

brain and nervous system. What light (we may ask) is thrown by neurophysiology, neurochemistry, and comparative neuroanatomy on such matters as mental illness, learning, three-dimensional vision, and the mental life of dolphins? The answer is, "Considerable light," although neuroscientists will be the first to admit that they have only scratched the surface.

I have included these chapters to provide at least an instructive sampling of the research currently under way in these fields. They are certainly not adequate to introduce an aspiring computer scientist or neuroscientist to these fields. But they will provide some real understanding of how empirical research bears on the philosophical issues discussed in this text. (That is important because, as I hope to make clear, most of those philosophical issues are ultimately empirical in character. They will be decided by the comparative success and the relative progress displayed by alternative scientific research programs.) These chapters will also provide a lasting conceptual framework from which to address future developments concerning the mind. And they may whet your appetite for more empirical information. If they do only that, they will have served their purpose.

The concluding chapter is overtly speculative, as befits a concluding chapter, and opens with an attempt to estimate the distribution of conscious intelligence in the universe at large. Intelligence appears likely to be a fairly widespread phenomenon in the universe, and all advanced instances of it will inevitably face the problem of constructing a useful conception of just what intelligence *is*. That process of self-discovery, to judge from our own case, need not be an easy one. Neither will it be completed in a short period, if indeed it can ever be truly *completed*. But progress is still possible, here, as elsewhere in the human endeavor; and we must be prepared to contemplate revolutions in our conception of what *we* are, just as we have successfully navigated repeated revolutions in our conception of the universe that embeds us. The final section scouts the consequences of such a conceptual revolution for the contents of human self-consciousness.

This concludes my set of promissory notes. Let us now turn to the issues themselves.

Chapter 2

The Ontological Problem (the Mind-Body Problem)

What is the real nature of mental states and processes? In what medium do they take place, and how are they related to the physical world? Will my consciousness survive the disintegration of my physical body? Or will it disappear forever as my brain ceases to function? Is it possible that a purely physical system such as a computer could be constructed so as to enjoy real conscious intelligence? Where do minds come from? What are they?

These are some of the questions we shall confront in this chapter. Which answers we should give to them depends on which theory of mind proves to be the most reasonable theory on the evidence, to have the greatest explanatory power, predictive power, coherence, and simplicity. Let us examine the available theories, and the considerations that weigh for and against each.

1. Dualism

The dualistic approach to mind encompasses several quite different theories, but they are all agreed that the essential nature of conscious intelligence resides in something *nonphysical*, in something forever beyond the scope of sciences like physics, neurophysiology, and computer science. Dualism is not the most widely held view in the current philosophical and scientific community, but it is the most common theory of mind in the public at large, it is deeply entrenched in most of the world's popular religions, and it has been the dominant theory of mind for most of Western history. It is thus an appropriate place to begin our discussion.

Substance Dualism

The distinguishing claim of this view is that each mind is a distinct nonphysical thing, an individual 'package' of nonphysical substance, a thing whose identity is independent of any physical body to which it may be temporarily 'attached'. Mental states and activities derive

their special character, on this view, from their being states and activities of this unique, nonphysical substance.

This leaves us wanting to ask for more in the way of a *positive* characterization of the proposed mind-stuff. It is a frequent complaint with the substance dualist's approach that his characterization of it is so far almost entirely negative. This need not be a fatal flaw, however, since we no doubt have much to learn about the underlying nature of mind, and perhaps the deficit here can eventually be made good. On this score, the philosopher René Descartes (1596–1650) has done as much as anyone to provide a positive account of the nature of the proposed mind-stuff, and his views are worthy of examination.

Descartes theorized that reality divides into two basic kinds of substance. The first is ordinary matter, and the essential feature of this kind of substance is that it is extended in space: any instance of it has length, breadth, height, and occupies a determinate position in space. Descartes did not attempt to play down the importance of this type of matter. On the contrary, he was one of the most imaginative physicists of his time, and he was an enthusiastic advocate of what was then called "the mechanical philosophy". But there was one isolated corner of reality he thought could not be accounted for in terms of the mechanics of matter: the conscious reason of Man. This was his motive for proposing a second and radically different kind of substance, a substance that has no spatial extension or spatial position whatever, a substance whose essential feature is the activity of *thinking*. This view is known as *Cartesian dualism*.

As Descartes saw it, the real *you* is not your material body, but rather a nonspatial thinking substance, an individual unit of mind-stuff quite distinct from your material body. This nonphysical mind is in systematic causal interaction with your body. The physical state of your body's sense organs, for example, causes visual/auditory/tactile experiences in your mind. And the desires and decisions of your nonphysical mind cause your body to behave in purposeful ways. Its causal connections to your mind are what make your body yours, and not someone else's.

The main reasons offered in support of this view were straightforward enough. First, Descartes thought that he could determine, by direct introspection alone, that he was essentially a thinking substance and nothing else. And second, he could not imagine how a purely physical system could ever use *language* in a relevant way, or engage in mathematical *reasoning*, as any normal human can. Whether these are good reasons, we shall discuss presently. Let us first notice a difficulty that even Descartes regarded as a problem.

If 'mind-stuff' is so utterly different from 'matter-stuff' in its nature—different to the point that it has no mass whatever, no shape whatever,

and no position anywhere in space—then how is it possible for my mind to have any causal influence on my body at all? As Descartes himself was aware (he was one of the first to formulate the law of the conservation of momentum), ordinary matter in space behaves according to rigid laws, and one cannot get bodily movement (= momentum) from nothing. How is this utterly insubstantial 'thinking substance' to have any influence on ponderous matter? How can two such different things be in any sort of causal contact? Descartes proposed a very subtle material substance—'animal spirits'—to convey the mind's influence to the body in general. But this does not provide us with a solution, since it leaves us with the same problem with which we started: how something ponderous and spatial (even 'animal spirits') can interact with something entirely nonspatial.

In any case, the basic principle of division used by Descartes is no longer as plausible as it was in his day. It is now neither useful nor accurate to characterize ordinary matter as that-which-has-extension-in-space. Electrons, for example, are bits of matter, but our best current theories describe the electron as a point-particle with no extension whatever (it even lacks a determinate spatial position). And according to Einstein's theory of gravity, an entire star can achieve this same status, if it undergoes a complete gravitational collapse. If there truly is a division between mind and body, it appears that Descartes did not put his finger on the dividing line.

Such difficulties with Cartesian dualism provide a motive for considering a less radical form of substance dualism, and that is what we find in a view I shall call *popular dualism*. This is the theory that a person is literally a 'ghost in a machine', where the machine is the human body, and the ghost is a spiritual substance, quite unlike physical matter in its internal constitution, but fully possessed of spatial properties even so. In particular, minds are commonly held to be *inside* the bodies they control: inside the head, on most views, in intimate contact with the brain.

This view need not have the difficulties of Descartes'. The mind is right there in contact with the brain, and their interaction can perhaps be understood in terms of their exchanging energy of a form that our science has not yet recognized or understood. Ordinary matter, you may recall, is just a form or manifestation of energy. (You may think of a grain of sand as a great deal of energy condensed or frozen into a small package, according to Einstein's relation, $E = mc^2$.) Perhaps mind-stuff is a well-behaved form or manifestation of energy also, but a different form of it. It is thus *possible* that a dualism of this alternative sort be consistent with familiar laws concerning the conservation of

momentum and energy. This is fortunate for dualism, since those particular laws are very well established indeed.

This view will appeal to many for the further reason that it at least holds out the possibility (though it certainly does not guarantee) that the mind might survive the death of the body. It does not guarantee the mind's survival because it remains possible that the peculiar form of energy here supposed to constitute a mind can be produced and sustained only in conjunction with the highly intricate form of matter we call the brain, and must disintegrate when the brain disintegrates. So the prospects for surviving death are quite unclear even on the assumption that popular dualism is true. But even if survival were a clear consequence of the theory, there is a pitfall to be avoided here. Its promise of survival might be a reason for *wishing* dualism to be true, but it does not constitute a reason for *believing* that it is true. For that, we would need independent empirical evidence that minds do indeed survive the permanent death of the body. Regrettably, and despite the exploitative blatherings of the supermarket tabloids (**TOP DOCS PROVE LIFE AFTER DEATH!!!**), we possess no such evidence.

As we shall see later in this section, when we turn to evaluation, positive evidence for the existence of this novel, nonmaterial, thinking *substance* is in general on the slim side. This has moved many dualists to articulate still less extreme forms of dualism, in hopes of narrowing further the gap between theory and available evidence.

Property Dualism

The basic idea of the theories under this heading is that while there is no *substance* to be dealt with here beyond the physical brain, the brain has a special set of *properties* possessed by no other kind of physical object. It is these special properties that are nonphysical: hence the term *property dualism*. The properties in question are the ones you would expect: the property of having a pain, of having a sensation of red, of thinking that *P*, of desiring that *Q*, and so forth. These are the properties that are characteristic of conscious intelligence. They are held to be nonphysical in the sense that they cannot ever be reduced to or explained solely in terms of the concepts of the familiar physical sciences. They will require a wholly new and autonomous science—the 'science of mental phenomena'—if they are ever to be adequately understood.

From here, important differences among the positions emerge. Let us begin with what is perhaps the oldest version of property dualism: *epiphenomenalism*. This term is rather a mouthful, but its meaning is simple. The Greek prefix "epi-" means "above", and the position at issue holds that mental phenomena are not a part of the physical

phenomena in the brain that ultimately determine our actions and behavior, but rather ride 'above the fray'. Mental phenomena are thus *epiphenomena*. They are held to just appear or emerge when the growing brain passes a certain level of complexity.

But there is more. The epiphenomenalist holds that while mental phenomena are caused to occur by the various activities of the brain, *they do not have any causal effects in turn*. They are entirely impotent with respect to causal effects on the physical world. They are *mere* epiphenomena. (To fix our ideas, a vague metaphor may be helpful here. Think of our conscious mental states as little sparkles of shimmering light that occur on the wrinkled surface of the brain, sparkles which are caused to occur by physical activity in the brain, but which have no causal effects on the brain in return.) This means that the universal conviction that one's actions are determined by one's desires, decisions, and volitions is false! One's actions are exhaustively determined by physical events in the brain, which events *also* cause the epiphenomena we call desires, decisions, and volitions. There is therefore a constant conjunction between volitions and actions. But according to the epiphenomenalist, it is mere illusion that the former cause the latter.

What could motivate such a strange view? In fact, it is not too difficult to understand why someone might take it seriously. Put yourself in the shoes of a neuroscientist who is concerned to trace the origins of behavior back up the motor nerves to the active cells in the motor cortex of the cerebrum, and to trace in turn their activity into inputs from other parts of the brain, and from the various sensory nerves. She finds a thoroughly physical system of awesome structure and delicacy, and much intricate activity, all of it unambiguously chemical or electrical in nature, and she finds no hint at all of any nonphysical inputs of the kind that substance dualism proposes. What is she to think? From the standpoint of her researches, human behavior is exhaustively a function of the activity of the physical brain. And this opinion is further supported by her confidence that the brain has the behavior-controlling features it does exactly because those features have been ruthlessly selected for during the brain's long evolutionary history. In sum, the seat of human behavior appears entirely physical in its constitution, in its origins, and in its internal activities.

On the other hand, our neuroscientist has the testimony of her own introspection to account for as well. She can hardly deny that she has experiences, beliefs, and desires, nor that they are connected in some way with her behavior. One bargain that can be struck here is to admit the *reality* of mental properties, as nonphysical properties, but demote them to the status of impotent epiphenomena that have nothing to do

with the scientific explanation of human and animal behavior. This is the position the epiphenomenalist takes, and the reader can now perceive the rationale behind it. It is a bargain struck between the desire to respect a rigorously scientific approach to the explanation of behavior, and the desire to respect the testimony of introspection.

The epiphenomenalist's 'demotion' of mental properties—to causally impotent by-products of brain activity—has seemed too extreme for most property dualists, and a theory closer to the convictions of common sense has enjoyed somewhat greater popularity. This view, which we may call *interactionist property dualism*, differs from the previous view in only one essential respect: the interactionist asserts that mental properties do indeed have causal effects on the brain, and thereby, on behavior. The mental properties of the brain are an integrated part of the general causal fray, in systematic interaction with the brain's physical properties. One's actions, therefore, are held to be caused by one's desires and volitions after all.

As before, mental properties are here said to be *emergent* properties, properties that do not appear at all until ordinary physical matter has managed to organize itself, through the evolutionary process, into a system of sufficient complexity. Examples of properties that are emergent in this sense would be the property of being *solid*, the property of being *colored*, and the property of being *alive*. All of these require matter to be suitably organized before they can be displayed. With this much, any materialist will agree. But any property dualist makes the further claim that mental states and properties are *irreducible*, in the sense that they are not just organizational features of physical matter, as are the examples cited. They are said to be novel properties beyond prediction or explanation by physical science.

This last condition—the irreducibility of mental properties—is an important one, since this is what makes the position a dualist position. But it sits poorly with the joint claim that mental properties emerge from nothing more than the organizational achievements of physical matter. If that is how mental properties are produced, then one would expect a physical account of them to be possible. The simultaneous claim of evolutionary emergence and physical irreducibility is *prima facie* puzzling.

A property dualist is not absolutely bound to insist on both claims. He could let go the thesis of evolutionary emergence, and claim that mental properties are *fundamental* properties of reality, properties that have been here from the universe's inception, properties on a par with length, mass, electric charge, and other fundamental properties. There is even an historical precedent for a position of this kind. At the turn of this century it was still widely believed that electromagnetic phe-

nomena (such as electric charge and magnetic attraction) were just an unusually subtle manifestation of purely *mechanical* phenomena. Some scientists thought that a reduction of electromagnetics to mechanics was more or less in the bag. They thought that radio waves, for example, would turn out to be just travelling oscillations in a very subtle but jellylike aether that fills space everywhere. But the aether turned out not to exist. So electromagnetic properties turned out to be fundamental properties in their own right, and we were forced to add electric charge to the existing list of fundamental properties (mass, length, and duration).

Perhaps mental properties enjoy a status like that of electromagnetic properties: irreducible, but not emergent. Such a view may be called *elemental-property dualism*, and it has the advantage of clarity over the previous view. Unfortunately, the parallel with electromagnetic phenomena has one very obvious failure. Unlike electromagnetic properties, which are displayed at all levels of reality from the subatomic level on up, mental properties are displayed only in large physical systems that have evolved a very complex internal organization. The case for the evolutionary emergence of mental properties through the organization of matter is extremely strong. They do not appear to be basic or elemental at all. This returns us, therefore, to the issue of their irreducibility. Why should we accept this most basic of the dualist's claims? Why be a dualist?

Arguments for Dualism

Here we shall examine some of the main considerations commonly offered in support of dualism. Criticism will be postponed for a moment so that we may appreciate the collective force of these supporting considerations.

A major source of dualistic convictions is the religious belief many of us bring to these issues. Each of the major religions is in its way a theory about the cause or purpose of the universe, and Man's place within it, and many of them are committed to the notion of an immortal soul—that is, to some form of substance dualism. Supposing that one is consistent, to consider disbelieving dualism is to consider disbelieving one's religious heritage, and some of us find that difficult to do. Call this the *argument from religion*.

A more universal consideration is the *argument from introspection*. The fact is, when you center your attention on the contents of your consciousness, you do not clearly apprehend a neural network pulsing with electrochemical activity: you apprehend a flux of thoughts, sensations, desires, and emotions. It seems that mental states and properties, as revealed in introspection, could hardly be more different from physical

states and properties if they tried. The verdict of introspection, therefore, seems strongly on the side of some form of dualism—on the side of property dualism, at a minimum.

A cluster of important considerations can be collected under the *argument from irreducibility*. Here one points to a variety of mental phenomena where it seems clear that no purely physical explanation could possibly account for what is going on. Descartes has already cited our ability to use language in a way that is relevant to our changing circumstances, and he was impressed also with our faculty of Reason, particularly as it is displayed in our capacity for mathematical reasoning. These abilities, he thought, must surely be beyond the capacity of any physical system. More recently, the introspectible qualities of our sensations (sensory 'qualia'), and the meaningful content of our thoughts and beliefs, have also been cited as phenomena that will forever resist reduction to the physical. Consider, for example, seeing the color or smelling the fragrance of a rose. A physicist or chemist might know everything about the molecular structure of the rose, and of the human brain, argues the dualist, but that knowledge would not enable him to predict or anticipate the quality of these inexpressible experiences.

Finally, parapsychological phenomena are occasionally cited in favor of dualism. Telepathy (mind reading), precognition (seeing the future), telekinesis (thought control of material objects), and clairvoyance (knowledge of distant objects) are all awkward to explain within the normal confines of psychology and physics. If these phenomena are real, they might well be reflecting the superphysical nature that the dualist ascribes to the mind. Trivially they are *mental* phenomena, and if they are also forever beyond physical explanation, then at least some mental phenomena must be irreducibly nonphysical.

Collectively, these considerations may seem compelling. But there are serious criticisms of each, and we must examine them as well. Consider first the argument from religion. There is certainly nothing wrong in principle with appealing to a more general theory that bears on the case at issue, which is what the appeal to religion amounts to. But the appeal can only be as good as the scientific credentials of the religion(s) being appealed to, and here the appeals tend to fall down rather badly. In general, attempts to decide scientific questions by appeal to religious orthodoxy have a very sorry history. That the stars are other suns, that the earth is not the center of the universe, that diseases are caused by microorganisms, that the earth is billions of years old, that life is a physicochemical phenomenon; all of these crucial insights were strongly and sometimes viciously resisted, because the dominant religion of the time happened to think otherwise. Giordano Bruno was

burned at the stake for urging the first view; Galileo was forced by threat of torture in the Vatican's basement to recant the second view; the firm belief that disease was a punishment visited by the Devil allowed public health practices that brought chronic plagues to most of the cities of Europe; and the age of the earth and the evolution of life were forced to fight an uphill battle against religious prejudice even in an age of supposed enlightenment.

History aside, the almost universal opinion that one's own religious convictions are the reasoned outcome of a dispassionate evaluation of all of the major alternatives is almost demonstrably false for humanity in general. If that really were the genesis of most people's convictions, then one would expect the major faiths to be distributed more or less randomly or evenly over the globe. But in fact they show a very strong tendency to cluster: Christianity is centered in Europe and the Americas, Islam in Africa and the Middle East, Hinduism in India, and Buddhism in the Orient. Which illustrates what we all suspected anyway: that *social forces* are the primary determinants of religious belief for people in general. To decide scientific questions by appeal to religious orthodoxy would therefore be to put social forces in place of empirical evidence. For all of these reasons, professional scientists and philosophers concerned with the nature of mind generally do their best to keep religious appeals out of the discussion entirely.

The argument from introspection is a much more interesting argument, since it tries to appeal to the direct experience of everyman. But the argument is deeply suspect, in that it assumes that our faculty of inner observation or introspection reveals things as they really are in their innermost nature. This assumption is suspect because we already know that our other forms of observation—sight, hearing, touch, and so on—do no such thing. The red surface of an apple does not *look* like a matrix of molecules reflecting photons at certain critical wavelengths, but that is what it is. The sound of a flute does not *sound* like a sinusoidal compression wave train in the atmosphere, but that is what it is. The warmth of the summer air does not *feel* like the mean kinetic energy of millions of tiny molecules, but that is what it is. If one's pains and hopes and beliefs do not *introspectively* seem like electrochemical states in a neural network, that may be only because our faculty of introspection, like our other senses, is not sufficiently penetrating to reveal such hidden details. Which is just what one would expect anyway. The argument from introspection is therefore entirely without force, unless we can somehow argue that the faculty of introspection is quite different from all other forms of observation.

The argument from irreducibility presents a more serious challenge, but here also its force is less than first impression suggests. Consider first our capacity for mathematical reasoning which so impressed Des-

cartes. The last ten years have made available, to anyone with fifty dollars to spend, electronic calculators whose capacity for mathematical reasoning—the calculational part, at least—far surpasses that of any normal human. The fact is, in the centuries since Descartes' writings, philosophers, logicians, mathematicians, and computer scientists have managed to isolate the general principles of mathematical reasoning, and electronics engineers have created machines that compute in accord with those principles. The result is a hand-held object that would have astonished Descartes. This outcome is impressive not just because machines have proved capable of some of the capacities boasted by human reason, but because some of those achievements invade areas of human reason that past dualistic philosophers have held up as forever closed to mere physical devices.

Although debate on the matter remains open, Descartes' argument from language use is equally dubious. The notion of a *computer language* is by now a commonplace: consider BASIC, PASCAL, FORTRAN, APL, LISP, and so on. Granted, these artificial 'languages' are much simpler in structure and content than human natural language, but the differences may be differences only of degree, and not of kind. As well, the theoretical work of Noam Chomsky and the generative grammar approach to linguistics have done a great deal to explain the human capacity for language use in terms that invite simulation by computers. I do not mean to suggest that truly conversational computers are just around the corner. We have a great deal yet to learn, and fundamental problems yet to solve (mostly having to do with our capacity for inductive or theoretical reasoning). But recent progress here does nothing to support the claim that language use must be forever impossible for a purely physical system. On the contrary, such a claim now appears rather arbitrary and dogmatic, as we shall see in chapter 6.

The next issue is also a live problem: How can we possibly hope to explain or to predict the intrinsic qualities of our sensations, or the meaningful content of our beliefs and desires, in purely physical terms? This is a major challenge to the materialist. But as we shall see in later sections, active research programs are already under way on both problems, and positive suggestions are being explored. It is in fact not impossible to imagine how such explanations might go, though the materialist cannot yet pretend to have solved either problem. Until he does, the dualist will retain a bargaining chip here, but that is about all. What the dualists need in order to establish their case is the conclusion that a physical reduction is outright impossible, and that is a conclusion they have failed to establish. Rhetorical questions, like the one that opens this paragraph, do not constitute arguments. And it is equally difficult, note, to imagine how the relevant phenomena could

be explained or predicted solely in terms of the substance dualist's nonphysical mind-stuff. The explanatory problem here is a major challenge to everybody, not just to the materialist. On this issue then, we have a rough standoff.

The final argument in support of dualism urged the existence of parapsychological phenomena such as telepathy and telekinesis, the point being that such mental phenomena are (a) real, and (b) beyond purely physical explanation. This argument is really another instance of the argument from irreducibility discussed above, and as before, it is not entirely clear that such phenomena, even if real, must forever escape a purely physical explanation. The materialist can already suggest a possible mechanism for telepathy, for example. On his view, thinking is an electrical activity within the brain. But according to electromagnetic theory, such changing motions of electric charges must produce electromagnetic waves radiating at the speed of light in all directions, waves that will contain information about the electrical activity that produced them. Such waves can subsequently have effects on the electrical activity of other brains, that is, on their thinking. Call this the 'radio transmitter/receiver' theory of telepathy.

I do not for a moment suggest that this theory is true: the electromagnetic waves emitted by the brain are fantastically weak (billions of times weaker than the ever present background electromagnetic flux produced by commercial radio stations), and they are almost certain to be hopelessly jumbled together as well. This is one reason why, in the absence of systematic, compelling, and repeatable evidence for the existence of telepathy, one must doubt its possibility. But it is significant that the materialist has the theoretical resources to suggest a detailed possible explanation of telepathy, if it were real, which is more than any dualist has so far done. It is not at all clear, then, that the materialist *must* be at an explanatory disadvantage in these matters. Quite the reverse.

Put the preceding aside, if you wish, for the main difficulty with the argument from parapsychological phenomena is much, much simpler. Despite the endless pronouncements and anecdotes in the popular press, and despite a steady trickle of serious research on such things, there is no significant or trustworthy evidence that such phenomena even exist. The wide gap between popular conviction on this matter, and the actual evidence, is something that itself calls for research. For there is not a single parapsychological effect that can be repeatedly or reliably produced in any laboratory suitably equipped to perform and control the experiment. Not one. Honest researchers have been repeatedly hoodwinked by 'psychic' charlatans with skills derived from the magician's trade, and the history of the subject is largely a history

of gullibility, selection of evidence, poor experimental controls, and outright fraud by the occasional researcher as well. If someone really does discover a repeatable parapsychological effect, then we shall have to reevaluate the situation, but as things stand, there is nothing here to support a dualist theory of mind.

Upon critical examination, the arguments in support of dualism lose much of their force. But we are not yet done: there are arguments against dualism, and these also require examination.

Arguments against Dualism

The first argument against dualism urged by the materialists appeals to the greater *simplicity* of their view. It is a principle of rational methodology that, if all else is equal, the simpler of two competing hypotheses should be preferred. This principle is sometimes called "Ockham's Razor"—after William of Ockham, the medieval philosopher who first enunciated it—and it can also be expressed as follows: "Do not multiply entities beyond what is strictly necessary to explain the phenomena." The materialist postulates only one kind of substance (physical matter), and one class of properties (physical properties), whereas the dualist postulates two kinds of matter and/or two classes of properties. And to no explanatory advantage, charges the materialist.

This is not yet a decisive point against dualism, since neither dualism nor materialism can yet explain all of the phenomena to be explained. But the objection does have some force, especially since there is no doubt at all that physical matter exists, while spiritual matter remains a tenuous hypothesis.

If this latter hypothesis brought us some definite explanatory advantage obtainable in no other way, then we would happily violate the demand for simplicity, and we would be right to do so. But it does not, claims the materialist. In fact, the advantage is just the other way around, he argues, and this brings us to the second objection to dualism: the relative *explanatory impotence* of dualism as compared to materialism.

Consider, very briefly, the explanatory resources already available to the neurosciences. We know that the brain exists and what it is made of. We know much of its microstructure: how the neurons are organized into systems and how distinct systems are connected to one another, to the motor nerves going out to the muscles, and to the sensory nerves coming in from the sense organs. We know much of their microchemistry: how the nerve cells fire tiny electrochemical pulses along their various fibers, and how they make other cells fire also, or cease firing. We know some of how such activity processes sensory information, selecting salient or subtle bits to be sent on to higher systems. And we know some of how such activity initiates and coordinates bodily be-

havior. Thanks mainly to neurology (the branch of medicine concerned with brain pathology), we know a great deal about the correlations between damage to various parts of the human brain, and various behavioral and cognitive deficits from which the victims suffer. There are a great many isolated deficits—some gross, some subtle—that are familiar to neurologists (inability to speak, or to read, or to understand speech, or to recognize faces, or to add/subtract, or to move a certain limb, or to put information into long-term memory, and so on), and their appearance is closely tied to the occurrence of damage to very specific parts of the brain.

Nor are we limited to cataloguing traumas. The growth and development of the brain's microstructure is also something that neuroscience has explored, and such development appears to be the basis of various kinds of learning by the organism. Learning, that is, involves lasting chemical and physical changes in the brain. In sum, the neuroscientist can tell us a great deal about the brain, about its constitution and the physical laws that govern it; he can already explain much of our behavior in terms of the physical, chemical, and electrical properties of the brain; and he has the theoretical resources available to explain a good deal more as our explorations continue. (We shall take a closer look at neurophysiology and neuropsychology in chapter 7.)

Compare now what the neuroscientist can tell us about the brain, and what he can do with that knowledge, with what the dualist can tell us about spiritual substance, and what he can do with those assumptions. Can the dualist tell us anything about the internal constitution of mind-stuff? Of the nonmaterial elements that make it up? Of the laws that govern their behavior? Of the mind's structural connections with the body? Of the manner of its operations? Can he explain human capacities and pathologies in terms of its structures and its defects? The fact is, the dualist can do none of these things, because no detailed theory of mind-stuff has ever been formulated. Compared to the rich resources and explanatory successes of current materialism, dualism is less a theory of mind than it is an empty space waiting for a genuine theory of mind to be put in it.

Thus argues the materialist. But again, this is not a completely decisive point against dualism. The dualist can admit that the brain plays a major role in the administration of both perception and behavior—on his view the brain is the *mediator* between the mind and the body—but he may attempt to argue that the materialist's current successes and future explanatory prospects concern only the mediative functions of the brain, not the *central* capacities of the nonphysical mind, capacities such as reason, emotion, and consciousness itself. On these latter topics, he may argue, both dualism *and* materialism currently draw a blank.

But this reply is not a very good one. So far as the capacity for reasoning is concerned, machines already exist that execute in minutes sophisticated deductive and mathematical calculations that would take a human a lifetime to execute. And so far as the other two mental capacities are concerned, studies of such things as depression, motivation, attention, and sleep have revealed many interesting and puzzling facts about the neurochemical and neurodynamical basis of both emotion and consciousness. The *central* capacities, no less than the peripheral, have been addressed with profit by various materialist research programs.

In any case, the (substance) dualist's attempt to draw a sharp distinction between the unique 'mental' capacities proper to the nonmaterial mind, and the merely mediative capacities of the brain, prompts an argument that comes close to being an outright refutation of (substance) dualism. If there really is a distinct entity in which reasoning, emotion, and consciousness take place, and if that entity is dependent on the brain for nothing more than sensory experiences as input and volitional executions as output, *then one would expect reason, emotion, and consciousness to be relatively invulnerable to direct control or pathology by manipulation or damage to the brain.* But in fact the exact opposite is true. Alcohol, narcotics, or senile degeneration of nerve tissue will impair, cripple, or even destroy one's capacity for rational thought. Psychiatry knows of hundreds of emotion-controlling chemicals (lithium, chlorpromazine, amphetamine, cocaine, and so on) that do their work when vectored into the brain. And the vulnerability of consciousness to the anesthetics, to caffeine, and to something as simple as a sharp blow to the head, shows its very close dependence on neural activity in the brain. All of this makes perfect sense if reason, emotion, and consciousness are activities of the brain itself. But it makes very little sense if they are activities of something else entirely.

We may call this the argument from the *neural dependence* of all known mental phenomena. Property dualism, note, is not threatened by this argument, since, like materialism, property dualism reckons the brain as the seat of all mental activity. We shall conclude this section, however, with an argument that cuts against both varieties of dualism: the argument from *evolutionary history*.

What is the origin of a complex and sophisticated species such as ours? What, for that matter, is the origin of the dolphin, the mouse, or the housefly? Thanks to the fossil record, comparative anatomy, and the biochemistry of proteins and nucleic acids, there is no longer any significant doubt on this matter. Each existing species is a surviving type from a number of variations on an earlier type of organism; each earlier type is in turn a surviving type from a number of variations on

a still earlier type of organism; and so on down the branches of the evolutionary tree until, some three billion years ago, we find a trunk of just one or a handful of very simple organisms. These organisms, like their more complex offspring, are just self-repairing, self-replicating, energy-driven molecular structures. (That evolutionary trunk has its own roots in an earlier era of purely chemical evolution, in which the molecular elements of life were themselves pieced together.) The mechanism of development that has structured this tree has two main elements: (1) the occasional blind variation in types of reproducing creature, and (2) the selective survival of some of these types due to the relative reproductive advantage enjoyed by individuals of those types. Over periods of geological time, such a process can produce an enormous variety of organisms, some of them very complex indeed.

For purposes of our discussion, the important point about the standard evolutionary story is that the human species and all of its features are the wholly physical outcome of a purely physical process. Like all but the simplest of organisms, we have a nervous system. And for the same reason: a nervous system permits the discriminative guidance of behavior. But a nervous system is just an active matrix of cells, and a cell is just an active matrix of molecules. We are notable only in that our nervous system is more complex and powerful than those of our fellow creatures. Our inner nature differs from that of simpler creatures in degree, but not in kind.

If this is the correct account of our origins, then there seems neither need, nor room, to fit any nonphysical substances or properties into our theoretical account of ourselves. We are creatures of matter. And we should learn to live with that fact.

Arguments like these have moved most (but not all) of the professional community to embrace some form of materialism. This has not produced much unanimity, however, since the differences between the several materialist positions are even wider than the differences that divide dualism. The next four sections explore these more recent positions.

Suggested Readings

On Substance Dualism

Descartes, René, *The Meditations*, meditation II.

Descartes, René, *Discourse on Method*, part 5.

Eccles, Sir John C., *The Self and Its Brain*, with Sir Karl Popper (New York: Springer-Verlag, 1977).

On Property Dualism

- Popper, Sir Karl, *The Self and Its Brain*, with Sir John C. Eccles (New York: Springer-Verlag, 1977).
- Margolis, Joseph, *Persons and Minds: The Prospects of Nonreductive Materialism* (Dordrecht-Holland: Reidel, 1978).
- Jackson, Frank, "Epiphenomenal Qualia," *The Philosophical Quarterly*, vol. 32, no. 127 (April, 1982).
- Nagel, Thomas, "What Is It Like to Be a Bat?" *Philosophical Review*, vol. LXXXIII (1974). Reprinted in *Readings in Philosophy of Psychology*, vol. I, ed. N. Block (Cambridge, MA: Harvard University Press, 1980).

2. Philosophical Behaviorism

Philosophical behaviorism reached the peak of its influence during the first and second decades after World War II. It was jointly motivated by at least three intellectual fashions. The first motivation was a reaction against dualism. The second motivation was the Logical Positivists' idea that the meaning of any sentence was ultimately a matter of the observable circumstances that would tend to verify or confirm that sentence. And the third motivation was a general assumption that most, if not all, philosophical problems are the result of linguistic or conceptual confusion, and are to be solved (or dissolved) by careful analysis of the language in which the problem is expressed.

In fact, philosophical behaviorism is not so much a theory about what mental states are (in their inner nature) as it is a theory about how to analyze or to understand the vocabulary we use to talk about them. Specifically, the claim is that talk about emotions and sensations and beliefs and desires is not talk about ghostly inner episodes, but is rather a shorthand way of talking about actual and potential patterns of *behavior*. In its strongest and most straightforward form, philosophical behaviorism claims that any sentence about a mental state can be paraphrased, without loss of meaning, into a long and complex sentence about what observable behavior *would* result if the person in question were in this, that, or the other observable circumstance.

A helpful analogy here is the dispositional property, *being soluble*. To say that a sugar cube is soluble is not to say that the sugar cube enjoys some ghostly inner state. It is just to say that *if* the sugar cube were put in water, then it *would* dissolve. More strictly,

"x is water soluble"

is equivalent by definition to

"if x were put in unsaturated water, x would dissolve."

This is one example of what is called an "operational definition". The term "soluble" is defined in terms of certain operations or tests that would reveal whether or not the term actually applies in the case to be tested.

According to the behaviorist, a similar analysis holds for mental states such as "wants a Caribbean holiday", save that the analysis is much richer. To say that Anne wants a Caribbean holiday is to say that (1) if asked whether that is what she wants, she would answer yes, and (2) if given new holiday brochures for Jamaica and Japan, she would peruse the ones for Jamaica first, and (3) if given a ticket on this Friday's flight to Jamaica, she would go, and so on and so on.

Unlike solubility, claims the behaviorist, most mental states are *multitracked* dispositions. But dispositions they remain.

There is therefore no point in worrying about the 'relation' between the mind and the body, on this view. To talk about Marie Curie's mind, for example, is not to talk about some 'thing' that she 'possesses'; it is to talk about certain of her extraordinary capacities and dispositions. The mind-body problem, concludes the behaviorist, is a pseudoproblem.

Behaviorism is clearly consistent with a materialist conception of human beings. Material objects can have dispositional properties, even multitracked ones, so there is no necessity to embrace dualism to make sense of our psychological vocabulary. (It should be pointed out, however, that behaviorism is strictly consistent with dualism also. Even if philosophical behaviorism were true, it would remain possible that our multitracked dispositions are grounded in immaterial mind-stuff rather than in molecular structures. This is not a possibility that most behaviorists took seriously, however, for the many reasons outlined at the end of the preceding section.)

Philosophical behaviorism, unfortunately, had two major flaws that made it awkward to believe, even for its defenders. It evidently ignored, and even denied, the 'inner' aspect of our mental states. To have a pain, for example, seems to be not merely a matter of being inclined to moan, to wince, to take aspirin, and so on. Pains also have an intrinsic qualitative nature (a horrible one) that is revealed in introspection, and any theory of mind that ignores or denies such *qualia* is simply derelict in its duty.

This problem received much attention from behaviorists, and serious attempts were made to solve it. The details take us deeply into semantical problems, however, so we shall postpone further discussion of this difficulty until chapter 3.

The second flaw emerged when behaviorists attempted to specify in detail the multitracked disposition said to constitute any given mental state. The list of conditionals necessary for an adequate analysis of "wants a Caribbean holiday", for example, seemed not just to be long, but to be indefinitely or even infinitely long, with no finite way of specifying the elements to be included. And no term can be well-defined whose *definiens* is open-ended and unspecific in this way. Further, each conditional of the long analysis was suspect on its own. Supposing that Anne does want a Caribbean holiday, conditional (1) above will be true only if she isn't *secretive* about her holiday fantasies; conditional (2) will be true only if she isn't already *bored* with the Jamaica brochures; conditional (3) will be true only if she doesn't *believe* the Friday flight will be hijacked, and so forth. But to repair each conditional by adding in the relevant qualification would be to rein-

roduce a series of *mental* elements into the business end of the definition, and we would no longer be defining the mental solely in terms of publicly observable circumstances and behavior.

So long as behaviorism seemed the only alternative to dualism, philosophers were prepared to struggle with these flaws in hopes of repairing or defusing them. However, three more materialist theories rose to prominence during the late fifties and sixties, and the flight from behaviorism was swift.

(I close this section with a cautionary note. The *philosophical* behaviorism discussed above is to be sharply distinguished from the *methodological* behaviorism that has enjoyed such a wide influence within psychology. In its bluntest form, this latter view urges that any new theoretical terms invented by the science of psychology *should be* operationally defined, in order to guarantee that psychology maintains a firm contact with empirical reality. Philosophical behaviorism, by contrast, claims that all of the common-sense psychological terms in our prescientific vocabulary *already* get whatever meaning they have from (tacit) operational definitions. The two views are logically distinct, and the methodology might be a wise one, for new theoretical terms, even though the correlative analysis of common-sense mental terms is wrong.)

Suggested Readings

- Ryle, Gilbert, *The Concept of Mind* (London: Hutchinson & Company, 1949), chapters I and V.
 Malcolm, Norman, "Wittgenstein's *Philosophical Investigations*," *Philosophical Review*, vol. XLVII (1956). Reprinted in *The Philosophy of Mind*, ed. V. C. Chappell (Englewood Cliffs, NJ: Prentice-Hall, 1962).