

no doubt, even in some human cases we might be puzzled as to whether we should take the ascription of intentionality literally or metaphorically. But marginal cases do not alter the distinction between the sort of facts corresponding to ascriptions of intrinsic intentionality and those corresponding to *as-if* metaphorical ascriptions of intentionality. There is nothing harmful, misleading, or philosophically mistaken about *as-if* metaphorical ascriptions. The only mistake is to take them literally.

I hope the distinctions I have been making are painfully obvious. However, I have to report, from the battlefronts as it were, that the neglect of these simple distinctions underlies some of the biggest mistakes in contemporary intellectual life. A common pattern of mistake is to suppose that because we can make *as-if* ascriptions of intentionality to systems that have no intrinsic intentionality, that somehow or other we have discovered the nature of intentionality.<sup>5</sup>

JOHN SEARLE - REDISCOVERY OF  
THE MIND

SL

## Chapter 4

### Consciousness and Its Place in Nature

#### I. Consciousness and the "Scientific" World View

As with most words, it is not possible to give a definition of "consciousness" in terms of necessary and sufficient conditions, nor is it possible to define it in the Aristotelian fashion by way of genus and differentia. However, though we cannot give a noncircular verbal definition, it is still essential for me to say what I mean by this notion, because it is often confused with several others. For example, for reasons of both etymology and usage, "consciousness" is often confused with "conscience," "self-consciousness," and "cognition."

What I mean by "consciousness" can best be illustrated by examples. When I wake up from a dreamless sleep, I enter a state of consciousness, a state that continues as long as I am awake. When I go to sleep or am put under a general anesthetic or die, my conscious states cease. If during sleep I have dreams, I become conscious, though dream forms of consciousness in general are of a much lower level of intensity and vividness than ordinary waking consciousness. Consciousness can vary in degree even during our waking hours, as for example when we move from being wide awake and alert to sleepy or drowsy, or simply bored and inattentive. Some people introduce chemical substances into their brains for the purpose of producing altered states of consciousness, but even without chemical assistance, it is possible in ordinary life to distinguish different degrees and forms of consciousness. Consciousness is an on/off switch: a system is either conscious or not. But once conscious, the system is a rheostat: there are different degrees of consciousness.

A near synonym for "consciousness," in my sense, is "~~awareness~~," but I do not think they are exactly equivalent in meaning because "awareness" is more closely connected to cognition, to knowledge, than is the general notion of consciousness. Furthermore, it seems possible that one might allow for cases in which one is aware of something unconsciously (cf. Weiskrantz et al. 1974). It is also worth emphasizing that there is nothing so far in my account of consciousness that implies self-consciousness. I will later (in chapter 6) discuss the connection between consciousness and self-consciousness.

Some philosophers (e.g., Block, "Two Concepts of Consciousness,") claim that there is a sense of this word that implies no sentience whatever, a sense in which a total zombie could be "conscious." I know of no such sense, but in any case that is not the sense in which I am using the word.

Conscious states always have a content. One can never just be conscious, rather when one is conscious, there must be an answer to the question, "What is one conscious of?" But the "of" of "conscious of" is not always the "of" of intentionality. If I am conscious of a knock on the door, my conscious state is intentional, because it makes reference to something beyond itself, the knock on the door. If I am conscious of a pain, the pain is not intentional, because it does not represent anything beyond itself.<sup>1</sup>

The main aim of this chapter is to locate consciousness within our overall "scientific" conception of the world. The reason for emphasizing consciousness in an account of the mind is that it is the central mental notion. In one way or another, all other mental notions—such as intentionality, subjectivity, mental causation, intelligence, etc.—can only be fully understood as mental by way of their relations to consciousness (more about this in chapter 7). Because at any given point in our waking lives only a tiny fraction of our mental states is conscious, it may seem paradoxical to think of consciousness as the central mental notion, but I intend in the course of this book to try to resolve the appearance of paradox. Once we have located the place of consciousness in our overall world

view, we can see that the materialist theories of the mind we discussed in chapter 2 are just as profoundly antiscientific as the dualism they thought they were attacking.

We will find that when we try to state the facts, the pressure on the traditional categories and terminology becomes almost unbearable and they begin to crack under the strain. What I say will sound almost self-contradictory: On the one hand I will claim that consciousness is just an ordinary biological feature of the world, but I will also try to show why we find it almost literally inconceivable that it should be so.

Our contemporary world view began to develop in the seventeenth century, and its development is continuing right through the late twentieth century. Historically, one of the keys to this development was the exclusion of consciousness from the subject matter of science by Descartes, Galileo, and others in the seventeenth century. On the Cartesian view, the natural sciences proper excluded "mind," *res cogitans*, and concerned themselves only with "matter," *res extensa*. The separation between mind and matter was a useful heuristic tool in the seventeenth century, a tool that facilitated a great deal of the progress that took place in the sciences. However, the separation is philosophically confused, and by the twentieth century it had become a massive obstacle to a scientific understanding of the place of consciousness within the natural world. One of the main aims of this book is to try to remove that obstacle, to bring consciousness back into the subject matter of science as a biological phenomenon like any other. To do that, we need to answer the dualistic objections of contemporary Cartesians.

It goes without saying that our "scientific" world view is extremely complex and includes all of our generally accepted theories about what sort of place the universe is and how it works. It includes, that is, theories ranging from quantum mechanics and relativity theory to the plate tectonic theory of geology and the DNA theory of hereditary transmission. At present, for example, it includes a belief in black holes, the germ theory of disease, and the heliocentric account of the

solar system. Some features of this world view are very tentative, others well established. At least two features of it are so fundamental and so well established as to be no longer optional for reasonably well-educated citizens of the present era; indeed they are in large part constitutive of the modern world view. These are the atomic theory of matter and the evolutionary theory of biology. Of course, like any other theory, they might be refuted by further investigation; but at present the evidence is so overwhelming that they are not simply up for grabs. To situate consciousness within our understanding of the world, we have to situate it with respect to these two theories.

According to the atomic theory of matter, the universe consists entirely of extremely small physical phenomena that we find it convenient, though not entirely accurate, to call "particles." All the big and middle-sized entities in the world, such as planets, galaxies, cars, and overcoats, are made up of smaller entities that are in turn made up of yet smaller entities until finally we reach the level of molecules, themselves composed of atoms, themselves composed of subatomic particles. Examples of particles are electrons, hydrogen atoms, and water molecules. As these examples illustrate, bigger particles are made up of smaller particles; and there is still much uncertainty and dispute about the identification of the ultimately smallest particles. We are somewhat embarrassed to use the word "particle" for at least two reasons. First, it seems more accurate to describe the more basic of these entities as points of mass/energy rather than as extended spatial entities. And second, more radically, according to quantum mechanics, as long as they are not being measured or interfered with in some way, "particles," such as electrons, behave more like waves than like particles. However, for convenience I will stick with the word "particle."

Particles, as our earlier examples illustrated, are organized into larger systems. It would be tricky to try to define the notion of a system, but the simple intuitive idea is that systems are collections of particles where the spatio-temporal boun-

daries of the system are set by causal relations. Thus, a raindrop is a system, but so is a glacier. Babies, elephants, and mountain ranges are also examples of systems. It should be obvious from these examples that systems can contain subsystems.

Essential to the explanatory apparatus of atomic theory is not only the idea that big systems are made up of little systems, but that many features of the big ones can be causally explained by the behavior of the little ones. This conception of explanation gives us the possibility, indeed the requirement, that many sorts of macrophenomena be explicable in terms of microphenomena. And this in turn has the consequence that there will be different levels of explanation of the same phenomenon, depending on whether we are going left to right from macro to macro, or micro to micro, or bottom up from micro to macro. We can illustrate these levels with a simple example. Suppose I wish to explain why this pot of water is boiling. One explanation, a left-right macro-macro explanation, would be that I put the pot on the stove and turned on the heat under it. I call this explanation "left-right" because it cites an earlier event to explain a later event,<sup>2</sup> and I call it "macro-macro" because both explanans and explanandum are at the macrolevel. Another explanation—bottom-up micro-macro—would be that the water is boiling because the kinetic energy transmitted by the oxidization of hydrocarbons to the H<sub>2</sub>O molecules has caused them to move so rapidly that the internal pressure of the molecule movements equals the external air pressure, which pressure in turn is explained by the movement of the molecules of which the external air is composed. I call this explanation "bottom-up micro-macro" because it explains the features and behavior of surface or macrophenomena in terms of lower-level microphenomena. I do not mean to imply that these are the only possible levels of explanation. There are also left-right micro-micro explanations, and further subdivisions can be made within each micro or macro level.

This, then, is one of the chief lessons of atomic theory: many features of big things are explained by the behavior of little

things. We regard the germ theory of disease or the DNA theory of genetic transmission as such major breakthroughs precisely because they fit this model. If someone had an explanation of diseases in terms of the movement of the planets we would never accept it as complete explanation, even if it worked for diagnoses and cures, until we understood how the macro causes and effects at the level of planets and symptoms were grounded in bottom-up micro-macro causal structures.

To these elementary notions of atomic theory let us now add the principles of evolutionary biology. Over long periods of time, certain *types* of living systems evolve in certain very special ways. On our little earth, the types of systems in question invariably contain carbon-based molecules, and they make extensive use of hydrogen, nitrogen, and oxygen. The ways in which they evolve are complicated, but the basic procedure is that token instances of the types cause similar tokens to come into existence. Thus, after the original tokens are destroyed, the type or pattern that they exemplify continues in other tokens and continues to be replicated as subsequent generations of tokens produce yet other tokens. Variations in the surface features, phenotypes, of the tokens give those tokens greater or lesser chances of survival, relative to the specific environments in which they find themselves. Those tokens that have a greater probability of survival relative to their environment will therefore have a greater probability of producing further tokens like themselves, tokens with the same genotype. And thus does the type evolve.

Part of the intellectual appeal of the theory of evolution, as supplemented by Mendelian and DNA genetics, is that it fits in with the explanatory model we have derived from atomic theory. Specifically, the grounding of genetic mechanisms in molecular biology allows for different levels of explanation of biological phenomena corresponding to the different levels of explanation we have for physical phenomena. In evolutionary biology, there are characteristically two levels of explanation, a "functional" level where we explain the survival of species in

terms of "inclusive fitness," which depends on the phenotypical traits possessed by members of the species, and a "causal" level where we explain the causal mechanisms by which the traits in question actually relate the organism to the environment. We can illustrate this with a simple example. Why do green plants turn their leaves toward the sun? The functional explanation:<sup>3</sup> This trait has survival value. By increasing the plant's capacity to perform photosynthesis, it increases the plant's capacity to survive and reproduce. The plant does not turn toward the sun to survive; rather, the plant tends to survive because it is predisposed to turn toward the sun anyway. The causal explanation: The plant's biochemical structure as determined by its genetic makeup causes it to secrete the growth hormone auxin, and the varying concentrations of auxin in turn cause the leaves to turn in the direction of the light source.

If you put these two levels of explanation together, you get the following result: Because the phenotype, as produced by the interaction of the genotype with the environment, has survival value relative to the environment, the genotype survives and reproduces. Such, in very brief form, are the mechanisms of natural selection.

The products of the evolutionary process, organisms, are made of subsystems called "cells," and some of these organisms develop subsystems of nerve cells, which we think of as "nervous systems." Furthermore, and this is the crucial point, some extremely complex nervous systems are capable of causing and sustaining conscious states and processes. Specifically, certain big collections of nerve cells, that is, brains, cause and sustain conscious states and processes. We do not know the detail of how brains cause consciousness, but we know for a fact that this occurs in human brains, and we have overwhelming evidence that it also occurs in the brains of many species of animals (Griffin 1981). We do not know at present how far down the evolutionary scale consciousness extends.

Basic to our world view is the idea that human beings and other higher animals are part of the biological order like any

other organisms. Humans are continuous with the rest of nature. But if so, the biologically specific characteristics of these animals—such as their possession of a rich system of consciousness, as well as their greater intelligence, their capacity for language, their capacity for extremely fine perceptual discriminations, their capacity for rational thought, etc.—are biological phenomena like any other biological phenomena. Furthermore, these features are all phenotypes. They are as much the result of biological evolution as any other phenotype.

*Consciousness, in short, is a biological feature of human and certain animal brains. It is caused by neurobiological processes and is as much a part of the natural biological order as any other biological features such as photosynthesis, digestion, or mitosis.* This principle is the first stage in understanding the place of consciousness within our world view.<sup>4</sup> The thesis of this chapter so far has been that once you see that atomic and evolutionary theories are central to the contemporary scientific world view, then consciousness falls into place naturally as an evolved phenotypical trait of certain types of organisms with highly developed nervous systems. I am not in this chapter concerned to defend this world view. Indeed, many thinkers whose opinions I respect, most notably Wittgenstein, regard it as in varying degrees repulsive, degrading, and disgusting. It seems to them to allow no place—or at most a subsidiary place—for religion, art, mysticism, and “spiritual” values generally. But, like it or not, it is the world view we have. Given what we know about the details of the world — about such things as the position of elements in the periodic table, the number of chromosomes in the cells of different species, and the nature of the chemical bond—this world view is not an option. It is not simply up for grabs along with a lot of competing world views. Our problem is not that somehow we have failed to come up with a convincing proof of the existence of God or that the hypothesis of an afterlife remains in serious doubt, it is rather that in our deepest reflections we cannot take such opinions seriously. When we encounter people who claim to believe such things, we may envy them the comfort

and security they claim to derive from these beliefs, but at bottom we remain convinced that either they have not heard the news or they are in the grip of faith. We remain convinced that somehow they must separate their minds into separate compartments to believe such things. When I lectured on the mind-body problem in India and was assured by several members of my audience that my views must be mistaken, because they personally had existed in their earlier lives as frogs or elephants, etc., I did not think, “Here is evidence for an alternative world view,” or even “Who knows, perhaps they are right.” And my insensitivity was much more than mere cultural provincialism: Given what I know about how the world works, I could not regard their views as serious candidates for truth.

And once you accept our world view the only obstacle to granting consciousness its status as a biological feature of organisms is the outmoded dualistic/materialistic assumption that the “mental” character of consciousness makes it impossible for it to be a “physical” property.

I have only discussed the relation of consciousness to carbon-based living systems of the sort we have on our earth, but of course, we cannot exclude the possibility that consciousness may have evolved on other planets in other solar systems in other parts of the universe. Given the sheer size of the universe, it would be statistically amazing if we were the only bearers of consciousness in it. Furthermore, we do not want to exclude the possibility that consciousness might have been evolved in systems that are not carbon-based, but use some other sort of chemistry altogether. For all we know at present, there might be no theoretical obstacle to developing consciousness in systems made up of other elements. We are at present very far from having an adequate theory of the neurophysiology of consciousness; but until we do, we have to keep an open mind about its possible chemical bases. My own hunch would be that the neurobiology of consciousness is likely to prove at least as restricted as, say, the biochemistry of digestion. There are different varieties of digestion, but not any-

thing can be digested by anything. And similarly, it seems to me we are likely to find that though there may be biochemically different varieties of consciousness, not anything goes.

Furthermore, because consciousness is entirely caused by the behavior of lower-level biological phenomena, it would in principle be possible to produce it artificially by duplicating the causal powers of the brain in a laboratory situation. We know that many biological phenomena have been created artificially. We can synthesize certain organic compounds, and even artificially create certain biological processes such as photosynthesis. If we can artificially create photosynthesis, why not consciousness as well? For photosynthesis, the artificial form of the phenomenon was created by actually duplicating the chemical processes in the laboratory. Similarly, if one were going to create consciousness artificially, the natural way to go about it would be to try to duplicate the actual neurobiological basis that consciousness has in organisms like ourselves. Because at present we do not know exactly what that neurobiological basis is, the prospects for such "artificial intelligence" are very remote. Furthermore, as I suggested earlier, it might be possible to produce consciousness using some altogether different sort of chemistry from the one that our brains in fact use. However, one thing we know before we even begin the investigation is that any system capable of causing consciousness must be capable of duplicating the causal powers of the brain. If, for example, it is done with silicon chips instead of neurons, it must be because the chemistry of the silicon chips is capable of duplicating the specific causal powers of neurons to cause consciousness. It is a trivial logical consequence of the fact that brains cause consciousness that any other system capable of causing consciousness, but using completely different mechanisms, would have to have at least the equivalent power of brains to do it. (Compare: airplanes don't have to have feathers to fly, but they do have to share with birds the causal capacity to overcome the force of gravity in the earth's atmosphere.)

To summarize: Our world picture, though extremely complicated in detail, provides a rather simple account of the mode of

existence of consciousness. According to the atomic theory, the world is made up of particles. These particles are organized into systems. Some of these systems are living, and these types of living systems have evolved over long periods of time. Among these, some have evolved brains that are capable of causing and sustaining consciousness. Consciousness is, thus, a biological feature of certain organisms in exactly the same sense of "biological" in which photosynthesis, mitosis, digestion, and reproduction are biological features of organisms.

I have tried to describe the position of consciousness in our overall world view in very simple terms, because I want it to seem absolutely obvious. Anyone who has had even a modicum of "scientific" education after about 1920 should find nothing at all contentious or controversial in what I have just said. It is worth emphasizing also that all of this has been said without any of the traditional Cartesian categories. There has been no question of dualism, monism, materialism, or anything of the sort. Furthermore, there has been no question of "naturalizing consciousness"; it already is completely natural. Consciousness, to repeat, is a natural biological phenomenon. The exclusion of consciousness from the natural world was a useful heuristic device in the seventeenth century, because it enabled scientists to concentrate on phenomena that were measurable, objective, and meaningless, that is, free of intentionality. But the exclusion was based on a falsehood. It was based on the false belief that consciousness is not part of the natural world. That single falsehood, more than anything else, more even than the sheer difficulty of studying consciousness with our available scientific tools, has prevented us from arriving at an understanding of consciousness.

## II. Subjectivity

Conscious mental states and processes have a special feature not possessed by other natural phenomena, namely, subjectivity. It is this feature of consciousness that makes its study so recalcitrant to the conventional methods of biological and psychological research, and most puzzling to philosophical

analysis. There are several different senses of "subjectivity," none of them entirely clear, and I need to say at least a little bit to clarify the sense in which I am claiming that consciousness is subjective.

We often speak of judgments as being "~~subjective~~" when we mean that their truth or falsity cannot be settled "objectively" because the truth or falsity is not a simple matter of fact, but depends on certain attitudes, feelings, and points of view of the makers and the hearers of the judgment. An example of such a judgment might be, "Van Gogh is a better artist than Matisse." In this sense of "subjectivity," we contrast such subjective judgments with completely objective judgments, such as the judgment, "Matisse lived in Nice during the year 1917." For such objective judgments, we can ascertain what sorts of facts in the world make them true or false independent of anybody's attitudes or feelings about them.

Now this sense in which we speak of "objective" and "subjective" judgments is not the sense of "subjective" in which I am speaking of consciousness as subjective. In the sense in which I am here using the term, "subjective" refers to an ontological category, not to an epistemic mode. Consider, for example, the statement, "I now have a pain in my lower back." That statement is completely objective in the sense that it is made true by the existence of an actual fact and is not dependent on any stance, attitudes, or opinions of observers. However, the phenomenon itself, the actual pain itself, has a subjective mode of existence, and it is in that sense which I am saying that consciousness is subjective.

What more can we say about this subjective mode of existence? Well, first it is essential to see that in consequence of its subjectivity, the pain is not equally accessible to any observer. Its existence, we might say, is a first-person existence. For it to be a pain, it must be somebody's pain; and this in a much stronger sense than the sense in which a leg must be somebody's leg, for example. Leg transplants are possible; in that sense, pain transplants are not. And what is true of pains is true of conscious states generally. Every conscious

state is always someone's conscious state. And just as I have a special relation to my conscious states, which is not like my relation to other people's conscious states, so they in turn have a relation to their conscious states, which is not like my relation to their conscious states.<sup>5</sup> Subjectivity has the further consequence that all of my conscious forms of intentionality that give me information about the world independent of myself are always from a special point of view. The world itself has no point of view, but my access to the world through my conscious states is always perspectival, always from my point of view.

It would be difficult to exaggerate the disastrous effects that the failure to come to terms with the subjectivity of consciousness has had on the philosophical and psychological work of the past half century. In ways that are not at all obvious on the surface, much of the bankruptcy of most work in the philosophy of mind and a great deal of the sterility of academic psychology over the past fifty years, over the whole of my intellectual lifetime, have come from a persistent failure to recognize and come to terms with the fact that the ontology of the mental is an irreducibly first-person ontology. There are very deep reasons, many of them embedded in our unconscious history, why we find it difficult if not impossible to accept the idea that the real world, the world described by physics and chemistry and biology, contains an ineliminably subjective element. How could such a thing be? How can we possibly get a coherent world picture if the world contains these mysterious conscious entities? Yet we all know that we are for most of our lives conscious, and that other people around us are conscious. And unless we are blinded by bad philosophy or some forms of academic psychology, we really don't have any doubts that dogs, cats, monkeys, and small children are conscious, and that their consciousness is just as subjective as our own.

So let us try to describe in a little more detail the world picture that contains subjectivity as a rock-bottom element, and then try to describe some of the difficulties we have in coming

to terms with this world picture. If we think of the world as consisting of particles, and those particles as organized into systems, and some of those systems as biological systems, and some of those biological systems as conscious, and consciousness as essentially subjective—then what is it that we are being asked to imagine when we imagine the subjectivity of consciousness? After all, all those other things we imagined—particles, systems, organisms, etc.—were completely objective. In consequence, they are equally accessible to all competent observers. So what are we being asked to imagine if we are now to throw into this metaphysical pot something that is irreducibly subjective?

Actually, what we are being asked to “imagine” is simply the world that we know to exist. I know, for example, that I am now conscious, and that this conscious state that I am in has the subjectivity I have been referring to, and I know that a very large number of other organisms like myself are similarly conscious and have similar subjective states. Then why does it seem that I am asking us to imagine something that is difficult or in some way counterintuitive, when all I am doing is reminding us of facts that are right in front of our face all along? Part—but only part—of the answer has to do with the fact that quite naively I invoked the word “observer” in the previous paragraph. When we are asked to form a world *view* or a world *picture*, we form these on the model of vision. We tend literally to form an image of reality as consisting of very small bits of matter, “the particles,” and then we imagine these organized into systems, again with gross visible features. But when we visualize the world with this inner eye, we can’t see consciousness. Indeed, it is the very subjectivity of consciousness that makes it invisible in the crucial way. *If we try to draw a picture of someone else’s consciousness, we just end up drawing the other person* (perhaps with a balloon growing out of his or her head). *If we try to draw our own consciousness, we end up drawing whatever it is that we are conscious of.* If consciousness is the rock-bottom epistemic basis for getting at reality, we cannot get at the reality of consciousness in that way. (Alternative formulation: We cannot get at the reality of consciousness in the

way that, using consciousness, we can get at the reality of other phenomena.)

It is important to go over this rather slowly and not just zip past it in the usual fashion, so let me go through it step by step in low gear. If I try to observe the consciousness of another, what I observe is not his subjectivity but simply his conscious behavior, his structure, and the causal relations between structure and behavior. Furthermore, I observe the causal relations between both structure and behavior, on the one hand, and the environment that impinges on him and on which he in turn impinges, on the other. So there is no way I can observe someone else’s consciousness as such; rather what I observe is him and his behavior and the relations between him, the behavior, the structure, and the environment. Well, what about my own inner goings-on? Can I not observe those? The very fact of subjectivity, which we were trying to observe, makes such an observation impossible. Why? Because where conscious subjectivity is concerned, there is no distinction between the observation and the thing observed, between the perception and the

*no real distinction place*

object perceived. The model of vision works on the presupposition that there is a distinction between the thing seen and the seeing of it. But for “introspection,” there is simply no way to make this separation. Any introspection I have of my own conscious state is itself that conscious state. This is not to say that my conscious mental phenomena don’t come in many different levels and varieties—we will have occasion to examine some of these in detail later—it is simply to say that the standard model of observation simply doesn’t work for conscious subjectivity. It doesn’t work for other people’s consciousness, and it doesn’t work for one’s own. For that reason, the idea that there might be a special method of investigating consciousness, namely “introspection,” which is supposed to be a kind of inner observation, was doomed to failure from the start, and it is not surprising that introspective psychology proved bankrupt.

We find it difficult to come to terms with subjectivity, not just because we have been brought up in an ideology that says that ultimately reality must be completely objective, but

because our idea of an objectively observable reality presupposes the notion of observation that is itself ineliminably subjective, and that cannot itself be made the object of observation in a way that objectively existing objects and states of affairs in the world can. There is, in short, no way for us to picture subjectivity as part of our world view because, so to speak, the subjectivity in question is the picturing. The solution is not to try to develop a special mode of picturing, a kind of super-introspection, but rather to stop picturing altogether at this point and just acknowledge the facts. The facts are that biological processes produce conscious mental phenomena, and these are irreducibly subjective.

Philosophers have invented another metaphor for describing certain features of subjectivity that seems to me even more confused than the commonsense metaphor of introspection, and that is "privileged access." For the *visual* metaphor of introspection, we are tempted to substitute the *spatial* metaphor of privileged access, a model that suggests that consciousness is like a private room into which only we are allowed to enter. Only I can go inside the space of my own consciousness. But this metaphor doesn't work either, because for there to be something to which I have privileged access, I would have to be different from the space in which I enter. But just as the metaphor of introspection broke down when the only thing to be observed was the observing itself, so the metaphor of a private inner space breaks down when we understand that there isn't anything like a space into which I can enter, because I cannot make the necessary distinctions between the three elements of myself, the act of entering, and the space in which I am supposed to enter.

We might summarize these points by saying that our modern model of reality and of the relation between reality and observation cannot accommodate the phenomenon of subjectivity. The model is one of objective (in the epistemic sense) observers observing an objectively (in the ontological sense) existing reality. But there is no way on that model to observe the act of observing itself. For the act of observing is the sub-

jective (ontological sense) access to objective reality. Though I can easily observe another person, I cannot observe his or her *subjectivity*. And worse yet, I cannot *observe* my own subjectivity, for any observation that I might care to make is itself that which was supposed to be observed. The whole idea of there being an observation of reality is precisely the idea of (ontologically) subjective representations of reality. The ontology of observation—as opposed to its epistemology—is precisely the ontology of subjectivity. Observation is always someone's observation; it is in general conscious; it is always from a point of view; it has a subjective feel to it; etc.

I want to make clear exactly what I am saying and what I am not saying. I am not making the old muddled point to the effect that there is a self-referential paradox involved in studying subjectivity. Such paradoxes do not worry me at all. We can use the eye to study the eye, the brain to study the brain, consciousness to study consciousness, language to study language, observation to study observation, and subjectivity to study subjectivity. There is no problem in any of these. The point is rather that because of the ontology of subjectivity, our models of "studying," models that rely on the distinction between observation and thing observed, do not work for subjectivity itself.

There is a sense, then, in which we find subjectivity difficult to conceive. Given our concept as to what reality must be like and what it would be like to find out about that reality, it seems inconceivable to us that there should be anything irreducibly subjective in the universe. Yet we all know that subjectivity exists.

I hope we can now see a little more clearly what happens if we try to describe the universe leaving out subjectivity. Suppose we insist on giving an account of the world that is completely objective, not just in the epistemic sense that its claims are independently checkable, but in the ontological sense that the phenomena it describes have an existence independent of any form of subjectivity. Once you adopt this strategy (the principal strategy in the philosophy of mind of the past fifty

years), it then becomes impossible to describe consciousness, because it becomes literally impossible to acknowledge the subjectivity of consciousness. Examples of this are really too numerous to mention, but I will cite two authors who explicitly address the problem of consciousness. Armstrong (1980) tacitly eliminates subjectivity by treating consciousness simply as a capacity for making discriminations about one's own inner states, and Changeux, the French neurobiologist, defines consciousness simply as a "global regulatory system dealing with mental objects and computations using these objects" (1985, p. 145). Both of these accounts presuppose a third-person conception of reality, a conception of a reality that is not merely epistemically objective but ontologically objective as well; and such a reality has no place for consciousness, because it has no place for ontological subjectivity.

### III. Consciousness and the Mind-Body Problem

I have said repeatedly that I think the mind-body problem has a rather simple solution, at least in broad outline, and that the only obstacles to our having a full understanding of mind-body relations are our philosophical prejudice in supposing that the mental and the physical are two distinct realms, and our ignorance of the workings of the brain. If we had an adequate science of the brain, an account of the brain that would give causal explanations of consciousness in all its forms and varieties, and if we overcame our conceptual mistakes, no mind-body problem would remain. However, the possibility of any solution to the mind-body problem has been very powerfully challenged over the years by the writings of Thomas Nagel (1974, 1986). He argues as follows: At present, we simply do not have the conceptual apparatus even to conceive of a solution to the mind-body problem. This is for the following reason: Causal explanations in the natural sciences have a kind of causal necessity. We understand, for example, how the behavior of H<sub>2</sub>O molecules causes water to be in a liquid form, because we see that the liquidity is a necessary consequence of

the molecular behavior. The molecular theory does more than show that systems of H<sub>2</sub>O molecules will be liquid under certain conditions; rather it shows why the system *has to be* in a liquid form. Given that we understand the physics in question, it is inconceivable that the molecules should behave in that fashion and the water not be in a liquid state. In short, Nagel argues that explanations in science imply necessity, and necessity implies inconceivability of the opposite.

Now, says Nagel, we cannot achieve this type of necessity for the relationship between matter and consciousness. No possible account of neuronal behavior would explain why, given that behavior, we *have to be*, for example, in pain. No account could explain why pain was a necessary consequence of certain sorts of neuron firings. The proof that the account does not give us causal necessity is that we can always conceive the opposite. We can always conceive of a state of affairs in which the neurophysiology behaves in whatever way you like to specify, but all the same, the system is not in pain. If adequate scientific explanation implies necessity and necessity implies inconceivability of the opposite, then by contraposition the conceivability of the opposite implies that we do not have necessity, and that in turn implies that we do not have an explanation. Nagel's despairing conclusion is that we would need a major overhaul of our conceptual apparatus if we were ever to be able to solve the mind-body problem.

I am not convinced by this argument. First we should note that not all explanations in science have the kind of necessity that we found in the relation between molecule movement and liquidity. For example the inverse square law is an account of gravity, but it does not show why bodies *have to have* gravitational attraction. Secondly, the apparent "necessity" of any scientific explanation may be just a function of the fact that we find the explanation so convincing that we cannot, for example, conceive of the molecules moving in a particular way and the H<sub>2</sub>O not being liquid. A person in antiquity or the Middle Ages might not have found the explanation a matter of "necessity." The "mystery" of consciousness today is in roughly the

same shape that the mystery of life was before the development of molecular biology or the mystery of electromagnetism was before Clerk-Maxwell's equations. It seems mysterious because we do not know how the system of neurophysiology/consciousness works, and an adequate knowledge of how it works would remove the mystery. Furthermore, the claim that we could always conceive of the possibility that certain brain states *may not* cause the appropriate conscious states might simply depend on our ignorance of how the brain works. Given a full understanding of the brain, it seems to me likely that we would think it obvious that if the brain was in a certain sort of state, it had to be conscious. Notice that we already accept this form of causal necessity of conscious states for gross molar phenomena. For example, if I see a screaming man with his foot caught in a punch press, then I know the man must be in terrible pain. It is, in a sense, inconceivable to me that a normal human being should be in such a situation and not feel a terrible pain. The physical causes necessitate the pain.

However, let us grant Nagel's point for the sake of argument. Nothing follows about how the world works in fact. The limitation that Nagel points out is only a limitation of our powers of conception. Even assuming he is right, what his argument shows is only that in the case of the relations between material and material phenomena, we can subjectively picture both sides of the relation, but in the case of the relations between material and mental phenomena, one side of the relation is already subjective, and hence we cannot picture its relation to the material in the way that we can picture the relations between liquidity and molecule movement, for example.

Nagel's argument, in short, only shows that we cannot get out of the subjectivity of our consciousness to see its necessary relation to its material basis. We form a picture of necessity based on our subjectivity, but we cannot in that way form a picture of the necessity of the relation between subjectivity and neurophysiological phenomena, because we are already in the subjectivity, and the picturing relation would require that we

get outside it. (If solidity were conscious, it would seem to it mysterious that it was caused by vibratory movements of molecules in lattice structures, but all the same those movements explain solidity.)

You can appreciate this objection to Nagel if you imagine other ways of detecting causally necessary relations. Suppose God or a machine could simply detect causally necessary relations, then for God or the machine there would be no difference between matter/matter forms of necessity and matter/mind forms of necessity. Furthermore, even if we grant that we cannot picture both sides of the relation for consciousness and the brain in a way that we can picture both sides of the relation for liquidity and molecule movement, we could nonetheless get at the causal relations involved in the production of consciousness by indirect means. Suppose we actually had an account of the neurophysiological processes in the brain that cause consciousness. It is not at all impossible that we should get such an account, because the usual tests for causal relations can be performed on brain/consciousness relations as they can on any natural phenomena. The knowledge of lawlike causal relations will give us all of the causal necessity we need. Indeed, we already have the beginnings of such lawlike relations. As I mentioned in chapter 3, standard textbooks of neurophysiology routinely explain, for example, the similarities and differences between how cats see things and how humans see things. There is no question that certain sorts of neurophysiological similarities and differences are causally sufficient for certain sorts of similarities and differences in visual experiences. Furthermore we can and will break down the big question—How does the brain cause consciousness?—into a lot of smaller questions (for example, How does cocaine produce certain characteristic experiences?). And the detailed answers we are already starting to give (for example, Cocaine impedes the capacity of certain synaptic receptors to reabsorb norepinephrine) already allow for the characteristic inferences that go with causal necessity (for example, If you increase the dose of cocaine, you increase the effect). I conclude that Nagel