CHAPTER TWO

The Turn to Materialism

I. TROUBLES WITH DUALISM

We now skip forward in time to the twentieth and twenty-first centuries. Because of the failures of Cartesian-style dualism, especially the failure to get an adequate or even coherent account of the relationship between the mind and the body, it is widely assumed that substance dualism in any form is out of the question. This is not to say that no serious professionals are substance dualists. But in my experience most substance dualists I know are people who hold this view for some religious reasons, or as part of a religious faith. It is a consequence of substance dualism that when our body is destroyed our soul can continue to survive; and this makes the view appealing to adherents of religions that believe in an afterlife. But among most of the professionals in the field, substance dualism is not regarded as a serious possibility. A prominent exception is the defense of dualism offered by Karl Popper and J. C. Eccles.[1] They claim that there are two quite distinct worlds, World

1 of physical objects and states and World 2 of states of consciousness. Each is a separate and distinct world that interacts with the other. Actually they go Descartes one better and also postulate World 3, a world of "culture in all its manifestations."[2]

All forms of substance dualism inherit Descartes' problem of how to give a coherent account of the causal relations between the soul and the body, but recent versions have an additional problem. It seems impossible to make substance dualism consistent with modern physics. Physics says that the amount of matter/energy in the universe is constant; but substance dualism seems to imply that there is another kind of energy, mental energy or spiritual energy, that is not fixed by physics. So if substance dualism is true then it seems that one of the most fundamental laws of physics, the law of conservation, must be false. Some substance dualists have attempted to cope with this problem by claiming that for each infusion of spiritual energy, there is a diminution of physical energy, thus preserving a constant amount of energy in the universe. Others have said that the mind rearranges the distribution of energy in the universe without adding to it or subtracting from it. Eccles says that the mind can affect the body by altering the probability of neuronal events without any energy input, and that quantum physics enables us to see how this can be done: "The hypothesis of mind-brain interaction is that mental events act by a quantal probability field to alter the probability of emission of vesicles from presynaptic vesicular grids."[3] There is something ad hoc about these maneuvers, in the sense that the authors are convinced in advance of the truth of dualism and are trying to find some way, any way, that will make dualism consistent with physics.

It is important to understand what an extreme doctrine substance dualism is. According to substance dualism our brains and bodies are not really conscious. Your body is just an unconscious machine like your car or your television set. Your body is alive in the way that plants are alive, but there is no consciousness to your body. Rather, your conscious soul is somehow attached to your body and remains attached to it until your body dies, at which time your soul departs. You are identical with your soul and only incidentally and temporarily inhabit this body.

The problem with this view is that, given what we know about how the world works, it is hard to take it seriously as a scientific hypothesis. We know that in humans consciousness cannot exist at all without certain sorts of physical processes going on in the brain. We might, in principle, be able to produce consciousness in some other physical substance, but right now we have no way of knowing how to do this. And the idea that consciousness might be produced apart from any physical substrate whatever, though conceivable, just seems out of the question as a scientific hypothesis.

It is not easy to make the idea that the mind is a separate substance consistent with the rest of what we know about the world. Here are three ways of trying to do it, each with a different conception of the mind.

First, divine intervention. Physical science is incomplete. Our souls are something in addition to the rest of the world. They are created by divine intervention and are not part of the physical world as described by science.

Second, quantum mechanics. The traditional mind-body problem arises only because of an obsolete Newtonian conception of the physical. On one interpretation of

quantum measurement, consciousness is required to complete the collapse of the wave function and thus create quantum particles and events. So some form of consciousness is not created by the rest of nature, rather it is essential for the creation of nature in the first place. It is a primitive part of nature required to explain brain processes and everything else.[4]

Third, idealism. The universe is entirely mental. What we think of as the physical world is just one of the forms that the underlying mental reality takes.[5]

I mention these for the sake of completeness. I do not agree with any of them, and I don't think I understand the second; but as none of them is an influential view in the philosophy of mind, and as I am trying to explain the philosophy of mind, I won't discuss them further in this book.

There is a weaker version of dualism called "property dualism," and that view is fairly widespread. The idea is this: Though there are not two kinds of substances in the world, there are two kinds of properties. Most properties, such as having an electrical charge, or having a certain mass, are physical properties; but some properties, such as feeling a pain or thinking about Kansas City, are mental properties. It is characteristic of human beings that though they are not composed of two different kinds of substances, their physical bodies, and in particular their brains, have not only physical properties, but mental properties as well.

Property dualism avoids postulating a separate mental substance, but it inherits some of the difficulties of substance dualism. What are the relationships between the mental and the physical supposed to be? How is it that physical events can ever cause mental properties? And there is a particular problem that property dualists seem to

be beset with, and that is the problem of how the mental properties, granted that they exist, can ever function causally to produce anything. How can my conscious states, which on this view are not even parts of an extra substance, but merely nonphysical features of my brain, function to cause any physical events in the world? This difficulty, how mental states can ever function causally to produce physical effects, I described in chapter 1 as the problem of "epiphenomenalism." According to epiphenomenalism, mental states do indeed exist but they are epiphenomena. They just go along for the ride; they do not actually have any causal effects. They are like the froth on the wave that comes up on the shore or the flashes of light that glisten off a lake—they are there all right, but they play no significant causal role in the physical world. Indeed, they are worse than the froth and the flash, because they could not play any causal role. The challenge is, How could they play any causal role in determining physical events when they are not themselves physical? If we assume, as it seems we must, that the physical universe is causally closed, in the sense that nothing outside it could have any effects inside; and if we assume, as it seems we must, that consciousness is not part of the physical universe, then it would seem to follow that consciousness can have no effect in the physical universe.

Property dualism does not force us to postulate the existence of a thing that is attached to the body but not really part of the body. But it still forces us to suppose that there are properties of the body, presumably properties of the brain, that are not ordinary physical properties like the rest of our biological makeup. And the problem with this is that we do not see how to fit an account of these properties into

our overall conception of the universe and of how it works. We really do not get out of the postulation of mental entities by calling them properties. We are still postulating nonmaterial mental things. It does not matter whether we say that my conscious pain is a mental property of my brain or that it is an event in my brain. Either way, we are stuck with the traditional difficulties of dualism. One antidualist philosopher characterized these leftover mental phenomena as "nomological danglers" ("nomological" means lawlike). They are produced by the brain in a lawlike fashion, but then they do nothing. They just dangle there.[6]

Many, probably most, philosophers have abandoned dualism, but the situation is odd because to many dualists, the arguments I have just presented do not look at all decisive against all forms of dualism. I think a typical property dualist would say, "OK, the mind is not a separate substance but all the same it is just a brute fact of nature that creatures like us do have pains and tickles and itches, as well as thoughts and emotions and these are not in any ordinary sense physical. Nor are they reducible to anything physical." And indeed some dualists bite the bullet and accept epiphenomenalism.

My guess is that dualism, in spite of being out of fashion, will not go away. Indeed in recent years dualism, at least property dualism, has been making something of a comeback, partly due to a renaissance of interest in consciousness. The insight that drives dualism is powerful. Here is the insight, at its most primitive: we all have real conscious experiences and we know that they are not the same sort of thing as the physical objects around us. This primitive insight can be given a more sophisticated formulation: the world consists almost entirely of physical parti-

cles and everything else is in some way an illusion (like colors and tastes) or a surface feature (like solidity and liquidity) that can be reduced to the behavior of the physical particles. At the level of molecular structure the table is not really solid. It is, as the physicist Eddington said, a cloud of molecules. It is just that from our point of view it seems solid. But at bottom the physical world consists entirely of microentities, the physical particles. However there is one exception. Consciousness is not just particles. In fact it is not particles at all. Whatever else it is, it is something "over and above" the particles. I believe this is the insight that drives contemporary property dualism.

David Chalmers[7] puts the point by saying that it is not logically possible that the course of the physical universe should be different if the course of microphysical facts is the same. Once you have the microphysics then everything else follows. But that is not true for consciousness. You could imagine the whole physical course of the universe exactly the same, minus consciousness. It is logically possible that the course of the physical universe should be exactly as it is, but with no consciousness.

It is such apparent basic differences between the mental and the physical that drives dualism. I think dualism can be answered and refuted, but we do not yet have the tools to do it. I will do it in chapter 4.

II. THE TURN TO MATERIALISM

The dualists said that there are two kinds of things or properties in the universe, and with the failure of dualism, it is natural to suppose that maybe there is only one kind of thing in the universe. Not surprisingly, this view is called

"monism" and it comes in two flavors, mentalist monism and materialist monism. These are called "idealism" and "materialism," respectively. Idealism says that the universe is entirely mental or spiritual; there exists nothing but "ideas" in the technical sense of the word, according to which any mental phenomenon at all is an idea. On some views—for example, Berkeley's—in addition to ideas there are minds that contain the ideas. Idealism had a prodigious influence in philosophy, literally for centuries, but as far as I can tell it has been dead as a doornail among nearly all of the philosophers whose opinions I respect, for many decades, so I will not say much about it. Some of the most famous idealists were Berkeley, Hegel, Bradley, and Royce.

The single most influential family of views in the philosophy of mind throughout the twentieth century and leading into the twenty-first century is one version or another of materialism. Materialism is the view that the only reality that exists is material or physical reality, and consequently if mental states have a real existence, they must in some sense be reducible to, they must be nothing but, physical states of some kind. There is a sense in which materialism is the religion of our time, at least among most of the professional experts in the fields of philosophy, psychology, cognitive science, and other disciplines that study the mind. Like more traditional religions, it is accepted without question and it provides the framework within which other questions can be posed, addressed, and answered. The history of materialism is fascinating, because though the materialists are convinced, with a quasi-religious faith, that their view must be right, they never seem to be able to formulate a version of it that they are completely satisfied with and that can be generally

accepted by other philosophers, even by other materialists. I think this is because they are constantly running up against the fact that the different versions of materialism seem to leave out some essential mental feature of the universe, which we know, independently of our philosophical commitments, to exist. The features they generally leave out are consciousness and intentionality. The problem is to give a completely satisfying materialist account of the mind that does not end up denying the obvious fact that we all intrinsically have conscious states and intentional states. In the next few pages I am going to sketch briefly the history of materialism in the twentieth century, up to the point where it finally reached its most sophisticated formulation in the computational theory of the mind, the theory that the brain is a computer and the mind is a computer program. This sketch is necessarily oversimplified. For reasons of space, I can only hit the high points, but I do want you to see those high points and how they relate to each other. There is a natural progression that leads from behaviorism to the computational theory of the mind and I want you to see that progression.

III. THE SAGA OF MATERIALISM: FROM BEHAVIORISM TO STRONG ARTIFICIAL INTELLIGENCE

Behaviorism

The earliest influential form of materialism in the twentieth century was called "behaviorism." In its crudest version, behaviorism says the mind just is the behavior of the body. There is nothing over and above the behavior of the body

statements about behavior, what the agent would do or would say under such and such circumstances.

According to a typical behaviorist analysis, to say that Jones believes it is going to rain just means the same as saying an indefinite number of statements such as the following: if the windows in Jones's house are open, he will close them; if the garden tools are left outside, he will put them indoors; if he goes for a walk he will carry an umbrella or wear a raincoat or both; and so forth. The idea was that having a mental state was just being disposed to certain sorts of behavior; and the notion of a disposition was to be analyzed in terms of hypothetical statements, statements of the form "If p then q." As applied to the problem of mental states, these statements would take the form, "If such-and-such conditions obtain, then such-and-such behavior will ensue."

Physicalism and the Identity Theory

By the middle decades of the twentieth century, the difficulties of behaviorism had led to its general weakening and eventual rejection. It was going nowhere as a methodological project in psychology, and indeed was under quite effective attack, especially from the linguist Noam Chomsky. Chomsky claimed that the idea that when we study psychology we are studying behavior is as unintelligent as the idea that when we study physics we are studying meter readings. Of course we use behavior as evidence in psychology, just as we use meter readings as evidence in physics, but it is a mistake to confuse the evidence that we have about a subject matter for the subject matter itself. The subject matter of psychology is the human mind, and

human behavior is evidence for the existence and features of the mind, but is not itself the mind.

The difficulties with the logical behaviorists were even more marked. No one had ever given a remotely plausible account of how you could translate statements about minds into statements about behavior. There were various technical difficulties about how to specify the antecedents of the hypotheticals, and especially about how to do it without circularity. I said earlier that the behaviorists would analyze Jones's belief that it is going to rain into sets of statements about his rain-avoidance behavior. But the difficulty with that is that we can only begin to make such a reduction on the assumption that Jones desires to stay dry. So the assumption that Jones will carry an umbrella if he believes that it is going to rain is only plausible if we suppose that Jones does not want to be rained on. But then if we are analyzing belief in terms of desire, it looks like there is a kind of circularity in the reduction. We did not really reduce the belief to behavior; we reduced it to behavior plus desire, which still leaves us with a mental state that needs to be analyzed. Analogous remarks could be made about the reduction of desire. To say that Jones's desire to stay dry consists in such things as his disposition to carry an umbrella will only seem remotely plausible if we suppose that Jones believes it is going to rain.

A second family of difficulties had to do with the causal relations between mental states and behavior. The logical behaviorists had argued that mental states consisted in nothing but behavior and dispositions to behavior, but this runs against our common sense intuition that there is a causal relation between our inner mental states and our outward behavior. My pain causes me to cry out and to take

aspirin; my belief that it is going to rain and my desire to stay dry cause me to take an umbrella, etc., and it seems that this apparent truth is denied by the behaviorists. They cannot account for the causal relations between the inner experience and the external behavior, because they are in effect denying that there is any internal experience in addition to the external behavior.

The real difficulty with behaviorism, though, is that its sheer implausibility became more and more embarrassing. We do have thoughts and feelings and pains and tickles and itches, but it does not seem reasonable to suppose that these are identical with our behavior or even with our dispositions to behavior. The feeling of pain is one thing, pain behavior is something else. Behaviorism is so intuitively implausible that unsympathetic commentators often made fun of it. As early as the 1920s, I. A. Richards pointed out that to be a behaviorist you have to "feign anesthesia."[9] And university lecturers have a stock repertoire of bad jokes about behaviorism. A typical joke: a behaviorist couple just after making love, he says to her "It was great for you. How was it for me?"

The sheer implausibility of behaviorism had become an embarrassment by the 1960s and it was gradually replaced among materialist-minded philosophers by a doctrine called "physicalism," sometimes called the "identity theory." The physicalists said that Descartes was not wrong, as the logical behaviorists had claimed, as a matter of logic, but just as a matter of fact. It might have turned out that we had souls in addition to bodies, but the way that nature in fact turned out, what we think of as minds are just brains, and what we think of as mental states, such as the feeling of pain or having a tickle or an itch, are just

states of the brain—and perhaps the rest of the central nervous system. This thesis was sometimes called the "identity thesis" because it asserted an identity between mental states and brain states. The identity theorists were anxious to insist on the contrast between their view and behaviorism. Behaviorism was supposed to be a logical thesis about the definition of mental concepts. The identity thesis was supposed to be a factual claim, not about the analysis of mental concepts, but rather about the mode of existence of mental states. The model for the behaviorists was one of definitional identities. Pains are dispositions to behavior in a way that triangles are three-sided plane figures. In each case it is a matter of definition. The identity theorists said no, the model is not definitions, but rather empirical discoveries of identities in science. We have discovered, as a matter of fact, that a bolt of lightning is identical with an electrical discharge; we have discovered, as a matter of fact, that water is identical with H_2O, and we are now discovering, and the discovery is proceeding daily, that mental states are really identical with brain states.[10]

Objections to the Identity Theory

There were a number of objections to the identity theory. I find it useful to distinguish between the technical objections and the common-sense objections. The first technical objection was that the theory seemed to violate a principle of logic called "Leibnitz's Law."[11] The law says that if any two things are identical, then they must have all their properties in common. So if you could show that mental states had properties that could not be attributed to brain

states, and brain states had properties that could not be attributed to mental states, it looks like you would refute the identity theory. And it did not seem difficult to provide such examples. So I can say, for example, that the brain state that corresponds to my thought that it is raining is 3 cm inside my left ear; but, according to the objectors, it does not make any sense to say that my thought that it is raining is 3 cm inside my left ear. Furthermore, even for conscious states that have a location, such as pain, the pain may be in my toe, but the brain state that corresponds to that pain is not in my toe, but in my brain. So the properties of the brain state are not the same as the properties of the mental state. Therefore, physicalism is false.

The identity theorists thought that they had an easy answer to these objections. The objections, they say, just rest on ignorance. When we come to know more about the brain, we will come to feel perfectly comfortable in attributing spatial location to mental states and attributing so-called mental properties to states of the brain. And, about the location of the pain in the toe, the identity theorists said that what we are interested in is not a putative object, the pain, but rather the total experience of having the pain. And that total experience extends all the way from the stimulation of the peripheral nerve endings in the toe to the brain itself. I think that the identity theorists were successful in answering this objection, but there were other objections that were more serious.

A common-sense objection to the identity theorists was that if the identity was indeed an empirical identity, something that could be discovered as a matter of fact, on the analogy with water and H_2O, or lightning and electrical discharge, then it seems there would have to be two kinds

of properties to nail down the two sides of the identity statement.[12] Thus, just as the statement, "lightning is identical with an electrical discharge" has to identify one and the same thing in terms of its lightning properties and in terms of its electrical discharge properties; and "water is identical with H_2O molecules" has to identify one and the same thing in terms of its water properties and in terms of its H_2O properties, so the claim that, for example, "pain is identical with a certain type of brain state" has to identify one and the same thing in terms of its pain properties, and in terms of its brain-state properties. But if there are to be two independent sets of properties in the identity statement, then it looks like we have two different types of properties left over: mental properties and physical properties. It looks, in short, as if in order to make the identity thesis work, we have to fall back into property dualism. If all mental states are brain states, then there are two kinds of brain states, those that are mental and those that are not. What is the difference? The mental states have mental properties. The others have only physical properties. And that view sounds like property dualism.

This was a decisive problem for the identity theorists. The whole point of the theory was to vindicate materialism, to show that mental states were really identical with, were nothing but, were reducible to material states of the brain. But if it turns out that the brain states in question have irreducible mental properties then the project fails. It leaves us with an irreducible mental element. In doing research for this book I found at least one philosopher who thought of himself as an identity theorist who seemed willing to embrace this result at least as a possibility.[13] Grover Maxwell calls his view the identity theory, but he says, "the

way is entirely open for speculating that some brain events just are our joys, sorrow, pains, thoughts, etc., in all of their qualitative, and mentalistic richness"(p. 235). This is quite similar to what I think is the correct view and will explain in chapter 4. But it was not a typical view among the identity theorists.

The standard identity theorists' answer to this objection was less plausible than their answer to the Leibnitz Law objections.[14] The answer they gave was that the phenomena in question could be specified without using any mental predicates. They could be specified in a topic-neutral vocabulary. Instead of saying, "There is a yellow-orange afterimage in me," they prefer to say "There is something going on in me that is like what goes on when I see an orange." Such a rephrasing of the identification of the mental states in a "topic-neutral" vocabulary was supposed to answer the objection because it enabled us to specify the mental element in a nonmental, neutral vocabulary: there is this thing going on in me and it can be specified in a way that is neutral between dualism and materialism, but it just turns out that the thing is a brain process. So we can specify the mental feature but in a way that is consistent with materialism.

I think this answer fails. The point that we can talk about mental phenomena without using a mental vocabulary does not change the fact that the mental phenomena continue to have mental properties. My yellow-orange after-image remains qualitative and subjective whether or not we choose to mention those features. If one wanted to refuse to talk about airplanes, one could just say, "some property belonging to United Airlines." But that does not eliminate the existence of airplanes. To put the point

succinctly, the fact that one can mention a phenomenon that is intrinsically qualitative and subjective in a vocabulary that does not reveal these features does not remove the features. In the end of course, the identity theorists wanted to deny that there were any such features, but that requires a separate argument.[15]

One slightly more technical objection that really did concern the identity theorists and indeed eventually forced a modification in their views was the accusation of "neuronal chauvinism."[16] If the claim of the identity theorists was that every pain is identical with a certain kind of neuronal stimulation, and every belief is identical with a certain type of brain state, then it seems that a being that did not have neurons or that did not have the right kind of neurons could not have pains and beliefs. But why can't animals that have brain structures different from ours have mental states? And indeed, why couldn't we build a machine that did not have neurons at all, but also had mental states? This objection led to an important shift in the identity theory from what came to be called "type-type identity theory" to "token-token identity theory." In order to explain this distinction I need to say a bit about the type-token distinction. If I write the word "dog" three times: "dog dog dog," have I written one word or three? Well, I have written three instances, or tokens, of one type of word. So we need a distinction between types, which are abstract general entities, and tokens, which are concrete particular objects and events. A token of a type is a particular concrete exemplification of that abstract general type.

Using this distinction we can see how the identity theorists were motivated to move from a type-type identity theory to a token-token identity theory. The type-type

identity theory says "Every type of mental state is identical with some type of physical state." By their own lights this is a bit sloppy, because the identity in question is between actual concrete tokens and not abstract universal types. What they meant is: for every mental-state type there is some brain-state type such that every token of the mental type is a token of the brain type. The token identity theorists simply said: for every token of a certain type of mental state, there is some token of some type of physical state or other with which that mental state token is identical. They, in short, did not require, for example, that all token pains had to exemplify exactly the same type of brain state. They might be tokens of different types of brain states even though they were all tokens of the same mental type, pain. For that reason they were called "token-token" identity theorists as opposed to "type-type" identity theorists. Token-token identity seems much more plausible than type-type identity theory. Suppose I believe that Denver is the capital of Colorado and suppose you believe that Denver is the capital of Colorado. It seems unnecessary to suppose that in order to have the same belief we must be in exactly the same type of neurobiological state. My neurobiological state of believing that Denver is the capital of Colorado might be at a certain point in my brain, and yours might be at another point, without these being different beliefs.

Unfortunately, the identity theorists were often rather feeble at giving examples. One of their favorite examples was to say that pains are identical with C-fiber stimulations. The idea was that according to the type identity theorists, every pain is identical with some C-fiber stimulation and according to the token identity theorists, this particular pain might be identical with this particular C-fiber stimu-

lation, but some other pain might be identical with some other state of a brain or some other state of a machine. Unfortunately, all of this is rather bad neurophysiology. A C-fiber is a type of axon; and it is true that certain types of pain signals, not all, are carried by C-fibers to the brain. But it would be ridiculous, neurophysiologically, to think there is nothing to pains except having your C-fibers stimulated. The C-fiber is just part of a complex pain mechanism in the brain and nervous system. Be that as it may, this was the sort of example that the identity theorists gave, and a good deal of the debate centered on whether or not we would get such type identities or whether token identities were all that we could hope for. In the long run the token identity theorists have been more influential than the type identity theorists.

But now they are faced with an interesting question. What is it that all of these tokens have in common that makes them tokens of the same mental-state type? If you and I both believe that Denver is the capital of Colorado, then what is it exactly that we share if there is nothing there but our brain states and we have different types of brain states? Notice that the two answers that would traditionally be given to this, the dualist answer and the type-type answer, will not do for the token physicalist. The token physicalists cannot say that what they have in common are the same irreducibly mental properties, because their whole idea was to eliminate, or get rid of, such irreducible mental properties. Nor can they say that they are the same type of brain state, because the whole move from type identity theory to token identity theory was to avoid having to say that every token of a particular mental-state type is identical with a token of a certain brain-state type.