion that one is not one's body. There may be a sense in which it seems to me that I am closer to my hands than to my feet, but this appearance might be mere appearance and not reality. Our sense organs—leaving aside the skin, our organ of touch—cluster around the brain. Is it not plausible to suppose that one might seem to oneself to be located at or near the place where one's sense-organs cluster? We seem to ourselves to be at the center of the environment our senses reveal to us, and if our sense-organs cluster around some small region, that region will seem to be at the center of our "subjective world." In fact, it is plausible to suppose that sighted persons would seem to themselves to be approximately where their eyes were, even if their ears and other sense-organs were moved to their elbows and ankles, for sighted people construct their internal model of their immediate environment mainly on the basis of visual data. (Consider Helen Keller, who was blind and deaf from very shortly after her birth. Her model of her immediate surroundings was based almost entirely on tactile data, the data of touch. Would she have felt it natural to say she was closer to her hands than to her feet? Well, perhaps she would have, given the central role her hands played in her knowledge of her immediate environment. But perhaps she would also have felt it natural to say she was closer to her arms than to her head. One can imag-

ine her touching her arms and saying, "My arms are right here . . .," and then reaching up to touch her head and saying, ". . . but my head is way up here.")

Our fourth argument for the conclusion that we are not physical things proceeds from the premise that whether or not there are other rational beings in the cosmos, there certainly could be: there is nothing intrinsically impossible in the notion. And there is nothing intrinsically impossible in the notion that such beings might be physically very different from us. Therefore, it is intrinsically possible for there to be beings that have thoughts and feelings very much like ours, even though they are *radically* different from us in their anatomy and physiology. Imagine a science-fiction story in which there are beings, the Scorpions, with whom we can carry on intelligent conversations about politics and philosophy and even art and who—it never even occurs to us to doubt this—experience pain when they are injured and pleasure when they relax at the end of a hard day in their sulfuric-acid baths. But there is nothing inside their chitinous shells resembling a human brain: there is only purple goo bearing no resemblance whatever, even on the chemical level, to any human tissue. Now suppose physicalism is correct. If that is so, and if we really do think and feel, then our thoughts and feelings are identical with certain physical processes that go on within our brains. But, obviously, none of the physical processes that go on in the grey matter inside our heads goes on in the purple Scorpionian goo.

Suppose, for example, that when one feels pain this event is identical with the firing of C-fibers in one's brain; pain (according to physicalism) has turned out to be the firing of C-fibers, just as bolts of lightning turned out to be massive electrical discharges and water turned out to be H<sub>2</sub>O. But there are no C-fibers, or anything remotely resembling them, inside the Scorpions. And, therefore, if pain is the firing of C-fibers, the Scorpions do not experience pain—just as, if there is no H<sub>2</sub>O on their planet, there is no water on their planet. It would therefore seem that if physicalism is true, neither the Scorpions nor any other beings radically unlike us in their physical structure can think and feel. Only a being that was either human or very similar to a human being could think and feel. But this conclusion can only be regarded as human (or mammalian or carbon) chauvinism. In any case, it is absurd.

A physicalist might well respond to this argument with a question: What makes you so sure it is possible for there to be creatures radically different from us in their physical structure and capable of thought and sensation? And it might not be easy to answer this question unless bluster about chauvinism counts as an answer. But there are two replies available to the physicalist that are consistent with the assumption that the possibility of beings like the Scorpions is a real one.

Each of these replies depends upon a distinction between *types* of events and *tokens* (that is, particular instances) of those types. This distinction is best introduced by example. *War* is a type of event (or an event-type, as philosophers sometimes say), and the First World War and the Seven Years' War and the War of the Austrian Succession are three "tokens" of this one type; Lincoln's death and Caesar's death and the death of Catherine the Great are three tokens of the event-type *death*. A particular, concrete event may be—in fact, all particular, concrete events *must* be—a token of more than one type. Thus, Lincoln's death and Caesar's death are tokens not only of the type *death* but also of the type *assassination*. But, fortunately, not all tokens of the former are tokens of the latter: not all deaths are deaths by assassination. If every

event is a token of various types, then every mental event is a token of various types, and every physical event is a token of various types.

Making use of the type-token distinction, we may distinguish two forms of physicalism (or two forms of the identity theory): type-type physicalism and token-token physicalism. Let us first examine type-type physicalism. Consider the physical event-type *a firing of C-fibers* and the mental event-type *feeling pain*. Suppose someone says that these event-types are identical, are one and the same event-type. This person's thesis could also be put this way, if we neglect some niceties about language some philosophers will not want to neglect: the phrase 'a firing of C-fibers' and the phrase 'feeling pain' are two different names for the same event-type, just as 'water' and 'the liquid that consists of H<sub>2</sub>O molecules' are two names for the same liquid—or just as 'the Morning Star' and 'the planet Venus' are two names for the same celestial object. Type-type physicalism is a generalization of this thesis; according to type-type physicalism, every mental event-type is identical with some physical event-type. (But, of course, only an idealist would suppose that the converse holds. Idealists aside, no one would suppose that, for example, the physical event-type *volcanic eruption* was identical with some mental event-type.)

Type-type physicalism is a very strong thesis, so strong that most physicalists decline to accept it; it is either known to be false (some physicalists will say) or at least it goes far beyond the available evidence. How (the enemies of type-type physicalism ask) can we be sure even that when identical twins experience pains that feel exactly the same there are physical events in the brains of each that are exactly alike—or even very much alike? How can we be sure there is any such pair of physical events to be found? Shouldn't it be left to the neurophysiologists to determine whether two such events exist? Should this question be settled by metaphysicians, by philosophers who have never made any neurophysiological investigations whatever? Fortunately (most physicalists believe) there is a weaker form of physicalism available, a form of physicalism adherence to which does not require philosophers to become armchair neurophysiologists: *token-token physicalism*.

According to token-token physicalism, each concrete mental event (such as my suddenly experiencing a sharp pain in my left arm at noon yesterday or Tim's gradual realization that Alice has been lying to him) is identical with a concrete physical event: a particular change in the physical state of someone's brain (at least in the case of human beings). But it may well be, the token-token physicalist holds, that no mental event-type is identical with any physical event-type. Perhaps, the token-token physicalist says, when Tim gradually realizes that Alice has been lying to him and his identical twin Tom gradually realizes that Alice has been lying to him, each of these two events is identical with a physical change in the respective brains of Tim and Tom, but these two physical changes bear little resemblance to each other (for example, it may be that they take place in different regions in the cerebral cortex). Token-token physicalism does not go so far as positively to deny that there are mental event-types that are identical with physical event-types; it simply refrains from asserting that such identities exist. If there are such identities, the token-token physicalist tells us, it is the business of observational sciences like psychology and neurophysiology to establish them; they are no more to be embraced on purely metaphysical grounds than are the chemical and astronomical identities mentioned above.

If token-token physicalism is correct, there is no problem in principle in saying, for example, that a Scorpion experiences a sensation very like the pain Jane experiences when she has a migraine. Jane's sensation of pain is, or let us suppose it is, identical with a certain pattern of C-fiber firings in her brain; the Scorpion's sensation is identical with some physical process that takes place in a reservoir of purple goo in the Scorpion's metathorax, a process that in none of its physical characteristics resembles the firing of C-fibers in a human brain.

This is the picture provided by token-token physicalism. There are many analogies that token-token physicalists have employed to make this picture a plausible one. The following analogy is typical of these. Suppose three radios are simultaneously receiving the same broadcast. One is an antique crystal set, one a vacuum-tube (valve) radio from the 1950s, and the third is the latest thing in solid-state technology. We may list three "reception events": radio A's receiving the CBS broadcast of the State of the Union Message, radio B's receiving this same broadcast, and, finally, radio C's receiving it. Each of these reception events is identical with a physical process going on inside one of the three radios, but the three physical processes are very different from one another. The thesis of "reception physicalism" may be defined as the thesis that reception events are physical events that go on inside radios. The thesis of type-type reception physicalism is the thesis that each reception event-type (like *receiving the CBS broadcast of the State of the Union Message*) is identical with some physical event-type. The thesis of token-token reception physicalism is the thesis that each reception event-token, or concrete event (like *radio B's receiving the CBS broadcast of the State of the Union Message yesterday*), is identical with some concrete physical event. No doubt everyone will accept token-token reception physicalism. But the fact that the physical events that go on inside a vacuum tube are quite different from the physical events that go on inside whatever the latest solid-state devices are called renders type-type reception physicalism at best doubtful.

Doubtful, perhaps, but not wholly indefensible. I said above that there were two replies available to the physicalist consistent with the assumption that the possibility of thinking, feeling beings like the Scorpions is a real one. The first was to distinguish type-type and token-token physicalism, and to maintain that, whatever the problems faced by type-type physicalism, token-token physicalism is consistent with this possibility. The second reply is an argument for the conclusion that even type-type physicalism is consistent with the possibility of thinking, feeling beings radically different from us in anatomy and physiology—or at least that this may be so, that it is true for all we know.

We may note that event-types may be more or less abstract. The more abstract an event-type is, the weaker the conditions are that an event has to satisfy to be a token of that type, and the less abstract an event-type is, the stronger the conditions are that an event has to satisfy to be a token of that type. Here are five event-types arranged in order of decreasing abstraction: *death, killing* (an untimely death caused by an external agency), *murder* (a deliberate and wrongful killing of one human being by another), *assassination* (the murder of a public figure from a political motive), and *terrorist assassination* (an assassination undertaken to create a politically useful climate of fear within some group). A defender of type-type physicalism could argue that the most that the example of the Scorpions shows is that if each mental event-type is identical