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.[1])

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 Sinews 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 Sinews 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 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.[2] 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.[3]

Our fifth and final argument for the superiority of dualism to physicalism is that dualism can account for the so-called identity of the human person across time and physicalism cannot. It would seem that we normally suppose that the same person can exist at two different times. You, for example, no doubt believe you existed ten years ago—not to mention last Tuesday. You exhibit this belief every time you say something like, "Ten years ago, I'd never have believed I'd be doing this today," or "Last Tuesday I finally decided it was time to buy a new car." And, of course, we rarely if ever believe that the present moment is the final moment of our existence. We therefore normally believe that we are going to exist at various times in the future, for the statement that one is *not* going to exist at any future time is equivalent to the statement that the present moment is the final moment of one's existence.

Some opponents of physicalism argue that physicalism must be false because it contradicts these facts (at least we all suppose them to be facts) about our identity across time. It may be, they argue, that although a "static" physical thing like a diamond or a fly in amber could exist on two dates that were, say, ten years apart, this could hardly be possible for anything made of living tissue. The Koh-i-Noor Diamond is, perhaps, composed of exactly the same matter (exactly the same carbon atoms) that composed it ten years ago, but I am not. If I am, as the physicalists say, a living organism or a part of one, then I have "lost" almost all the atoms that composed me ten years ago and I am now made almost entirely of atoms that existed ten years ago but were then parts of other things or parts of nothing at all. It is true that I have the same brain-cells I had ten years ago (minus those that

have died in the interval), but each of those brain-cells is now made of atoms that were not parts of it ten years ago.

If, therefore, I am a physical thing, the matter that now composes me is not the matter that composed the physical thing that bore the name 'Peter van Inwagen' ten years ago. The physicalist is forced to say that any statement implying that I existed ten years ago must be, strictly speaking, false. Of course—the physicalists could say this much—such statements may be useful statements even if they are false. After all, there are useful statements that imply that the apparent motion of the sun across the sky is real, as when we say, "It was cooler in the garden after the sun had moved behind the elms." We know that the sun did not really move behind the elms, but we talk as if it did because it is usually too complicated to describe the actual state of affairs that accounts for the apparent motion of the sun. And (the physicalist must say) when Alice says to Jack, "It's hard to believe that ten years have passed since you and I last saw each other," this statement must be understood in a similar way: it is useful—because a metaphysically accurate description of the actual state of affairs would be too complicated—but, strictly speaking, false.

And why is the physicalist committed to this conclusion? Let us consider the famous story (famous among metaphysicians, anyway) of the Ship of Theseus. The hero Theseus has a ship, which is entirely composed of wooden planks. Very gradually, over the course of years, the planks are removed from the ship and replaced. The replacement is so gradual that Theseus and his crew are able to be almost continuously at sea, engaged the while in a long series of adventures with a nautical setting. The planks that have been removed from the ship are not destroyed but are rather stacked in a certain field. When all the original planks have been replaced, Stilpo the shipwright notices that the field contains all the components needed to build a ship. Stilpo puts the planks together and puts them together in such a way that they are arranged exactly as they were when they composed Theseus' ship on the day he first took command of her. Stilpo takes his new ship to sea for a shakedown cruise, and his ship and Theseus' ship pass each other at sea.

Call "Theseus' ship on the day he first took command of her" the *Original Ship*. Call the ship in which Stilpo is now sailing the *Reconstructed Ship*. Call the ship in which Theseus is now sailing the *Continuous Ship* (because on any given day after Theseus took command of the Original Ship, the ship he was sailing on that day was made of the same or *almost the same* planks as the ship he was sailing the day before).[1] Is it the Reconstructed Ship or the Continuous Ship that is the Original Ship? The Continuous Ship has the name *Ariadne* painted on her bow, and Theseus swears he has been sailing one ship, the *Ariadne*, these many years.

And the Athenian Office of Marine Registry agrees with him. But isn't it evident that Theseus and the Registry Office bureaucrats are wrong? (Wrong strictly speaking, that is. The statement that Theseus has been sailing one ship, the *Ariadne*, for many years may be a very useful statement for legal and other practical purposes. The statement might be considered a "legal fiction," like the statement that a corporation is a person.) What is a ship but a certain "hunk of matter," a certain assemblage of planks or of atoms or of elementary particles? And the hunk of matter, the assemblage of planks, atoms, and elementary particles that Theseus is standing on now is not the same hunk of matter, the same assemblage of planks, atoms, and elementary particles, that he stood on on the day he took command. But the hunk of matter, the assemblage of planks, atoms, and elementary particles, that *Stilpo* is standing on now is the same hunk of matter, the same assemblage of planks, atoms, and elementary particles, that Theseus stood on on the day he took command. It is, therefore, the Reconstructed Ship and not the Continuous ship that is the Original Ship.

Now let us make one change in the story we have told. Suppose Stilpo had never assembled the planks he found in that field into a ship. Suppose, in fact, that each of the planks that had been a part of the Original Ship was burned to ashes the moment it was removed. Then, of course, it would not be true that the Reconstructed Ship was the Original Ship, since the Reconstructed Ship would not be there at all. But it would still be true that the Continuous Ship was *not* the Original Ship, since it would not be the same hunk of matter as the Original Ship, and a ship is nothing but a hunk of matter. It would also seem to follow that there is no such ship as the *Ariadne*: that is, no one ship that has been under Theseus' feet every day since the day he first took command of the Original Ship. For every time a plank was replaced, a different "shipshape" (so to speak) hunk of matter was under Theseus' feet—*almost* the same shipshape hunk of matter as before the replacement, to be sure, but when it comes to identity, a miss is as good as a mile. A miss is as good as a mile because identity is, as logicians say, *transitive*. This means that if A is identical with B (if A and B are one and the same thing), and B is identical with C, then it follows that A is identical with C. Therefore, if the "before" and "after" ships for each individual replacement of a plank are one and the same ship, it follows that the Original Ship and the Continuous Ship are one and the same ship. And, therefore, if the Original Ship and the Continuous Ship are *not* one and the same ship, the replacement of a single plank must have produced a new ship in at least one case; and if in one case, then presumably in every case.

Let us now return to the question of physicalism and the identity of the person across time. If physicalism is true, each of us is a hunk of matter, a certain assemblage of atoms or elementary particles. But you are not, strictly speaking, the same

hunk of matter as the one that bore your name ten years ago: the atoms that composed *that* hunk of matter are now pretty well scattered throughout the terrestrial biosphere. It follows that physicalism is incompatible with the reality of personal identity across time; physicalism is therefore absurd. And it cannot be argued that dualism is in the same boat—or ship. It cannot be argued that the identity of persons across time is a mystery that confronts both the physicalist and the dualist in the same way and to the same degree. This reply might conceivably be effective against a dualist who held that each human person was a composite of a large number of "smaller" non-physical things. But, while this position is, in theory, compatible with dualism, no actual dualist holds it. According to all actual dualists, a human person is not only a non-physical thing but a metaphysical simple, a thing without parts. The problem facing the physicalist is that if a human person can really exist at two times that are, say, ten years apart, that person must be made of entirely different parts at those two times. The physicalist must be able to explain how a thing can persist (can continue to exist and retain its identity) through a complete change of parts. And this problem is not one that confronts Plato or Descartes or any other actual dualist: if a thing has no parts, then, obviously, there is no problem of how it can persist through a change of parts. Therefore, the dualist concludes, dualism is to be preferred to physicalism.[2]

What can the physicalists say in response to this? In the above argument, it was suggested that the only course open to them is to accept the incompatibility of physicalism and the reality of personal identity across time and to try to live with it—to treat personal identity across time as some sort of useful fiction. And this is a course some physicalists have taken. (One influential philosopher has gone so far as to agree with the Buddhists that the idea of personal identity across time is not so much a useful fiction as a pernicious fiction and that realizing that this idea is a fiction is a kind of liberation.) I myself am unable to take this proposal seriously. But that is only a fact about my psychology and it proves nothing. Nevertheless, the idea of personal identity across time—the idea that it is a fact that one and the same human person can strictly and literally exist at different times and that, moreover, this fact is a feature of reality and not of mere appearance—is so central to a vast array of ways of thinking that have served us and our ancestors for millennia that we should abandon it only in the face of an unanswerable argument. And it is clear that we have no such unanswerable argument, for, even if physicalism is inconsistent with the reality of the identity of the human person across time, one could always be a dualist. And if that were the price we had to pay for a belief in personal identity, the belief would be cheap at the price. That the price is reasonable is a premise of the above argument against physicalism, and that premise I accept. But does this reasonable (but very high) price have to be paid? That is,

is there any way—or is there more than one way—for a physicalist consistently to believe in the identity of the human person across time?

Suppose one wants to be a physicalist, to believe that there is one all-comprehending identity-relation (so that, for example, '*x* is the same horse as y' is equivalent to '*x* and *y* are horses, and they stand in the single, all-comprehending identity-relation'), and to believe that people and the other objects of our experience are three-dimensional things that persist through time. Is this combination of beliefs consistent? In my view, it is. To believe all these things, it is necessary to believe that a thing can change its parts with the passage of time. And to say that it is possible for a thing to change its parts with the passage of time is to say that situations like the following are possible: Where A, B, C, and D are four distinct things (things with no common parts) it is possible for something—Alice, for example—to be made of A, B, and C on Monday and of B, C, and D on Wednesday. The "Ship of Theseus" argument involved a strand of reasoning that, if it is correct, shows that this is impossible. If we adapt it to the present case, we get something like this: "Look, suppose A still exists on Wednesday, as it very well might. Then all three of A, B, and C still exist on Wednesday. Let us use 'A+B+C' to abbreviate 'the thing made up of A, B, and C' and similarly for 'B+C+D'. You are saying that Alice was A+B+C on Monday and B+C+D on Wednesday. But if A, B, and C exist on Wednesday, A+B+C exists on Wednesday. And if Alice was A+B+C on Monday and was not A+B+C on Wednesday, A+B+C was a different thing on Monday from the thing it was on Wednesday: it was Alice on Monday and something else on Wednesday. But that is just impossible. Whenever something is made of A, B, and C, it is the same thing."

This argument has at least two doubtful premises:

- If A, B, and C exist on Wednesday, something made up of A, B, and C exists on Wednesday.
- Whenever there is something made up of A, B, and C, it is the same thing.

As to the former, why should we suppose that if there are three particular things, there is necessarily something of which those things are parts? As to the latter, why should we suppose that the same three parts could not "add up to" different wholes at different times? Neither of these premises is consistent with our everyday assumptions about parts and wholes. This inconsistency is particularly evident when we consider living organisms.

Let us consider my cat, Taffy—alas, now deceased. Let us call the atoms Taffy was composed of at noon on the first day of 1985 "the New Year's atoms." (There are, of course, a lot more than three New Year's atoms, but the point we are considering does not depend on numbers.) The New Year's atoms are now presumably scattered throughout the biosphere. If, therefore, anything is made of just exactly those atoms right now, it is a sort of rarefied spherical shell about thirteen thousand kilometers across and a few hundreds or thousands of meters thick. We do not normally suppose any such thing exists. And why should we?

Now imagine that, by some fantastic chance, the New Year's atoms were all and only the atoms that at some moment composed a certain fish—it, of course, weighed the same as Taffy did in 1985—that lived in the Indian Ocean four million years ago. When I stroked Taffy on New Year's Day, 1985, was the physical thing I was touching identical with a physical thing that swam in the Indian Ocean four million years ago? We should not normally say that these two things were identical. We should normally say that one was a prehistoric fish, a thing that no longer existed in 1985, and that the other was a cat that was alive and well in 1985. And why shouldn't we say these things?

Now consider the conclusion of the argument, namely, that a thing cannot change its parts. This conclusion does not fare any better than the two premises when we look at it in the light of the things we are normally inclined to believe about living organisms. Here is an interesting statement about living organisms from a book by the great physiologist J. Z. Young:

> The essence of a living thing is that it consists of atoms of the ordinary chemical elements we have listed, caught up into the living system and made part of it for a while. The living activity takes them up and organizes them in its characteristic way. The life of a man consists essentially in the activity he imposes upon that stuff.[3]

Professor Young goes on to describe the imaginary but typical biochemical adventures of a carbon atom that is taken up into the "living system" that is a particular human being and then expelled from that system. When this series of adventures begins, the atom is a part of a sugar molecule that is ingested by the human being. The adventures end when the atom, now a part of a carbon dioxide molecule, leaves that human being in an exhaled breath of air. These adventures—being carried to different parts of the body and participating in an elaborate sequence of chemical reactions—are of great complexity and yet last only a few minutes.

After telling the story of the atom's brief association with a human body, Young asks, "Can we say that [the carbon atom] has ever formed a part of the living tissue of the body?" and goes on to observe, "Many people when asked this question

quickly answer 'No'." But this quick reaction seems to me—as it does to Young—to be wrong. The story of the carbon atom describes just what it is for an atom to come to be and then cease to be a part of a living animal. The life of an animal is a kind of storm of atoms that is constantly, and very rapidly, changing its "membership." Whatever may be true of other physical objects, a living organism would seem not only to be a thing whose parts change with the passage of time, but to be a thing whose very nature demands that it change its parts with the passage of time.

I conclude that physicalists may consistently believe in the identity of the human person across time even if they do not believe in the relativity of identity or in four-dimensionalism. They need only assert that the human person is identical with the human organism and subscribe to the thesis that an organism can change its parts with the passage of time. It is less clear whether a physicalist can, in the end, consistently hold that a human person is some *part* of a human organism, such as a human brain. But we do not really need to investigate this question. We have done enough to show that the "personal-identity" argument against physicalism does not refute physicalism.

Before leaving the topic of the identity of the person across time, let us address a question that may have been troubling some readers. Does not the thesis of physicalism imply that there is no such thing as "life after death"—that it is impossible for the human person to survive biological death? Isn't it clear that if the human person is a physical thing, then its death is the end of it? Isn't it obvious that, whatever physical thing a human person may be (the human organism, the cerebral cortex, . . . ), that physical thing comes to an end with the person's death? It is undeniable that physicalism is incompatible with *certain* beliefs about life after death. It is incompatible with the belief that after our deaths we are going to be "reincarnated" as human beings or as beasts, and it is incompatible with the belief that each of us is a soul that "goes" to heaven or hell after death. And it is true that most beliefs about life after death resemble these two beliefs in some way that renders them incompatible with physicalism.

There is, however, one belief about life after death that is, or at least may be, compatible with physicalism, and that is the Judeo-Christian doctrine of the resurrection of the dead.[4] Unlike many other doctrines of life after death, the doctrine of the resurrection of the dead implies that our future life is not something that happens in the natural course of events but is possible only as the result of a miracle in the strictest and most literal sense of the word. It must be pointed out, however, that the doctrine of the resurrection of the dead, if it is not combined with a belief in a non-physical soul, faces grave metaphysical problems concerning the identity of the resurrected person. If, for example, someone is burned to ashes in a cremation oven, how can any living organism existing at any time in the future be *that* person? Wouldn't the future organism (if it is a human person) have to

be someone else, some new person? If I am to be burned to ashes, and if I do not have a non-physical soul that cannot be destroyed by fire and "carries" my identity with it, how can anything that exists thereafter be *me*? How could even an all-powerful being bring about the identity of myself and something that exists subsequently to my total destruction? This is not the place to discuss these questions, for this is not a book about the philosophy of religion. The thesis of this brief digression is that if the physicalist accepts *any* doctrine of life after death, it must be the doctrine of the resurrection of the dead. It may be that further investigation would show that the physicalist cannot accept even that doctrine and must instead conclude that death is the end of the human person.

This completes our examination of arguments against physicalism. We now turn to arguments for physicalism. There are, I believe, four good arguments for physicalism. Like all philosophical arguments, these arguments are not decisive. To my mind, however, they tip the scale in favor of physicalism. (I do not distinguish between arguments for physicalism and arguments against dualism, since, to my mind, physicalism and dualism are the two most plausible theories about our nature, and an argument against dualism—unless it also tells against physicalism—is therefore an argument for physicalism.)

First, there is the *interaction argument*. We briefly mentioned in Chapter 10 some difficulties with the idea that a non-physical thing could affect a physical thing. Wouldn't that require a violation of well-established physical conservation laws like the law of the conservation of energy or the law of the conservation of momentum? And isn't it also far from clear how a physical thing could affect a non-physical thing? Here is another sort of "interaction" difficulty. The World, the dualist says, contains both non-physical persons and physical organisms. But how do a particular person and a particular organism become "associated"? What brings it about that Jane Tyler interacts with *this* human organism (the one we label 'Jane Tyler's body' precisely because it is the one she interacts with)? The interaction argument comprises these difficulties, together with the observation that by far the most plausible form of dualism is dualistic interactionism.

Secondly, there is the *argument from common speech*. We usually talk and act as if we were visible and tangible. We say things like, "I didn't like the way he was looking at me," or "She reached for the seat belt and buckled herself in." We don't say, "She caused her body's hands to reach for the seat belt and buckle her body in." And, while someone might say, "I didn't like the way he was looking at my body," this would mean something rather special (perhaps, 'I thought he was exhibiting undue sexual interest in me') and it couldn't always be substituted for 'I didn't like the way he was looking at me'. This suggests that our concept of a human person (or our concept of ourselves) is the concept of a thing possessing cer-

tain physical characteristics: we normally conceive of ourselves as things made of flesh and blood and bone and shaped roughly like statues of human beings.

Thirdly, there is an argument I like to call the *remote-control argument*. If dualism is true, our relation to our bodies is analogous to the relation of the operator of a remotely controlled device (such as a radio-controlled model airplane) to that device. Now consider Alfred, who is operating a model airplane by remote control. Suppose that something—an unwary bird or a large hailstone—strikes a heavy blow to the model in midair. If the blow does significant damage to the model, we can expect that both the performance of the model and Alfred's ability to control the model will be impaired. But the blow will have no effect at all on *Alfred*, or no effect beyond his becoming aware of the blow or of some of its effects on the performance of the model and his ability to control it. But if Alfred's *body* were struck a heavy blow, and particularly if it were a blow to the head, this might have an effect on *him*, an effect that goes beyond his becoming aware of the blow and its damaging effects on his body and his ability to control his body: Alfred might well become unconscious.

This is just the sort of effect we should expect if Alfred were a certain human organism, for if the processes of consciousness are certain physical processes within the organism, a damaging blow might well cause those processes to cease, at least temporarily. But what effects should *dualism* lead us to expect from a blow to the body? I submit that if we are non-physical things and if the processes of consciousness are non-physical processes that do not occur within the body, the most natural thing to expect is that (at the worst) we should lose control of our bodies while continuing to be conscious. The blow to the base of Alfred's skull that in fact produces unconsciousness should, according to dualism, produce the following effects on Alfred: he experiences a sharp pain at the base of his skull; he then notes that his body is falling to the floor and that it no longer responds to his will; his visual sensations and the pain at the base of his skull and all the other sensations he has been experiencing fade away; and he is left, as it were, floating in darkness, isolated, but fully conscious and able to contemplate his isolated situation and to speculate about its probable causes and its duration. But this is not what happens when one receives a blow at the base of the skull. One never finds oneself conscious but isolated from one's body.

Dualism, therefore, seems, on the face of it, to make wrong predictions about what the human person will experience in certain situations. Here is another wrong prediction that dualism seems to make: if dualism were correct, we should expect that the ingestion of large quantities of alcohol would result in a partial or complete loss of motor control but leave the mind clear. Physicalism, however, would predict the former effect and it would also strongly suggest that the drinker's mental processes

would be impaired. Because dualism makes (or seems on the face of it to make) these wrong predictions, it is doubtful. I say 'doubtful' rather than 'false', because the defender of dualism will not have too much difficulty in contriving a hypothesis to explain away the fact that a blow to the base of the skull causes one to lose consciousness or the fact that the ingestion of alcohol impairs one's mental processes. For example, the dualist might suggest that a temporary interruption of the normal causal interaction between the person and the body has a traumatic effect on the person, a salient feature of which is loss of consciousness. But this does not change the fact that the typical effects of a blow to the base of the skull are something that has to be *explained away* by dualists and are therefore an embarrassment to them. I say 'is doubtful' rather than 'faces a difficulty' because it is my hope that the reader will find all the hypotheses by which the dualist explains away the observed effects of a blow to the base of the skull (or the ingestion of alcohol) to be implausible and *ad hoc*. I find them so; if I am wrong about the typical reaction of the disinterested reader to these hypotheses, I have claimed too much by using the word 'doubtful'.

Finally, there is the *duplication argument*. This is the single argument for physicalism that I find the most powerful and persuasive. Recall the "duplicating machine" we imagined in Chapter 2, in connection with our discussion of the concept of an intrinsic property. Let us imagine this machine and its operations in a little more detail. The duplicating machine consists of two chambers connected by an impressive mass of science-fictional gadgetry. If you place any physical object inside one of the chambers and press the big red button, a perfect physical duplicate of the object appears in the other chamber. The notion of a perfect physical duplicate may be explained as follows. A physical thing is composed entirely of quarks and electrons. A perfect physical duplicate of the physical thing *x* is a thing composed entirely of quarks and electrons arranged in the same way in relation to one another as the quarks and electrons composing *x* are, and each of the quarks and electrons composing a perfect physical duplicate of *x* will be in the same physical state as the corresponding particle in *x*. If, for example, you place the Koh-i-Noor diamond in one of the chambers and press the button, a thing *absolutely indistinguishable from* the Koh-i-Noor (since it is a perfect physical duplicate of the Koh-i-Noor) will appear in the other. If the two objects are placed side by side and then moved in a rapid and confusing way, so that everyone loses track of which was the original and which the duplicate, no one, no jeweler, mineralogist, or physicist, will ever be able to tell, by any test whatever, which of the two played an important role in the history of the British Raj in the nineteenth century and which was created a moment ago in the duplicating machine.

Now let us consider a second case of duplication. A marble is slowly rolling across the floor of one of the chambers. The button is pressed. There appears on

the floor of the other chamber a marble of the same shape and size and weight and color, rolling in the same direction and at the same speed: our machine reproduces not only the "static" properties of a thing, but also its "dynamic" properties.

Now let us place a living mouse in the chamber and press the button. What will appear in the other chamber? Another living mouse, surely? And wouldn't it be a mouse in every respect interchangeable with the original? If, for example, the original mouse had been taught to get cheese from a cheese dispenser by pressing a lever when a light flashed, wouldn't the new mouse know this trick too? Knowledge of how and when to press the lever to get cheese must somehow be stored in the mouse's little brain, and since the duplicate mouse's brain is a *perfect* duplicate of the original's brain, right down to the sub-atomic level, the same knowledge must be stored in the duplicate brain. (If you put a computer disk containing your novel into the machine, you wouldn't get a blank disk in the other chamber; you'd get another disk containing your novel: in duplicating *every* physical characteristic of the original disk, the machine automatically duplicates those characteristics of the disk that encode a record of the sequence of keystrokes that form your novel.)

And now, finally, let us put *Alfred* into one of the chambers of the duplicating machine and press the button. What do we find in the other chamber? A very intelligent Muslim student of mine once assured me that what one would find would be a dead human body—since the duplicating machine would not reproduce Alfred's soul, which was the principle of life. This dead body, at the instant of its appearance, would be standing just as Alfred stood, and on its face would be an expression just like the expression on Alfred's face. Even in that first instant, however, the body would not be alive, and, having appeared, it would immediately collapse and lie unmoving, its face the blank mask of a corpse. (As a testimony to the general intellectual capacity of my student, I will mention that he was the salutatorian of his graduating class and went on to earn a Ph.D. in nuclear engineering.) I think Plato would have agreed with my student. Descartes, however, would not have agreed. Descartes would have contended that a *living* human body would have appeared in the other chamber. But, Descartes would have said, this body would immediately crumple to the floor. It would then lie there breathing and perhaps drooling, and, if you force-fed it, it would digest the food and in time produce excreta. But it would not *do* anything much. It would just lie there breathing and drooling and digesting and excreting. And this, of course, would be because there was no mind or soul or person in interaction with it. As a consequence, no thought or sensation would be in any way associated with the duplicate body. Life, in the strict, biological sense, was for Descartes (as it was not for Plato or for my student) a purely physical phenomenon; thought and sensation were not. Modern molecular biology, I think, has shown that Descartes was right

about life—or has at least rendered the thesis that life is a complex physical process vastly more probable than its denial. But what about thought and sensation?

That is the question. It is essentially the question whether physicalism is true. The story of the duplicating machine is a device to focus our thoughts as we consider this question. Dualists must say that since thought and sensation are not physical processes occurring within a living human organism, the human body the duplicating machine creates will crumple mindlessly, just as Descartes would have predicted. (I doubt whether many people raised and educated in a European or "European-descended" culture would agree with my Muslim student that the duplicating machine would produce a corpse.) But is this really what any of us believes? Aren't we strongly inclined to believe—at least when we are not considering the consequences of what we believe for the metaphysics of the human person—that the duplicate would "have" thoughts and feelings and beliefs and memories (or what felt like memories; they would not, of course, be connected with past events in the way a real memory is) and desires and emotions? Aren't we strongly inclined to believe that the duplicate would have a conscious mental life like our own and would display the content of this conscious mental life in his observable behavior?

Those who do believe this will concede, after a moment's reflection, that, just as most of the duplicate's memories will not be real memories, so most of his beliefs about himself and his history will be false. The duplicate will, for example, believe that he is Alfred, and he is not. That is, he is not a man who has existed for such-and-such a number of years (he is only a few minutes old) and is married to Winifred (he has never met her), and so on. The duplicate is in no sense Alfred. He is someone else, for if you stick a pin into Alfred, the duplicate feels no pain. Nevertheless, it *seems* to the duplicate that he is Alfred. What it is *like* to be the duplicate is just exactly what it is like to be Alfred. If the two men were "scrambled" (like the two diamonds in our earlier example), no one, including Alfred and the duplicate, could ever know which was Alfred and which was the duplicate. Alfred himself would have to say—at least if he were fully, and perhaps inhumanly, reasonable—"For all I know, I am the duplicate." And if by some chance it were the duplicate that went home to Winifred, she would never suspect that he was not her husband. And just as Winifred would never suspect that anything was amiss, neither would Alfred's children or his mother or his closest friend or his confessor or his psychiatrist.

If this were indeed the outcome of running Alfred through the duplicating machine, dualism would be effectively refuted. The dualist *could*—this sort of thing is almost always possible—contrive some hypothesis that would explain away this outcome. The dualist might, for example, propose that whenever a human body is perfectly duplicated, God creates a perfect duplicate of the non-physical person

who had been interacting with the original body and so arranges matters that the duplicate person is in interaction with the duplicate body. But this would be a desperate move. It would be far more reasonable, even for theists, to conclude that the observed result of our "experiment" should be explained as follows: the thoughts and feelings of a human person are physical processes within a human organism, and, in making a perfect physical duplicate of a human organism, we produce a human organism with the same thoughts and feelings. (The same, that is, at the first moment of the new organism's existence. The thoughts and feelings of the two organisms would probably diverge almost immediately, since the two organisms would probably find themselves almost immediately in different situations.) It would be reasonable to conclude that the mental properties of a human person are related to the physical properties of that person in a way somewhat analogous to the way in which the software associated with a particular computer is related to the physical properties of that computer.

The fact that certain software is associated with (is present in, has been programmed into, is embodied by) a particular computer is as much a physical fact about that computer as are any facts about the hardware constituting the "architecture" of that computer. If I were to take the computer with which I am writing these words and place it (complete with an internal power source and turned on) in the duplicating machine, the computer the machine produced would not be simply another computer of the same make and model; immediately after the duplication, the same words would be visible on its screen, and, like the original, it would be running Microsoft Word X for Mac®, and it would respond in exactly the same way as the original to anything done at the keyboard.

And we have—don't we?—a strong tendency to believe that duplicating a living human organism would have the analogous result as regards the mental life of the human person whose body that organism is: just as, in making a perfect physical duplicate of a working computer, we duplicate all the software programmed into that computer, so, in making a perfect physical duplicate of a living human organism, we duplicate the entire psychology associated with that organism—everything from a neurotic fear of snakes and the ability to speak Russian to a hardly noticeable pain in the left elbow.

Anyone who can honestly reply to this question by saying something along the lines of, "Well, *I* don't observe any such tendency in myself. Like Descartes, I think the duplicate would crumple and fall to the floor and drool," will not be moved by the duplication argument. Anyone who, on reflection, decides that the duplicate would exhibit behavior indistinguishable from Alfred's (in the same situations) should conclude that the duplicate has a mental life like Alfred's and that physicalism is therefore true and dualism false.

This concludes our discussion of the nature of rational beings—or at any rate, of human beings, the only rational beings whose existence is uncontroversial. This discussion has been highly tentative. We should remember that even if we have succeeded in showing that physicalism is the most reasonable theory about the nature of human beings, we have not done anything to dispel the mystery of that nature. Thought and feeling remain as we found them: impenetrable mysteries.

## Suggestions for Further Reading

There are two excellent collections of essays devoted to the problem of personal identity: Perry's *Personal Identity* and Rorty's *The Identities of Persons*. For Judith Jarvis Thomson's reasons for thinking that an explanation of identity across time in terms of four-dimensional objects constitutes "a crazy metaphysic," see her "Parthood and Identity across Time" (a very difficult essay for those who are not formally trained in philosophy). The idea that there is a close analogy between computer hardware and software, on the one hand, and the physical and mental aspects of human beings, on the other, has been extremely influential in philosophy since about the middle of the 1960s. Parts II, III, and IV of Hoffstadter and Dennett's *The Mind's I* provide an excellent introduction to the use philosophers have made of this fascinating idea.

## Notes

1. The three terms 'the Original Ship', 'the Reconstructed Ship', and 'the Continuous Ship' were invented by Jonathan Bennett.

2. This argument is not watertight, even given that a physical thing cannot survive a change of parts. A physicalist *could* maintain that we are physical simples, or that each of us is some composite but very small thing (presumably located inside the brain) that does not change its parts. In fact, one physicalist has maintained this. But few have found it an attractive position. If *I* were convinced that the only way to render physicalism consistent with personal identity across time was to postulate that each person was a tiny object inside that person's brain, I'd become a dualist.

3. J. Z. Young, *An Introduction to the Study of Man* (Oxford: Clarendon Press, 1971), pp. 86–87.

4. In the Hebrew Bible (Daniel 12:2) we read "And many of them that sleep in the dust of the earth shall awake, some to everlasting life and some to shame and everlasting contempt." The Christian "Athanasian Creed" speaks of the resurrection of the dead in these words: "all human beings shall rise again with their bodies and shall give account for their own works. . . ."