CHAPTER 3

Arguments against Materialism

In the last chapter I presented some of the history of recent materialism, and I considered arguments against some versions, especially against behaviorism, type identity theory, and eliminative materialism. In this chapter I will present the most common arguments against materialism, concentrating on functionalism, because it is currently the most influential version of materialism. In general, these attacks have the same logical structure: the materialist account leaves out some essential feature of the mind such as consciousness or intentionality. In the jargon of philosophers, the materialist analysis fails to give *sufficient* conditions for mental phenomena, because it is possible to satisfy the materialist analysis and not have the appropriate mental phenomena. Strictly speaking, functionalism does not require materialism. The functionalist defines mental states in terms of causal relations and the causal relations could in principle be in anything. It just happens, as the world turned out, that they are in physical brains, physical computers, and other physical systems. The functionalist analysis is supposed to be a conceptual truth that analyzes mental concepts in causal terms. The fact that these causal relations are realized in human brains is an empirical discovery, not a conceptual truth. But the driving motivation for functionalism was a materialist rejection of dualism. Functionalists want to analyze mental phenomena in a way that avoids any reference to anything intrinsically subjective and nonphysical.

I. EIGHT (AND ONE HALF) ARGUMENTS AGAINST MATERIALISM

1. Absent Qualia

Conscious experiences have a qualitative aspect. There is a qualitative feel to drinking beer, which is quite different from the qualitative feel of listening to Beethoven's Ninth Symphony. Several philosophers have found it useful to introduce a technical term to describe this qualitative aspect of consciousness. The term for qualitative states is "qualia," of which the singular is "quale." Each conscious state is a quale, because there is a certain qualitative feel to each state. Now, say the anti-functionalists, the problem with functionalism is that it leaves out qualia. It leaves out the qualitative aspect of our conscious experiences, and thus qualia are absent from the functionalist account. Qualia really exist, so any theory like functionalism that denies their existence, either explicitly or implicitly, is false.

2. Spectrum Inversion

A related argument was advanced by a number of philosophers, and it relies on an old thought experiment, which has occurred to many people in the history of the subject, and to many people outside of philosophy as well.

Let us suppose that neither you nor I is color blind. We both make exactly the same color discriminations. If asked to pick out the red pencils from the green pencils, you and I will both pick out the red pencils. When the traffic light changes from red to green, we both go at once. But let us suppose that, in fact, the inner experiences we have are quite different. If I could have the experience you call "seeing green," I would call it "seeing red." And similarly, if you could have the experience I call "seeing green," you would call it "seeing red." We have, in short, a red-green inversion. This is totally undetectable by any behavioral tests, because the tests identify powers to make discriminations among objects in the world, and not the power to label inner experiences. The inner experiences might be different, even though the external behavior is exactly the same. But if that is possible, then functionalism cannot be giving an account of inner experience, for the inner experience is left out of any functionalist account. The functionalist would give exactly the same account of my experience described by "I see something

green" and your experience described by "I see something green," but the experiences are different, so functionalism is false.

3. Thomas Nagel: What Is It Like to Be a Bat?

One of the earliest well-known arguments against functionalist types of materialism was advanced in an article by Thomas Nagel called, "What It Is Like to Be a Bat?"[1] According to Nagel, the really difficult part of the mind-body problem is the problem of consciousness. Suppose we had a fully satisfactory functionalist, materialist, neurobiological account of various mental states: beliefs, desires, hopes, fears, etc. All the same, such an account would not explain consciousness. Nagel illustrates this with the example of a bat. Bats have a different lifestyle from ours. They sleep all day long, hanging upside down from rafters, and then they fly around at night, navigating by detecting echoes from sonar they bounce off of solid objects. Now, says Nagel, someone might have a complete knowledge of a bat's neurophysiology; he might have a complete knowledge of all the functional mechanisms that enable bats to live and navigate; but all the same, there would be something left out of this person's knowledge: What is it like to be a bat? What does it feel like? And this is the essence of consciousness. For any conscious being, there is a what-it-is-like aspect to his existence. And this is left out of any objective account of consciousness because an objective account cannot explain the subjective character of consciousness.

4. Frank Jackson: What Mary Didn't Know

A similar argument was advanced by the Australian philosopher, Frank Jackson.[2] Jackson imagines a neurobiologist, Mary, who knows all there is to know about color perception. She has a total and complete knowledge of the neurophysiology of our color-perceiving apparatus, and she also has a complete knowledge of the physics of light and of the color spectrum. But, says Jackson, let us imagine that she has been brought up entirely in a black and white environment. She has never seen anything colored, only black, white, and shades of gray. Now, says Jackson, it seems clear that there is something left out of her knowledge. What is left out, for example, is what the color red actually looks like. But, then, it seems that a functionalist or a materialist account of the mind would leave something out, because a person might have the

complete knowledge of all there was to know on a functionalist or materialist account, without knowing what colors look like. And the problem with colors is only a special case of the problem of qualitative experiences generally. Any account of the mind that leaves out these qualitative experiences is inadequate.

5. Ned Block: The Chinese Nation

A fifth argument for the same general antifunctionalist view was advanced by Ned Block.[3] Block says that we might imagine a large population carrying out the steps in a functionalist program of the sort that is presumably carried out by the brain. So, for example, imagine that there are a billion neurons in the brain, and imagine that there are a billion citizens of China. (The figure of a billion neurons is, of course, ludicrously small for the brain, but it does not matter for this argument.) Now we might imagine that just as the brain carries out certain functionalist steps, so we could get the population of China to carry out exactly those steps. But, all the same, the population of China does not thereby have any mental states as a total population in the way that the brain *does* have mental states.

6. Saul Kripke: Rigid Designators

A purely logical argument was advanced by Saul Kripke[4] against any version of the identity theory. Kripke's argument appeals to the concept of a "rigid designator." A rigid designator is defined as an expression that always refers to the same object in any possible state of affairs. Thus, the expression, "Benjamin Franklin," is a rigid designator because in the usage that I am now invoking, it always refers to the same man. This is not to say, of course, that I cannot name my dog "Benjamin Franklin," but, then, that is a different usage, a different meaning of the expression. On the standard meaning, "Benjamin Franklin" is a rigid designator. But the expression, "The inventor of daylight saving time," though it also refers to Benjamin Franklin, is not a rigid designator because it is easy to imagine a world in which Benjamin Franklin was not the inventor of daylight saving time. It makes sense to say that someone else, other than the actual inventor, might have been the inventor of daylight saving time, but it makes no sense to say that someone else, other than Benjamin Franklin, might have been Benjamin Franklin. For these reasons, "Benjamin Franklin" is a rigid designator, but "the inventor of daylight saving time" is nonrigid.

With the notion of rigid designators in hand, Kripke then proceeds to examine identity statements. His claim is that identity statements, where one term is rigid and the other not rigid, are in general not necessarily true; they might turn out to be false. Thus, the sentence, "Benjamin Franklin is identical with the inventor of daylight saving time," is true, but only contingently true. We can imagine a world in which it is false. But, says Kripke, where both sides of the identity statement are rigid, the statement, if true, must be necessarily true. Thus, the statement, "Samuel Clemens is identical with Mark Twain," is necessarily true because there cannot be a world in which Samuel Clemens exists, and Mark Twain exists, but they are two different people. Similarly with words naming kinds of things. Water is identical with H_2O, and because both expressions are rigid, the identity must be necessary. And here is the relevance to the mind-body problem: if we have on the left hand side of our identity statement an expression referring to a type of mental state rigidly, and on the right hand side, an expression referring to a type of brain state rigidly, then the statement, if true, would have to be necessarily true. Thus, if pains really were identical with C-fiber stimulations, then the statement, "Pain = C-fiber stimulation," would have to be necessarily true, if it were to be true at all. But, it is clearly not necessarily true. For even if there is a strict correlation between pains and C-fiber stimulations, all the same, it is easy to imagine that a pain might exist without a C-fiber stimulation existing, and a C-fiber stimulation might exist without a corresponding pain. But, if that is so, then the identity statement is not necessarily true, and if it is not necessarily true, it cannot be true at all. Therefore, it is false. And what goes for the identification of pains with neurobiological events goes for any identification of conscious mental states with physical events.

7. John Searle: The Chinese Room

An argument explicitly directed against Strong AI was put forth by the present author.[5] The strategy of the argument is to appeal to one's first person experiences in testing any theory of the mind. If Strong AI were true, then anybody should be able to acquire any cognitive capacity just by implementing the computer program simulating that cognitive capacity. Let us try this with Chinese. I do not, as a matter of fact, understand any Chinese at all. I cannot even tell Chinese writing from Japanese writing. But, we imagine that I am locked in a room with boxes full of Chinese symbols, and I have a

rule book, in effect, a computer program, that enables me to answer questions put to me in Chinese. I receive symbols that, unknown to me, are questions; I look up in the rule book what I am supposed to do; I pick up symbols from the boxes, manipulate them according to the rules in the program, and hand out the required symbols, which are interpreted as answers. We can suppose that I pass the Turing test for understanding Chinese, but, all the same, I do not understand a word of Chinese. And if I do not understand Chinese on the basis of implementing the right computer program, then neither does any other computer just on the basis of implementing the program, because no computer has anything that I do not have.

You can see the difference between computation and real understanding if you imagine what it is like for me also to answer questions in English. Imagine that in the same room I am given questions in English, which I then answer. From the outside my answers to the English and the Chinese questions are equally good. I pass the Turing test for both. But from the inside, there is a tremendous difference. What is the difference exactly? In English, I understand what the words mean; in Chinese, I understand nothing. In Chinese, I am just a computer.

The Chinese Room Argument struck at the heart of the Strong AI project. Prior to its publication, attacks on artificial intelligence usually took the form of saying that the human mind has certain abilities that the computer does not have and could not acquire.[6] This is always a dangerous strategy, because as soon as someone says that there is a certain sort of task that computers cannot do, the temptation is very strong to design a program that performs precisely that task. And this has often happened. When it happens, the critics of artificial intelligence usually say that the task was not all that important anyway and the computer successes do not really count. The defenders of artificial intelligence feel, with some justice, that the goal posts are being constantly moved. The Chinese Room Argument adopted a totally different strategy. It assumes complete success on the part of artificial intelligence in simulating human cognition. It assumes that AI researchers can design a program that passes the Turing test for understanding Chinese or anything else. All the same, as far as human cognition is concerned, such achievements are simply irrelevant. And they are irrelevant for a deep reason: the computer operates by manipulating symbols. Its processes are defined purely syntactically, whereas the human mind has more than just uninterpreted symbols, it attaches meanings to the symbols.

There is a further development of the argument that seems to me more powerful though it received much less attention than the original Chinese Room Argument. In the original argument I assumed that the attribution of syntax and computation to the system was unproblematic. But if you think about it you will see that *computation and syntax are observer relative.* Except for cases where a person is actually computing in his own mind there are no intrinsic or original computations in nature. When I add two plus two to get four, that computation is not observer relative. I am doing that regardless of what anybody thinks. But when I punch "2+2 =" on my pocket calculator and it prints out "4" it knows nothing of computation, arithmetic, or symbols, because it knows nothing about anything. Intrinsically it is a complex electronic circuit that we *use* to compute with. The electrical state transitions are intrinsic to the machine, but the computation is in the eye of the beholder. What goes for the calculator goes for any commercial computer. The sense in which computation is in the machine is the sense in which information is in a book. It is there alright, but it is observer relative and not intrinsic. For this reason you could not discover that the brain is a digital computer, because computation is not discovered in nature, it is assigned to it. So the question, Is the brain a digital computer? is ill defined. If it asks, Is the brain intrinsically a digital computer? the answer is that nothing is intrinsically a digital computer except for conscious agents thinking through computations. If it asks, Could we assign a computational interpretation to the brain? the answer is that we can assign a computational interpretation to anything.

I do not develop the argument here but I want you to know at least the bare bones of the argument. For a fuller statement of it see *The Rediscovery of the Mind*, chapter 9.[7]

8. The Conceivability of Zombies

One of the oldest arguments, and in a way the underlying argument in several of the others, is this: it is conceivable that there could be a being who was physically exactly like me in every respect but who was totally without any mental life at all. On one version of this argument it is logically possible that there might be a zombie who was exactly like me, molecule for molecule, but who had no mental life at all. In philosophy a zombie is a system that behaves just like humans but has no mental life, no consciousness or real intentionality; and this argument claims that zombies are logically possible. And if zombies are even logically possible, that

is, if it is logically possible that a system might have all the right behavior and all the right functional mechanisms and even the right physical structure while still having no mental life, then the behaviorist and functionalist analyses are mistaken. They do not state logically sufficient conditions for having a mind.

This argument occurs in various forms. One of the earliest contemporary statements is by Thomas Nagel.[8] Nagel argues, "I can conceive of my body doing precisely what it is doing now, inside and out, with complete physical causation of its behavior (including typically self-conscious behavior), but without any of the mental states which I am now experiencing, or any others, for that matter. If that is really conceivable, then the mental states must be distinct from the body's physical state." This is a kind of mirror image of Descartes' argument. Descartes argued that it is conceivable that my mind could exist without my body, therefore my mind cannot be identical with my body. And this argument says it is conceivable that my body could exist and be exactly as it is, but without my mind, therefore my mind is not identical with my body, or any part of, or any functioning of my body.

9. The Aspectual Shape of Intentionality

The final argument I can present only in an abbreviated form (hence I call it half an argument) because I haven't yet explained intentionality in enough detail to spell it out fully. But I think I can give you a clear enough idea of how it goes. Intentional states, like beliefs and desires, represent the world under some aspects and not others. For example, the desire for water is not the same as the desire for H_2O, because a person might desire water without knowing that it is H_2O and even believing that it is not H_2O. Because all intentional states represent under aspects we might say that all intentional states have an aspectual shape. But a causal account of intentionality such as the one given by functionalists cannot capture differences in aspectual shape because causation does not have this kind of aspectual shape. Whatever water causes, H_2O causes; and whatever causes water, causes H_2O. The functionalist analysis of my belief that this stuff is water and my desire for water given in causal terms can't distinguish this belief and desire from my belief that this stuff is H_2O and my desire for H_2O. But they are clearly distinct, so functionalism fails.

And you cannot answer this argument by saying that we could ask the person, "Do you believe that this stuff is water? Do you believe

that this stuff is H_2O?" because the problem we had about belief and desire now arises for *meaning*. How do we know that the person means by "H_2O" what we mean by "H_2O," and by "water" what we mean by "water"? If all we have to go on is behavior and causal relations, they are not enough to distinguish different meanings in the head of the agent. In short, alternative and inconsistent translations will be consistent with all the causal and behavioral facts.[9]

I have not seen this argument stated before and it only occurred to me when writing this book. To summarize it in the jargon I will explain in chapter 6, intentionality essentially involves aspectual shape. All mental representation is under representational aspects. Causation also has aspects but they are not representational aspects. You can't analyze mental concepts in causal terms because the representational aspectual shape of the intentional gets lost in the translation. This is why statements about intentionality are intensional-with-an-s, but statements about causation, of the form A caused B, are extensional. (Don't worry if you don't understand this paragraph. We will get there in chapter 6.)