24

THOMAS NAGEL

What Is It Like to Be a Bat?

Consciousness is what makes the mind–body problem really intractable. Perhaps that is why current discussions of the problem give it little attention or get it obviously wrong. The recent wave of reductionist euphoria has produced several analyses of mental phenomena and mental concepts designed to explain the possibility of some variety of materialism, psychophysical identification, or reduction.* But the problems dealt with are those common to this type of reduction and other types, and what makes the mind–body problem unique, and unlike the water–H_2O problem or the Turing machine–IBM machine problem or the lightning–electrical discharge problem or the gene–DNA problem or the oak tree–hydrocarbon problem, is ignored.

Every reductionist has his favorite analogy from modern science. It is most unlikely that any of these unrelated examples of successful reduction will shed light on the relation of mind to brain. But philosophers share the general human weakness for explanations of what is incomprehensible in terms suited for what is familiar and well understood, though entirely different. This has led to the acceptance of implausible accounts of the mental largely because they would permit familiar kinds of reduction. I shall try to explain why the usual examples do not help us to understand the relation between mind and body—why, indeed, we have at present no conception of what an explanation of the physical nature

"What Is It Like to Be a Bat?" by Thomas Nagel appeared in *The Philosophical Review*, October 1974. It is reprinted by permission of the author.

*See "Further Reading" for Nagel's references.

of a mental phenomenon would be. Without consciousness the mind–body problem would be much less interesting. With consciousness it seems hopeless. The most important and characteristic feature of conscious mental phenomena is very poorly understood. Most reductionist theories do not even try to explain it. And careful examination will show that no currently available concept of reduction is applicable to it. Perhaps a new theoretical form can be devised for the purpose, but such a solution, if it exists, lies in the distant intellectual future.

Conscious experience is a widespread phenomenon. It occurs at many levels of animal life, though we cannot be sure of its presence in the simpler organisms, and it is very difficult to say in general what provides evidence of it. (Some extremists have been prepared to deny it even of mammals other than man.) No doubt it occurs in countless forms totally unimaginable to us, on other planets in other solar systems throughout the universe. But no matter how the form may vary, the fact that an organism has conscious experience *at all* means, basically, that there is something it is like to *be* that organism. There may be further implications about the form of the experience; there may even (though I doubt it) be implications about the behavior of the organism. But fundamentally an organism has conscious mental states if and only if there is something that it is like to *be* that organism—something it is like *for* the organism.

We may call this the subjective character of experience. It is not captured by any of the familiar, recently devised reductive analyses of the mental, for all of them are logically compatible with its absence. It is not analyzable in terms of any explanatory system of functional states, or intentional states, since these could be ascribed to robots or automata that behaved like people though they experienced nothing.* It is not analyzable in terms of the causal role of experiences in relation to typical human behavior—for similar reasons.† I do not deny that conscious mental states and events cause behavior, nor that they may be given functional characterizations. I deny only that this kind of thing exhausts their analysis. Any reductionist program has to be based on an analysis of what is to be reduced. If the analysis leaves something out, the problem will be falsely posed. It is useless to base the defense of materialism on any analysis of mental phenomena that fails to deal explicitly with their

*Perhaps there could not actually be such robots. Perhaps anything complex enough to behave like a person would have experiences. But that, if true, is a fact which cannot be discovered merely by analyzing the concept of experience.

†It is not equivalent to that about which we are incorrigible, both because we are not incorrigible about experience and because experience is present in animals lacking language and thought, who have no beliefs at all about their experiences.

subjective character. For there is no reason to suppose that a reduction which seems plausible when no attempt is made to account for consciousness can be extended to include consciousness. Without some idea, therefore, of what the subjective character of experience is, we cannot know what is required of physicalist theory.

While an account of the physical basis of mind must explain many things, this appears to be the most difficult. It is impossible to exclude the phenomenological features of experience from a reduction in the same way that one excludes the phenomenal features of an ordinary substance from a physical or chemical reduction of it—namely, by explaining them as effects on the minds of human observers (cf. Rorty 1965). If physicalism is to be defended, the phenomenological features must themselves be given a physical account. But when we examine their subjective character it seems that such a result is impossible. The reason is that every subjective phenomenon is essentially connected with a single point of view, and it seems inevitable that an objective, physical theory will abandon that point of view.

Let me first try to state the issue somewhat more fully than by referring to the relation between the subjective and the objective, or between the *pour soi* and the *en soi.* This is far from easy. Facts about what it is like to be an X are very peculiar, so peculiar that some may be inclined to doubt their reality, or the significance of claims about them. To illustrate the connection between subjectivity and a point of view, and to make evident the importance of subjective features, it will help to explore the matter in relation to an example that brings out clearly the divergence between the two types of conception, subjective and objective.

I assume we all believe that bats have experience. After all, they are mammals, and there is no more doubt that they have experience than that mice or pigeons or whales have experience. I have chosen bats instead of wasps or flounders because if one travels too far down the phylogenetic tree, people gradually shed their faith that there is experience there at all. Bats, although more closely related to us than those other species, nevertheless present a range of activity and a sensory apparatus so different from ours that the problem I want to pose is exceptionally vivid (though it certainly could be raised with other species). Even without the benefit of philosophical reflection, anyone who has spent some time in an enclosed space with an excited bat knows what it is to encounter a fundamentally *alien* form of life.

I have said that the essence of the belief that bats have experience is that there is something that it is like to be a bat. Now we know that most bats (the microchiroptera, to be precise) perceive the external world primarily by sonar, or echolocation, detecting the reflections, from ob-

jects within range, of their own rapid, subtly modulated, high-frequency shrieks. Their brains are designed to correlate the outgoing impulses with the subsequent echoes, and the information thus acquired enables bats to make precise discriminations of distance, size, shape, motion, and texture comparable to those we make by vision. But bat sonar, though clearly a form of perception, is not similar in its operation to any sense that we possess, and there is no reason to suppose that it is subjectively like anything we can experience or imagine. This appears to create difficulties for the notion of what it is like to be a bat. We must consider whether any method will permit us to extrapolate to the inner life of the bat from our own case,* and if not, what alternative methods there may be for understanding the notion.

Our own experience provides the basic material for our imagination, whose range is therefore limited. It will not help to try to imagine that one has webbing on one's arms, which enables one to fly around at dusk and dawn catching insects in one's mouth; that one has very poor vision, and perceives the surrounding world by a system of reflected high-frequency sound signals; and that one spends the day hanging upside down by one's feet in an attic. Insofar as I can imagine this (which is not very far), it tells me only what it would be like for *me* to behave as a bat behaves. But that is not the question. I want to know what it is like for a *bat* to be a bat. Yet if I try to imagine this, I am restricted to the resources of my own mind, and those resources are inadequate to the task. I cannot perform it either by imagining additions to my present experience, or by imagining segments gradually subtracted from it, or by imagining some combination of additions, subtractions, and modifications.

To the extent that I could look and behave like a wasp or a bat without changing my fundamental structure, my experiences would not be anything like the experiences of those animals. On the other hand, it is doubtful that any meaning can be attached to the supposition that I should possess the internal neurophysiological constitution of a bat. Even if I could by gradual degrees be transformed into a bat, nothing in my present constitution enables me to imagine what the experiences of such a future stage of myself thus metamorphosed would be like. The best evidence would come from the experiences of bats, if we only knew what they were like.

So if extrapolation from our own case is involved in the idea of what it is like to be a bat, the extrapolation must be incompletable. We cannot

*By "our own case" I do not mean just "my own case," but rather the mentalistic ideas that we apply unproblematically to ourselves and other human beings.

form more than a schematic conception of what it *is* like. For example, we may ascribe general *types* of experience on the basis of the animal's structure and behavior. Thus we describe bat sonar as a form of three-dimensional forward perception; we believe that bats feel some versions of pain, fear, hunger, and lust, and that they have other, more familiar types of perception besides sonar. But we believe that these experiences also have in each case a specific subjective character, which it is beyond our ability to conceive. And if there is conscious life elsewhere in the universe, it is likely that some of it will not be describable even in the most general experiential terms available to us.* (The problem is not confined to exotic cases, however, for it exists between one person and another. The subjective character of the experience of a person deaf and blind from birth is not accessible to me, for example, nor presumably is mine to him. This does not prevent us each from believing that the other's experience has such a subjective character.)

If anyone is inclined to deny that we can believe in the existence of facts like this whose exact nature we cannot possibly conceive, he should reflect that in contemplating the bats we are in much the same position that intelligent bats or Martians† would occupy if they tried to form a conception of what it was like to be us. The structure of their own minds might make it impossible for them to succeed, but we know they would be wrong to conclude that there is not anything precise that it is like to be us: that only certain general types of mental state could be ascribed to us (perhaps perception and appetite would be concepts common to us both; perhaps not). We know they would be wrong to draw such a skeptical conclusion because we know what it is like to be us. And we know that while it includes an enormous amount of variation and complexity, and while we do not possess the vocabulary to describe it adequately, its subjective character is highly specific, and in some respects describable in terms that can be understood only by creatures like us. The fact that we cannot expect ever to accommodate in our language a detailed description of Martian or bat phenomenology should not lead us to dismiss as meaningless the claim that bats and Martians have experiences fully comparable in richness of detail to our own. It would be fine if someone were to develop concepts and a theory that enabled us to think about those things; but such an understanding may be permanently denied to us by the limits of our nature. And to deny the reality or logical significance of what we can never

*Therefore the analogical form of the English expression "what it is *like*" is misleading. It does not mean "what (in our experience) it *resembles*," but rather "how it is for the subject himself."

†Any intelligent extraterrestrial beings totally different from us.

describe or understand is the crudest form of cognitive dissonance.

This brings us to the edge of a topic that requires much more discussion than I can give it here: namely, the relation between facts on the one hand and conceptual schemes or systems of representation on the other. My realism about the subjective domain in all its forms implies a belief in the existence of facts beyond the reach of human concepts. Certainly it is possible for a human being to believe that there are facts which humans never *will* possess the requisite concepts to represent or comprehend. Indeed, it would be foolish to doubt this, given the finiteness of humanity's expectations. After all, there would have been transfinite numbers even if everyone had been wiped out by the Black Death before Cantor discovered them. But one might also believe that there are facts which *could* not ever be represented or comprehended by human beings, even if the species lasted forever—simply because our structure does not permit us to operate with concepts of the requisite type. This impossibility might even be observed by other beings, but it is not clear that the existence of such beings, or the possibility of their existence, is a precondition of the significance of the hypothesis that there are humanly inaccessible facts. (After all, the nature of beings with access to humanly inaccessible facts is presumably itself a humanly inaccessible fact.) Reflection on what it is like to be a bat seems to lead us, therefore, to the conclusion that there are facts that do not consist in the truth of propositions expressible in a human language. We can be compelled to recognize the existence of such facts without being able to state or comprehend them.

I shall not pursue this subject, however. Its bearing on the topic before us (namely, the mind–body problem) is that it enables us to make a general observation about the subjective character of experience. Whatever may be the status of facts about what it is like to be a human being, or a bat, or a Martian, these appear to be facts that embody a particular point of view.

I am not adverting here to the alleged privacy of experience to its possessor. The point of view in question is not one accessible only to a single individual. Rather it is a *type.* It is often possible to take up a point of view other than one's own, so the comprehension of such facts is not limited to one's own case. There is a sense in which phenomenological facts are perfectly objective: One person can know or say of another what the quality of the other's experience is. They are subjective, however, in the sense that even this objective ascription of experience is possible only for someone sufficiently similar to the object of ascription to be able to adopt his point of view—to understand the ascription in the first person as well as in the third, so to speak. The more different from oneself the

other experiencer is, the less success one can expect with this enterprise. In our own case we occupy the relevant point of view, but we will have as much difficulty understanding our own experience properly if we approach it from another point of view as we would if we tried to understand the experience of another species without taking up *its* point of view.*

This bears directly on the mind–body problem. For if the facts of experience—facts about what it is like *for* the experiencing organism—are accessible only from one point of view, then it is a mystery how the true character of experiences could be revealed in the physical operation of that organism. The latter is a domain of objective facts par excellence—the kind that can be observed and understood from many points of view and by individuals with differing perceptual systems. There are no comparable imaginative obstacles to the acquisition of knowledge about bat neurophysiology by human scientists, and intelligent bats or Martians might learn more about the human brain than we ever will.

This is not by itself an argument against reduction. A Martian scientist with no understanding of visual perception could understand the rainbow, or lightning, or clouds as physical phenomena, though he would never be able to understand the human concepts of rainbow, lightning, or cloud, or the place these things occupy in our phenomenal world. The objective nature of the things picked out by these concepts could be apprehended by him because, although the concepts themselves are connected with a particular point of view and a particular visual phenomenology, the things apprehended from that point of view are not: they are observable from the point of view but external to it; hence they can be comprehended from other points of view also, either by the same organisms or by others. Lightning has an objective character that is not exhausted by its visual appearance, and this can be investigated by a Martian without vision. To be precise, it has a *more* objective character than is revealed in its visual appearance. In speaking of the move from subjective to objective characterization, I wish to remain noncommittal about the existence of an end point, the completely objective intrinsic

*It may be easier than I suppose to transcend interspecies barriers with the aid of the imagination. For example, blind people are able to detect objects near them by a form of sonar, using vocal clicks or taps of a cane. Perhaps if one knew what that was like, one could by extension imagine roughly what it was like to possess the much more refined sonar of a bat. The distance between oneself and other persons and other species can fall anywhere on a continuum. Even for other persons the understanding of what it is like to be them is only partial, and when one moves to species very different from oneself, a lesser degree of partial understanding may still be available. The imagination is remarkably flexible. My point, however, is not that we cannot *know* what it is like to be a bat. I am not raising that epistemological problem. My point is rather that even to form a *conception* of what it is like to be a bat (and *a fortiori* to know what it is like to be a bat) one must take up the bat's point of view. If one can take it up roughly, or partially, then one's conception will also be rough or partial. Or so it seems in our present state of understanding.

nature of the thing, which one might or might not be able to reach. It may be more accurate to think of objectivity as a direction in which the understanding can travel. And in understanding a phenomenon like lightning, it is legitimate to go as far away as one can from a strictly human viewpoint.*

In the case of experience, on the other hand, the connection with a particular point of view seems much closer. It is difficult to understand what could be meant by the *objective* character of an experience, apart from the particular point of view from which its subject apprehends it. After all, what would be left of what it was like to be a bat if one removed the viewpoint of the bat? But if experience does not have, in addition to its subjective character, an objective nature that can be apprehended from many different points of view, then how can it be supposed that a Martian investigating my brain might be observing physical processes which were my mental processes (as he might observe physical processes which were bolts of lightning), only from a different point of view? How, for that matter, could a human physiologist observe them from another point of view?†

We appear to be faced with a general difficulty about psychophysical reduction. In other areas the process of reduction is a move in the direction of greater objectivity, toward a more accurate view of the real nature of things. This is accomplished by reducing our dependence on individual or species-specific points of view toward the object of investigation. We describe it not in terms of the impressions it makes on our senses, but in terms of its more general effects and of properties detectable by means other than the human senses. The less it depends on a specifically human viewpoint, the more objective is our description. It is possible to follow this path because although the concepts and ideas we employ in thinking about the external world are initially applied from a point of view that involves our perceptual apparatus, they are used by us to refer to things beyond themselves—toward which we *have* the phenomenal point of view. Therefore we can abandon it in favor of another, and still be thinking about the same things.

Experience itself, however, does not seem to fit the pattern. The idea

*The problem I am going to raise can therefore be posed even if the distinction between more subjective and more objective descriptions or viewpoints can itself be made only within a larger human point of view. I do not accept this kind of conceptual relativism, but it need not be refuted to make the point that psychophysical reduction cannot be accommodated by the subjective-to-objective model familiar from other cases.

†The problem is not just that when I look at the Mona Lisa, my visual experience has a certain quality, no trace of which is to be found by someone looking into my brain. For even if he did observe there a tiny image of the Mona Lisa, he would have no reason to identify it with the experience.

of moving from appearance to reality seems to make no sense here. What is the analogue in this case to pursuing a more objective understanding of the same phenomena by abandoning the initial subjective viewpoint toward them in favor of another that is more objective but concerns the same thing? Certainly it *appears* unlikely that we will get closer to the real nature of human experience by leaving behind the particularity of our human point of view and striving for a description in terms accessible to beings that could not imagine what it was like to be us. If the subjective character of experience is fully comprehensible only from one point of view, then any shift to greater objectivity—that is, less attachment to a specific viewpoint—does not take us nearer to the real nature of the phenomenon: It takes us farther away from it.

In a sense, the seeds of this objection to the reducibility of experience are already detectable in successful cases of reduction; for in discovering sound to be, in reality, a wave phenomenon in air or other media, we leave behind one viewpoint to take up another, and the auditory, human or animal viewpoint that we leave behind remains unreduced. Members of radically different species may both understand the same physical events in objective terms, and this does not require that they understand the phenomenal forms in which those events appear to the senses of members of the other species. Thus it is a condition of their referring to a common reality that their more particular viewpoints are not part of the common reality that they both apprehend. The reduction can succeed only if the species-specific viewpoint is omitted from what is to be reduced.

But while we are right to leave this point of view aside in seeking a fuller understanding of the external world, we cannot ignore it permanently, since it is the essence of the internal world, and not merely a point of view on it. Most of the neobehaviorism of recent philosophical psychology results from the effort to substitute an objective concept of mind for the real thing, in order to have nothing left over which cannot be reduced. If we acknowledge that a physical theory of mind must account for the subjective character of experience, we must admit that no presently available conception gives us a clue how this could be done. The problem is unique. If mental processes are indeed physical processes, then there is something it is like, intrinsically,* to undergo certain physical pro-

*The relation would therefore not be a contingent one, like that of a cause and its distinct effect. It would be necessarily true that a certain physical state felt a certain way. Kripke (1972) argues that causal behaviorist and related analyses of the mental fail because they construe, for example, "pain" as a merely contingent name of pains. The subjective character of an experience ("its immediate phenomenological quality," Kripke calls it [p. 340]) is the essential property left out by such analyses, and the one in virtue of which it is, necessarily, the experience it is. My view is closely related to his. Like Kripke, I find the hypothesis that a certain brain state should *necessarily* have a certain subjective character

cesses. What it is for such a thing to be the case remains a mystery.

What moral should be drawn from these reflections, and what should be done next? It would be a mistake to conclude that physicalism must be false. Nothing is proved by the inadequacy of physicalist hypotheses that assume a faulty objective analysis of mind. It would be truer to say that physicalism is a position we cannot understand because we do not at present have any conception of how it might be true. Perhaps it will be thought unreasonable to require such a conception as a condition of understanding. After all, it might be said, the meaning of physicalism is clear enough: mental states are states of the body; mental events are physical events. We do not know *which* physical states and events they are, but that should not prevent us from understanding the hypothesis. What could be clearer than the words "is" and "are"?

But I believe it is precisely this apparent clarity of the word "is" that is deceptive. Usually, when we are told that X is Y we know *how* it is supposed to be true, but that depends on a conceptual or theoretical background and is not conveyed by the "is" alone. We know how both "X" and "Y" refer, and the kinds of things to which they refer, and we have a rough idea how the two referential paths might converge on a single thing, be it an object, a person, a process, an event or whatever. But when the two terms of the identification are very disparate it may not

incomprehensible without further explanation. No such explanation emerges from theories which view the mind–brain relation as contingent, but perhaps there are other alternatives, not yet discovered.

A theory that explained how the mind–brain relation was necessary would still leave us with Kripke's problem of explaining why it nevertheless appears contingent. That difficulty seems to me surmountable, in the following way. We may imagine something by representing it to ourselves either perceptually, sympathetically, or symbolically. I shall not try to say how symbolic imagination works, but part of what happens in the other two cases is this. To imagine something perceptually, we put ourselves in a conscious state resembling the state we would be in if we perceived it. To imagine something sympathetically, we put ourselves in a conscious state resembling the thing itself. (This method can be used only to imagine mental events and states—our own or another's.) When we try to imagine a mental state occurring without its associated brain state, we first sympathetically imagine the occurrence of the mental state: that is, we put ourselves into a state that resembles it mentally. At the same time, we attempt perceptually to imagine the nonoccurrence of the associated physical state, by putting ourselves into another state unconnected with the first: one resembling that which we would be in if we perceived the nonoccurrence of the physical state. Where the imagination of physical features is perceptual and the imagination of mental features is sympathetic, it appears to us that we can imagine any experience occurring without its associated brain state, and vice versa. The relation between them will appear contingent even if it is necessary, because of the independence of the disparate types of imagination.

(Solipsism, incidentally, results if one misinterprets sympathetic imagination as if it worked like perceptual imagination: It then seems impossible to imagine any experience that is not one's own.)

be so clear how it could be true. We may not have even a rough idea of how the two referential paths could converge, or what kind of things they might converge on, and a theoretical framework may have to be supplied to enable us to understand this. Without the framework, an air of mysticism surrounds the identification.

This explains the magical flavor of popular presentations of fundamental scientific discoveries, given out as propositions to which one must subscribe without really understanding them. For example, people are now told at an early age that all matter is really energy. But despite the fact that they know what "is" means, most of them never form a conception of what makes this claim true, because they lack the theoretical background.

At the present time the status of physicalism is similar to that which the hypothesis that matter is energy would have had if uttered by a pre-Socratic philosopher. We do not have the beginnings of a conception of how it might be true. In order to understand the hypothesis that a mental event is a physical event, we require more than an understanding of the word "is." The idea of how a mental and a physical term might refer to the same thing is lacking, and the usual analogies with theoretical identification in other fields fail to supply it. They fail because if we construe the reference of mental terms to physical events on the usual model, we either get a reappearance of separate subjective events as the effects through which mental reference to physical events is secured, or else we get a false account of how mental terms refer (for example, a causal behaviorist one).

Strangely enough, we may have evidence for the truth of something we cannot really understand. Suppose a caterpillar is locked in a sterile safe by someone unfamiliar with insect metamorphosis, and weeks later the safe is reopened, revealing a butterfly. If the person knows that the safe has been shut the whole time, he has reason to believe that the butterfly is or was once the caterpillar, without having any idea in what sense this might be so. (One possibility is that the caterpillar contained a tiny winged parasite that devoured it and grew into the butterfly.)

It is conceivable that we are in such a position with regard to physicalism. Donald Davidson has argued that if mental events have physical causes and effects, they must have physical descriptions. He holds that we have reason to believe this even though we do not—and in fact *could* not—have a general psychophysical theory.* His argument applies to intentional mental events, but I think we also have some reason to believe that

*See Davidson (1970); though I do not understand the argument against psychophysical laws.

sensations are physical processes, without being in a position to understand how. Davidson's position is that certain physical events have irreducibly mental properties, and perhaps some view describable in this way is correct. But nothing of which we can now form a conception corresponds to it; nor have we any idea what a theory would be like that enabled us to conceive of it.*

Very little work has been done on the basic question (from which mention of the brain can be entirely omitted) whether any sense can be made of experiences' having an objective character at all. Does it make sense, in other words, to ask what my experiences are *really* like, as opposed to how they appear to me? We cannot genuinely understand the hypothesis that their nature is captured in a physical description unless we understand the more fundamental idea that they *have* an objective nature (or that objective processes can have a subjective nature).†

I should like to close with a speculative proposal. It may be possible to approach the gap between subjective and objective from another direction. Setting aside temporarily the relation between the mind and the brain, we can pursue a more objective understanding of the mental in its own right. At present we are completely unequipped to think about the subjective character of experience without relying on the imagination—without taking up the point of view of the experiential subject. This should be regarded as a challenge to form new concepts and devise a new method—an objective phenomenology not dependent on empathy or the imagination. Though presumably it would not capture everything, its goal would be to describe, at least in part, the subjective character of experiences in a form comprehensible to beings incapable of having those experiences.

We would have to develop such a phenomenology to describe the sonar experiences of bats; but it would also be possible to begin with humans. One might try, for example, to develop concepts that could be used to explain to a person blind from birth what it was like to see. One would reach a blank wall eventually, but it should be possible to devise a method of expressing in objective terms much more than we can at present, and with much greater precision. The loose intermodal analogies—for example, "Red is like the sound of a trumpet"—which crop up in discussions of this subject are of little use. That should be clear to anyone who has both heard a trumpet and seen red. But structural features of perception might be more accessible to objective description, even though something would be left out. And concepts alternative to those we learn in the first person may enable us to arrive at a kind of understanding even of our own experience which is denied us by the very ease of description and lack of distance that subjective concepts afford.

Apart from its own interest, a phenomenology that is in this sense objective may permit questions about the physical* basis of experience to assume a more intelligible form. Aspects of subjective experience that admitted this kind of objective description might be better candidates for objective explanations of a more familiar sort. But whether or not this guess is correct, it seems unlikely that any physical theory of mind can be contemplated until more thought has been given to the general problem of subjective and objective. Otherwise we cannot even pose the mind–body problem without sidestepping it.

*Similar remarks apply to Nagel (1965).

†This question also lies at the heart of the problem of other minds, whose close connection with the mind–body problem is often overlooked. If one understood how subjective experience could have an objective nature, one would understand the existence of subjects other than oneself.

Reflections

He does all the things that you would never do;
He loves me, too—
His love is true.
Why can't he be you?
—Hank Cochran, ca. 1955

Twinkle, twinkle, little bat,
How I wonder what you're at,
Up above the world you fly,
Like a tea-tray in the sky.
—Lewis Carroll, ca. 1865

There is a famous puzzle in mathematics and physics courses. It asks, "Why does a mirror reverse left and right, but not up and down?" It gives many people pause for thought, and if you don't want to be told the answer, skip the next two paragraphs.

The answer hinges on what we consider a suitable way to project ourselves onto our mirror images. Our first reaction is that by walking forward a few steps and then spinning around on our heels, we could step into the shoes of "that person" there in the mirror—forgetting that the heart, appendix, and so forth of "that person" are on the wrong side. The

*I have not defined the term "physical." Obviously it does not apply just to what can be described by the concepts of contemporary physics, since we expect further developments. Some may think there is nothing to prevent mental phenomena from eventually being recognized as physical in their own right. But whatever else may be said of the physical, it has to be objective. So if our idea of the physical ever expands to include mental phenomena, it will have to assign them an objective character—whether or not this is done by analyzing them in terms of other phenomena already regarded as physical. It seems to me more likely, however, that mental–physical relations will eventually be expressed in a theory whose fundamental terms cannot be placed clearly in either category.

language hemisphere of the brain is, in all probability, on the nonstandard side. On a gross anatomical level, that image is actually of a nonperson. Microscopically, the situation is even worse. The DNA molecules coil the wrong way, and the mirror-"person" could no more mate with a real person than could a nosrep!

But wait—you can get your heart to stay on the proper side if, instead, you flip yourself head over heels, as if swinging over a waist-high horizontal bar in front of you. Now your heart is on the same side as the mirror-person's heart—but your feet and head are in the wrong places, and your stomach, although at approximately the right height, is upside-down. So it seems a mirror *can* be perceived as reversing up and down, provided you're willing to map yourself onto a creature whose feet are above its head. It all depends on the ways that you are willing to slip yourself onto another entity. You have a choice of twirling around a horizontal or a vertical bar, and getting the heart right but not the head and feet, or getting the head and feet right but not the heart. It's simply that, because of the external vertical symmetry of the human body, the vertical self-twirling yields a more plausible-seeming you-to-image mapping. But mirrors intrinsically don't care which way you interpret what they do. And in fact, all they really reverse is back and front!

There is something very beguiling about this concept of mapping, projection, identification, empathy—whatever you want to call it. It is a basic human trait, practically irresistible. Yet it can lead us down very strange conceptual pathways. The preceding puzzle shows the dangers of over facile self-projection. The refrain quoted from the country-western ballad reminds us more poignantly of the futility of taking such mapping too seriously. Yet we can't stop our minds from doing it. So since we can't, let's go whole hog and indulge ourselves in an orgy of extravagant variations on the theme set by Nagel in his title.

What is it like to work at McDonald's? To be thirty-eight? To be in London today?

What is it like to climb Mount Everest? To be an Olympic gold-medal winner in gymnastics?

What would it be like to be a good musician? To be able to improvise fugues at the keyboard? To be J. S. Bach? To be J. S. Bach writing the last movement of the Italian Concerto?

What is it like to believe the earth is flat?

What is it like to be someone inconceivably more intelligent than yourself? Inconceivably less intelligent?

What is it like to hate chocolate (or your personal favorite flavor)?

What is it like to bat a bee? What is it like to be a bee being batted? What is it like to be a batted bee? (Illustration by Jim Hull.)

What is it like to hear English (or one's native language) without understanding it?

What is it like to be of the opposite sex? (See selection 15, "Beyond Rejection")

What would it be like to be your mirror image? (See the movie *Journey to the Far Side of the Sun*)

What would it be like to be Chopin's brother (he had none)? The present King of France?

What is it like to be a dreamed person? To be a dreamed person when the alarm rings? To be Holden Caulfield? To be the subsystem of J. D. Salinger's brain that represents the character of Holden Caulfield?

What is it like to be a molecule? A collection of molecules? A microbe? A mosquito? An ant? An ant colony? A beehive? China? The United States? Detroit? General Motors? A concert audience? A basketball team? A married couple? A two-headed cow? Siamese twins? A split-brain person? One half of a split-brain person? The head of a guillotined person? The body? The visual cortex of Picasso? The pleasure center of a rat? The jerking leg of a dissected frog? A bee's eye? A retinal cell in Picasso? A DNA molecule of Picasso?

What is it like to be a running AI program? An operating system in a computer? An operating system at the moment the system "crashes"?

What is it like to be under a general anesthetic? To be electrocuted? To be a Zen master who has attained a satori-like state in which no more subject ("I," ego, self) exists?

What is it like to be a pebble? A wind chime? A human body? The Rock of Gibraltar? The Andromeda Galaxy? God?

The image conjured up by the phrase "What is it like to be *X*"? is so seductive and tempting. . . . Our minds are so flexible, so willing to accept this notion, this idea that there is "something it is like to be a bat." Furthermore, we also willingly buy the idea that there are certain things that it is "like something to be"—"be-able things," or "BATs" for short —such as bats, cows, people; and other things for which this doesn't hold —such as balls, steaks, galaxies (even though a galaxy may contain innumerable be-able things). What is the criterion for "BAT-itude"?

In philosophical literature, many phrases have been used to try to evoke the right flavors for what being sentient really is ("being sentient" is one of them). Two old terms are "soul" and "anima." These days, an "in" word is "intentionality." There is the old standby, "consciousness." Then there is "being a subject," "having an inner life," "having experience," "having a point of view," having "perceptual aboutness" or "personhood" or a "self" or "free will." In some people's eyes, "having a mind," "being intelligent," and just plain old "thinking" have the right flavors. In Searle's article (selection 22), the contrast was drawn between "form" (hollow and mechanical) and "content" (alive and intentional); the words "syntactic" and "semantic" (or "meaningless" and "meaningful") were also used to characterize this distinction. All of the terms in this huge showcase are nearly synonymous. They all have to do with the emotional issue of whether it makes sense to project ourselves onto the object in question: "Is this object a BAT, or not?" But is there really some *thing* to which they refer?

Nagel makes it clear that the "thing" he is after is a distillation of that which is common to the experiences of all bats; it is not the set of experiences of some particular bat. Thus, Searle might say Nagel is a "dualist," since Nagel believes in some abstraction made from all those individuals' experiences.

Surprisingly enough, a look at the grammar of sentences that invite the reader to perform a mental mapping yields some insights into these tricky matters. Consider, for instance, the contrast between the questions "What *would it be like* to be Indira Gandhi?" and "What *is it like* to be Indira Gandhi?" The conditional sentence forces you to project yourself into the "skin," so to speak, of another human, whereas the indicative sentence seems to be asking what it is like for Indira Gandhi to be Indira Gandhi. The question might still be asked, "Described in whose terms?" Were Indira Gandhi to try to tell you what it is like to be Indira Gandhi, she might try to explain matters of political life in India by referring to things she considered vaguely analogous in your own experience. Would you protest and say, "No, don't translate it into *my* terms! Say it in your

own terms! Tell me what it is like—to Indira Gandhi—for Indira Gandhi to be Indira Gandhi!" In that case, of course, she might as well speak in Hindi and leave it to you to learn the language. And yet even then you would just be in the position of millions of native Hindi speakers who have no idea what it would be like to be Indira Gandhi—much less what it is like for Indira Gandhi to be Indira Gandhi. . . .

Something seems very wrong here. Nagel is insistent that he wants his verb "be" to be subjectless, in effect. Not "What would it be like *for me* to be *X*"? but "What is it like, *objectively,* to be X?" There is a "be-ee" here, with no "be-er"—a living beast without a head, as it were. Perhaps we ought to go back to the conditional version: "What would it be like to be Indira Gandhi?" Well, for me, or for her? Poor Indira—where does *she* go while I'm being her? Or if we turn it around (identity being a symmetric relationship), we get "What would it be like for Indira Gandhi to be me?" Once again, where would I be if she were me? Would we have traded places? Or would we have temporarily collapsed two separate "souls" into one?

Note that we tend to say "If she were me" rather than "If she were I." Many European languages are somewhat skittish about equations of this type. It sounds funny to use the nominative case in both the subject and complement positions. People prefer to use "be" with the accusative case, as if it were somehow a transitive verb! "Be" is not a transitive verb, but a symmetric one—yet language tilts us away from that symmetric vision.

We can see this in German, where one has interesting alternatives for constructing such identity-asserting sentences. Two examples follow, adapted from the German translation of a Stanislaw Lem dialogue in which an exact molecule-for-molecule replica of a doomed person is about to be constructed. In that spirit, we provide (nearly) exact word-for-word replicas in English of the German originals:

1. *Ob die Kopie wirklich du bist, dafür muß der Beweis noch erbracht werden.* (As-to-whether the copy really you are, thereof must the proof still provided be.)

2. *Die Kopie wird behaupten, daß sie du ist.* (The copy will claim that it you is.)

Observe that in both identity-asserting clauses, "the copy" (or "it") appears first, then "you," then the verb. But notice—in the first clause, "are" is the verb, which retroactively implies that "you" was the subject and "the copy" was the complement, whereas in the second clause, the verb is "is," retroactively implying that the subject was "it" and the

complement was "you." The fact that the verb comes at the end gives these clauses a sort of surprise-ending quality. In English we can't achieve precisely the same effect comfortably, but we can ask for the difference in shades of meaning between the sentences "Is the copy really you?" and "Are you really the copy?" These two questions "slip" in our minds along different dimensions. The former slips into "Or is the copy really someone else—or perhaps no one at all?" The latter slips into "Or are you somewhere else—or are you anywhere?" Our book's title, incidentally, can be construed not only as a possessive, but equally as a short full-sentence reply to the two questions "Who am I?" and "Who is me?" Notice how the transitive usage—strictly speaking, an ungrammatical usage of "to be"—gives the second question a quite different "flavor" from the first.

[D.C.D. to D.R.H.: If I were you, I'd mention how curious it would be to preface some advice with "If you were me, I'd . . ." but if you were me, would I suggest that you mention it?]

All of these examples show how suggestible we are. We just fall like a ton of bricks for the notion that there's a "soul" in there—a flamelike soul that can flicker on or off, or even be transferred between bodies as a flame between candles. If a candle blows out and is relit, is it "the same flame"? Or, if it stays lit, is it even "the same flame" from moment to moment? The Olympic Torch is carefully kept burning as it is carried by runners thousands of miles from Athens to its destination every four years. There is powerful symbolism to the idea that this is "the very flame that was lit in Athens." Even the shortest break in the chain, however, would ruin the symbolism for people who knew. For people who didn't know, of course, no harm done! How on earth could it possibly matter? Yet emotionally it seems to. It will not easily be extinguished, that "soul-flame" notion. Yet it leads us into so much hot water.

We certainly intuit that only things of approximately the "same-sized souls" can slip into each other. The science-fiction story *Flowers for Algernon* by Daniel Keyes is about a retarded young man who by a miracle medical treatment slowly gains in intelligence and becomes a great genius —but then it turns out that the effects of the treatment cannot last, and "he" witnesses his own mental crumbling back into his retarded state. This fictional story has its counterpart in the real-life tragedy of people who, having grown from a state of zero mind to normal adult intelligence, witness themselves growing senile or who suffer serious brain damage. Can they answer for us the question "What is it like to have your soul slip out from under you?" any better than someone with vivid imagination can, though?

Franz Kafka's *Metamorphosis* is the story of a young man who wakes

up one morning transformed into a giant beetle. But the beetle thinks like a person. It would be interesting to combine the *Flowers for Algernon* idea with the *Metamorphosis* idea and imagine the experiences of an insect whose intelligence rises to the level of a human genius (why not superhuman, for that matter?), then sinks back to the insect level. Yet this is virtually impossible for us to conceive. To borrow electrical-engineering jargon, the "impedance match" of the minds involved is too poor. In fact, impedance match may well be the main criterion for the plausibility of questions of the form Nagel poses. Which is it easier for you to imagine being—the totally fictional character Holden Caulfield or some particular, actual bat? Of course it is much easier to map yourself onto a fictional human than onto a real bat—much easier, much *realer.* This is slightly surprising. It seems that Nagel's verb "be" acts very strangely sometimes. Perhaps, as was suggested in the dialogue on the Turing test, the verb "be" is being extended. Perhaps it is even being stretched beyond its limits!

There's something very fishy about this whole idea. How can something *be* something that it *isn't*? And how is it rendered any more plausible when both things can "have experience"? It makes almost no sense for us to ask ourselves such questions as, "What would it be like for that black spider over there to be that mosquito trapped in its web?" Or worse yet, "What would it be like for my violin to be my guitar?" or "What would this sentence be like if it were a hippopotamus?" Like *for whom*? For the various objects concerned, sentient or not? For us the perceivers? Or, again, "objectively"?

This is the sticking-point of Nagel's article. He wants to know if it is possible to give, in his own words, "a description [of the real nature of human experience] in terms accessible to beings that could not imagine what it was like to be us." Put so starkly, it sounds like a blatant contradiction—and indeed, that is his point. He doesn't want to know what it's like *for him* to be a bat. He wants to know *objectively* what it is *subjectively* like. It wouldn't be enough for him to have had the experience of donning a "batter's helmet"—a helmet with electrodes that would stimulate his brain into batlike experiences—and to have thereby experienced "bat-itude." This would, after all, merely be what it would be like for *Nagel* to be a bat. What, then, would satisfy him? He's not sure that anything would, and that's what worries him. He fears that this notion of "having experience" is beyond the realm of the objective.

Now perhaps the most objective-sounding of the various synonyms earlier listed for BAT-itude is "having a point of view." After all, even the most dogmatic of disbelievers in machine intelligence would probably begrudgingly impute a "point of view" to a computer program that

represents some facts about the world and about its own relationship to the world. There is no arguing with the fact that a computer can be programmed to describe the world around it in terms of a frame of reference centered on the machine itself, as in this: "Three minutes ago, the Teddy bear was thirty-five leagues due east of here." Such a "here-centered, now-centered" frame of reference constitutes a rudimentary "egocentric" point of view. "Being here now" is a central experience for any "I." Yet how can you define "now" and "here" without making reference to some "I"? Is circularity inevitable?

Let us ponder for a moment on the connection of "I" and "now." What would it be like to be a person who had grown up normally, thus with ordinary perceptual and linguistic capacities, but who then suffered some brain damage and was left without the capacity to convert the reverberating neural circuits of short-term memory into long-term memories? Such a person's sense of existence would extend to only a few seconds on either side of "now." There would be no large-scale sense of continuity of self—no internal vision of a chain of selves stretching both directions in time, making one coherent person.

When you get a concussion, the few instants before it happened are obliterated from your mind, as if you had never been conscious at that time. Just think—if you were knocked on the head at this moment, there would be no permanent trace left in your brain of your having read these past few sentences. Who, then, has been experiencing them? Does an experience only become part of *you* once it has been committed to long-term memory? Who is it that has dreamt all those many dreams you don't remember one bit of?

Just as "now" and "I" are closely related terms, so are "here" and "I." Consider the fact that you are now experiencing death, in a curious way. Not being in Paris right now, you know what it is like to be *dead in Paris.* No lights, no sounds—nothing. The same goes for Timbuctu. In fact, you are dead *everywhere*—except for one small spot. Just think how close you are to being dead everywhere! And you are also dead in all other moments than *right now.* That one small piece of space-time you are alive in doesn't just *happen* to be where your body is now—it is *defined* by your body and by the concept of "now." Our languages all have words that incorporate a rich set of associations with "here" and "now"—namely, "I" and "me" and so on.

Now to program a computer to use words like "I" and "me" and "my" in describing its own relation to the world is a common thing. Of course, behind those words there need not stand any sophisticated self-concept—but there may. In essence, any physical representational system, as defined earlier in the commentary on the "Prelude, Ant Fugue"

(selection 11), is an embodiment of some point of view, however modest. This explicit connection between "having a point of view" and "being a representational system" now provides a step forward in thinking about BAT-itude, for if we can equate BATs with physical representational systems of sufficient richness in their repertoire of categories and sufficiently well-indexed memories of their worldlines, we will have objectified at least some of subjectivity.

It should be pointed out that what is strange about the idea of "being a bat" is *not* that bats sense the outside world in a bizarre way—it is that bats clearly have a highly reduced collection of conceptual and perceptual categories, compared to what we humans have. Sensory modalities are surprisingly interchangeable and equivalent, in some sense. For instance, it is possible to induce visual experiences in both blind and sighted people through the sensation of touch. A grid of over a thousand stimulators driven by a television camera is placed against a person's back. The sensations are carried to the brain where their processing can induce the having of visual experiences. A sighted woman reports on her experience of prosthetic vision:

> I sat blindfolded in the chair, the TSR cones cold against my back. At first I felt only formless waves of sensation. Collins said he was just waving his hand in front of me so that I could get used to the feeling. Suddenly I felt or saw, I wasn't sure which, a black triangle in the lower left corner of a square. The sensation was hard to pinpoint. I felt vibrations on my back, but the triangle appeared in a square frame inside my head. (Nancy Hechinger, "Seeing Without Eyes," *Science 81,* March 1981, p. 43.)

Similar transcending of modality in sensory input is well known. As has been pointed out in earlier selections, people who wear prism-shaped glasses that turn everything upside down can, after two or three weeks, get quite used to seeing the world this way. And, on a more abstract plane, people who learn a new language still experience the world of ideas in pretty much the same way.

So it is really not the mode of transduction of stimuli into percepts or the nature of the thought-supporting medium that makes the "bat *Weltanschauung*" different from ours. It is the severely limited set of categories, together with the stress on what is important in life and what is not. It is the fact that bats cannot form notions such as "the human *Weltanschauung*" and joke about them, because they are too busy, always being in raw-survival mode.

What Nagel's question forces us to think about—and think very hard about—is how we can map our *mind* onto that of a bat. What kind of representational system is the mind of a bat? Can we empathize with a

bat? In this view, Nagel's question seems intimately connected to the way in which one representational system emulates another, as discussed in the Reflections on selection 22. Would we learn something by asking a Sigma-5, "What is it like to be a DEC?" No, that would be a silly question. The reason it would be silly is this. An unprogrammed computer is not a representational system. Even when one computer has a program allowing it to emulate another, this does not give it the representational power to deal with the concepts involved in such a question. For that it would need a very sophisticated AI program—one that, among other things, could use the verb "be" in all the ways we do (including Nagel's extended sense). The question to ask would be, rather, "What is it like for you, as a self-understanding AI program, to emulate another such program?" But then this question starts to resemble very strongly the question "What is it like for one person to empathize strongly with another?"

As we pointed out earlier, people do not have the patience or accuracy to emulate a computer for any length of time. When trying to put themselves in the shoes of other BATs, people tend to empathize, not to emulate. They "subvert" their own internal symbol systems by voluntarily adopting a global set of biases that modify the cascades of symbolic activity in their brains. It is not quite the same as taking LSD, although that too creates radical changes in the way that neurons communicate with one another. LSD does so unpredictably. Its effects depend on how it is spread about inside the brain, and that has nothing to do with what symbolizes what. LSD affects thought in somewhat the same way that having a bullet shot through your brain would affect thought—neither intrusive substance pays any regard to the symbolic power of the stuff in the brain.

But a bias established through *symbolic* channels—"Hey, let me think about how it would feel to be a bat"—sets up a mental context. Translated into less mentalistic and more physical terms, the act of trying to project yourself into a bat's point of view activates some symbols in your brain. These symbols, as long as they remain activated, will contribute to the triggering patterns of all the other symbols that are activated. And the brain is sufficiently sophisticated that it can treat certain activations as stable—that is, as *contexts*—and other symbols then are activated in a subordinate manner. So when we attempt to "think bat," we subvert our brains by setting up neural contexts that channel our thoughts along different pathways than they usually follow. (Too bad we can't just "think Einstein" when we want!)

All this richness, however, cannot get us all the way to batitude. Each person's self-symbol—the "personal nucleus," or "gemma" in Lem's personetics—has become, over his or her life, so large and complicated

and idiosyncratic that it can no longer, chameleonlike, just assume the identity of another person or being. Its individual history is just too wound up in that little "knot" of a self-symbol.

It is interesting to think about two systems that are so alike that they have isomorphic, or identical, self-symbols—say a woman and an atom-by-atom replica of her. If she thinks about herself, is she also thinking about her replica? Many people fantasize that somewhere out there in the heavens, there is another person just like them. When you think about yourself, are you also thinking, without being aware of it, about that person? Who is that person thinking about right now? What would it be like to be that person? Are you that person? If you had a choice, would you let that person be killed, or yourself?

The one thing that Nagel seems not to have acknowledged in his article is that language (among other things) is a bridge that allows us to cross over into territory that is not ours. Bats don't have any idea of "what it is like to be another bat" and don't wonder about it, either. And that is because bats do not have a universal currency for the exchange of ideas, which is what language, movies, music, gestures, and so on give us. These media aid in our projection, aid us in absorbing foreign points of view. Through a universal currency, points of view become more *modular,* more transferable, less personal and idiosyncratic.

Knowledge is a curious blend of objective and subjective. Verbalizable knowledge can be passed around and shared, to the extent that words really "mean the same thing" to different people. Do two people ever speak the same language? What we mean by "speak the same language" is a prickly issue. We accept and take for granted that the hidden subterranean flavors are not shared. We know what comes with and what is left out of linguistic transactions, more or less. Language is a public medium for the exchange of the most private experiences. Each word is surrounded, in each mind, by a rich and inimitable cluster of concepts, and we know that no matter how much we try to bring to the surface, we always miss something. All we can do is approximate. (See George Steiner's *After Babel* for an extended discussion of this idea.)

By means of meme-exchange media (see selection 10, "Selfish Genes and Selfish Memes") such as language and gestures, we *can* experience (vicariously sometimes) what it is like to be or do *X.* It's never genuine, but then what is genuine knowledge of what it is like to be *X*? We don't even quite know what it was like to be ourselves ten years ago. Only by rereading diaries can we tell—and then, only by projection! It is still vicarious. Worse yet, we often don't even know how we could possibly have done what we did yesterday. And, when you come right down to it, it's not so clear just what it is like to be me, right now.

Language is what gets us into this problem (by allowing us to see the question) and what helps to get us out as well (by being a universal thought-exchange medium, allowing experiences to become sharable and more objective). However, it can't pull us all the way.

In a sense, Gödel's Theorem is a mathematical analogue of the fact that I cannot understand what it is like not to like chocolate, or to be a bat, except by an infinite sequence of ever-more-accurate simulation processes that converge toward, but never reach, emulation. I am trapped inside myself and therefore can't see how other systems are. Gödel's Theorem follows from a consequence of that general fact: I am trapped inside myself and therefore can't see how other systems see me. Thus the objectivity-subjectivity dilemmas that Nagel has sharply posed are somehow related to epistemological problems in both mathematical logic, and as we saw earlier, the foundations of physics. These ideas are developed in more detail in the last chapter of *Gödel, Escher, Bach* by Hofstadter.

D.R.H.

25

RAYMOND M. SMULLYAN

An Epistemological Nightmare

Scene 1. Frank is in the office of an eye doctor. The doctor holds up a book and asks "What color is it?" Frank answers, "Red." The doctor says, "Aha, just as I thought! Your whole color mechanism has gone out of kilter. But fortunately your condition is curable, and I will have you in perfect shape in a couple of weeks."

Scene 2. (A few weeks later.) Frank is in a laboratory in the home of an experimental epistemologist. (You will soon find out what that means!) The epistemologist holds up a book and also asks, "What color is this book?" Now, Frank has been earlier dismissed by the eye doctor as "cured." However, he is now of a very analytical and cautious temperament, and will not make any statement that can possibly be refuted. So Frank answers, "It seems red to me."

EPISTEMOLOGIST: Wrong!

FRANK: I don't think you heard what I said. I merely said that it *seems* red to me.

EPISTEMOLOGIST: I heard you, and you were wrong.

From *Philosophical Fantasies* by Raymond M. Smullyan, to be published by St. Martins Press, N.Y., in 1982.