

Consciousness, Naturalism, and Nagel

Owen Flanagan

Wellesley College

and

Duke University

In this paper I criticize Thomas Nagel's (1979) claim that consciousness is intractable from a naturalistic point of view. First, I show that there is a coherent conception of consciousness available to the naturalist which is both compatible with evolutionary theory and with certain widely acknowledged phenomenological features of conscious experience. Second, I discuss the adjustments that the naturalistic point of view requires to the traditional Cartesian conception of consciousness, in particular, to the doctrines of the unity of consciousness and privileged access. Third, I argue that the emerging picture of the mind within cognitive science as comprised of a variety of modular, serial, and parallel processors undermines the thesis that conscious awareness is a unified kind with a standard causal role. Finally, I take up Nagel's argument directly and disarm it by arguing that although Nagel is right that no theory can capture exactly the first person qualitative character of experience he is wrong to think this undermines the naturalistic picture of things. Indeed, I show that the naturalist easily can account for the fact Nagel makes so much of, namely, that conscious experience attaches uniquely to a single point of view.

The easiest mistake to make in trying to divine the nature of mind is to take what is phenomenologically most prominent for what is most essential or most real—to take what the mind seems to be like for what it is like. Descartes made this mistake when he announced:

I knew that I was a substance the whole essence or nature of which is to think, and that for its existence there is no need of any place, nor does it depend on any material thing; so this "me," that is to say, the soul by which I am what I am, is entirely distinct from body, and is even more easy to know than the latter; and even if body were not, the soul would not cease to be what it is. (1637/1975, p. 101)

This paper was written for a talk at a conference on "Consciousness and the Empiricist Tradition" at the University of South Carolina, April, 1985. An earlier version was read for the Faculty of Science at the National Autonomous University of Mexico, Mexico City, Mexico in August 1984. Special thanks to Stevan Harnad for some perceptive criticisms and to Paul Churchland for his appreciative comments. Conversations with Patricia Churchland, Barbara Von Eckardt, and Ken Winkler were also very helpful in clarifying some of my ideas. Finally, thanks to the National Endowment for the Humanities for support of this project. Requests for reprints should be sent to Owen Flanagan, Ph.D., Department of Philosophy, Duke University, Durham, North Carolina 27708 until July 1986. After this time, write to Department of Philosophy, Wellesley College, Wellesley, Massachusetts 02181.

Descartes' mistake was not so much in using the ubiquity of conscious experience as a way around skepticism regarding his own existence, as in thinking that the ubiquity of conscious experience had implications for his essential nature, for what he really was.

The temptation here, to leap from phenomenological claims to existential ones, is extraordinarily seductive. Julian Jaynes begins his popular book, *The Origin of Consciousness in the Breakdown of the Bicameral Mind* (1976), in the sort of rhapsodic Cartesian mood that makes it almost impossible to resist the temptation to think that one has come upon something of an altogether different ontological order. Jaynes sets what he calls "the problem of consciousness" this way:

O, what a world of unseen visions and heard silences, this insubstantial country of the mind! What ineffable essences, these touchless rememberings and unshowable reveries! And the privacy of it all! A speechless theater of speechless monologue and convenient counsel, an invisible mansion of moods, musings, and mysteries, an infinite resort of disappointments and discoveries. A whole kingdom where each of us reigns reclusively alone, questioning what we will, commanding what we can. A hidden hermitage where we may study out the troubled book of what we have done and yet may do. An introcosm that is more myself than anything I can find in the mirror. This consciousness that is myself of selves, that is everything, and yet nothing at all. (p. 1)

The doctrines that one's conscious self is one's essence, that this self is incorporeal, that it has omniscient reflexive access to itself, as well as complete control over volition, are all doctrines which result from taking Cartesian-like intuitions too seriously—from taking them, as Rorty (1982a) puts it, "ontologically." But this is something Descartes himself could not have been expected to know.

Knowing that intuitions about the nature of mind must be constrained—which is not to say defeated—required the advent of the sciences of the mind in the last century. It required trying to bring the Cartesian view of mind into equilibrium with an emerging naturalistic picture of mind; it required, to use Dennett's (1982) phrase, seeing how the autophenomenology, that is, the way things look from the first person point of view, meshes with the heterophenomenology, the way things look from the third person point of view, the point of view of an experimental science of the mind. The "phenomenology" suffix is by way of emphasizing that both the auto- and hetero- stories deal in appearances, and thus that both stories are subject to revision at every turn.

The attempt to bring the two stories into equilibrium has not been possible, even provisionally, without almost all the classical Cartesian intuitions losing some of their evidentiary status. Some behaviorists, and more recently some eliminativists, made the mistake of thinking that the intuitions themselves had to go, and reached the utterly implausible conclusion that conscious mental life was itself illusory—or epiphenomenal at best. Although the naturalist rejects the idea that the autophenomenology accurately mirrors the mind, he or she

rejects the hasty, and self-defeating, conclusion that there is nothing it is like to have a conscious mental life.

Despite the coherence and many successes of the naturalistic study of mind over the course of the last century, one hears again and again that consciousness itself remains a deep dark mystery, that there is no remotely plausible naturalistic account of consciousness, that something essential is, and always will be, left out of even our best theories. Thomas Nagel is the most well-known proponent of this line of argument. Nagel tells us that "Consciousness is what makes the mind-body problem really intractable Without consciousness the mind-body problem would be much less interesting. With consciousness it seems hopeless" (1979, pp. 165-166).

The sort of pessimism voiced by Nagel and others, it seems to me, is largely unwarranted. My strategy to show this is to sketch two related ways of understanding consciousness as a natural phenomenon. The two approaches share, as a deliberate tactic, a certain indirectness. The indirectness involves deliberately not trying to say what exactly consciousness is, not trying to generate necessary and sufficient conditions for the ascription of consciousness, but rather to, as it were, illuminate conscious experience by locating it within a general theory of the systems in which it figures. The first way involves trying to frame a coherent and explanatory naturalistic picture of the biological world in which mind and consciousness have a place. The second approach presupposes the first and then goes more fine-grained. It consists in trying to locate specifically the role(s) of conscious mental life within the overall economy of mental life itself. Here the facts are very sketchy, but the little we know about what the more fine-grained picture will look like indicates that the molar phenomenon of consciousness will dissolve into a large variety of different kinds of awarenesses with many different causal roles. Eventually I will explain how my analysis affects Nagel's argument.

Consciousness and Evolution

The first strategy involves a variation on the method of reflective equilibrium, the method of regimenting, organizing, and making consistent our intuitions and considered judgments about some domain. With respect to consciousness this will involve regimenting answers to several different sorts of fairly general questions. Among these questions are: What is conscious mental experience like phenomenologically?—or to be precise: what it is like autophenomenologically? How well and in what domains does the autophenomenology accurately reflect underlying processes, underlying functioning? How does our folk psychological taxonomy of mental states map onto the taxonomy needed by scientific psychology? What is conscious mental life for? That is, what function or functions, if any, does consciousness serve in the overall economy of organisms that possess it? How did consciousness evolve? How is it realized?

It seems to me that we have provisional answers to some of these questions—answers which taken together illuminate the allegedly unilluminable phenomena. With respect to the autophenomenological question, what we want in the first instance is a description as uncorrupted as possible by philosophical or ordinary folk theories of how experience seems. Of course, keeping this story uncorrupted will take lots of work. Not only do we learn the concepts and categories in terms of which we understand ourselves at society's knee, but we also theorize about our particular selves in terms of the folk psychology that dominates our particular locale. These folk theories incorporate in systematic and unsystematic ways prior philosophical wisdom—and the lack thereof.

The fact that we have no virginal intuitions about the mind as such does not imply, however, as Rorty suggests, that the mind is simply a "blur" (1982a). It does not imply that there is nothing it is like for members of our species to have conscious mental lives. It does not imply that there are no shared experiential features of conscious mental life which reveal themselves in spite of whatever folk psychology is dominant in some particular community. Here as elsewhere the project is one of seeking equilibrium—we will simply need to listen carefully to autophenomenologies from various and sundry places and try to detect the shared undercurrents, if any. The task is an awesome undertaking for philosophical anthropology; but it is possible, at least in principle.

James's famous analysis in the *Principles of Psychology* (1890/1970) is a not implausible candidate of an analysis of what conscious mental life might seem like for members of all human communities. According to James, conscious mental life feels continuous, like a "stream;" it is directed to objects both inside and outside oneself; it seems both to figure in the initiation of action and other times just to passively notice what is going on; it feels personal, and temporally sensitive, and both focussed and fuzzy.

How temporally sensitive, how focussed, and how personal conscious mental life feels will presumably be determined locally but that it, quite possibly, feels these ways across temporally and culturally distinct persons is the important thing. (On the other hand, one might be struck—not implausibly—with how thin and meager the shared autophenomenology ultimately is).

One of the virtues of James's analysis, it seems to me, is that it is not an analysis of what consciousness is, nor is it a theory of what features each token mental state must have to qualify as conscious. Token states, for example, do not feel continuous; it is just that overall our experiences tend to hang together. Nor, as we now know, are all conscious states intentional or all intentional states conscious. Pains, for example, are experienced but they do not have intentional contents, and, as we shall see in a moment, there is reasonable evidence of causally efficacious representational states of a non-Freudian sort. Such facts do not mean that James's account fails to isolate what conscious mental life is like even from an autophenomenological point of view, but rather that even from that point of view there is no one simple thing conscious experience is like.

Another attraction of James's account is that it can be made to fit with both an evolutionary account of the origin of consciousness and a materialistic view of how mental events are realized. Conscious mental activity, as part of mental life as a whole, must have initially emerged out of the complex organization of the nervous system — itself the long-term production of evolutionary processes.

Given this assumption that consciousness is a product of evolution the question remains as to whether it is an utterly baroque product, playing no important causal role in the lives of organisms which possess it, or whether it plays the important causal role it seems to. Although both the power and ontological uniqueness of consciousness have been dramatically overstated traditionally, the evidence (to everyone but the unflinching skeptic) points in the direction of conscious states having some causal efficacy as well as some not insignificant degree of representational reliability. Not only do we have the ability to consciously represent past, present, and future states of affairs to ourselves but we are also — just as the autophenomenology indicates — designed to use these representations in framing and pursuing our goals. This is not to say that consciousness could not have turned out to be epiphenomenal, just that it did not.

Accepting that conscious mental life has an evolutionary origin and is causally efficacious does not require accepting that all our conscious abilities, for example, our abilities to do calculus or philosophy, were selected by nature. The best bet is that almost all of our specific conscious abilities came as free-riders on capacities that were more directly selected for. Nor is it to say that a species very much like our own, but without consciousness, could not have been as successful or even more successful than us at conquering nature and proliferating, so long as its unconscious information-processing abilities were improvements on our own.

Consciousness and Cartesianism

The most famous and influential autophenomenological story is, of course, still the Cartesian story. Whereas James's account can withstand the sort of reflective equilibration I advocate — that is, the way things seem on James's story conflicts not at all with any claims we need, thus far, to make about the way things really are from the point of view of evolutionary theory or neuroscience — the Cartesian story cannot withstand the equilibration attempt. Cartesian intuitions about incorporeality must go because they conflict with the most fundamental ontological commitment of naturalism, and the intuitions about the unity of mind and privileged access must go because the evidence suggests that they are false. The doctrines of the unity of mind and of privileged access are worth considering in some detail since they, and the intuitions which support them, are largely responsible for the mystification of conscious mental life.

The doctrine of the unity of mind is more than the claim that conscious mental life hangs together. The doctrine, as I understand it, ascribes *omnipresence* to consciousness: it has access to all mental activity—it can get to the vicinity of all (truly) mental happenings. The doctrine of privileged access goes one step further and ascribes incorrigibility to this omnipresent force. It takes mere omnipresence and gives it the ingredient it needs for omniscience.

Indeed standard Cartesian doctrine has it that we always know exactly what *mental state* we are in, and in the case of the propositional attitudes, but not in the case of pains and moods, we also know the exact *representational contents* embedded in these states. It is possible to carry the privileged access doctrine beyond mental states and their contents to the domain of psychological science itself by claiming that one has reliable access to *how the mind actually operates*—to how it processes, retrieves, and stores information, to how it moves from one state to the next, and to how it activates the body. Descartes himself held something like this view: he thought that the power of each mental faculty—imagination, will, and understanding—as well as the relationship among them, is transparent once we pay attention. If one takes seriously the idea that we are omniscient with respect to our own minds then one has effectively bought the doctrine that psychology is a first-person exercise, a purely autophenomenological matter.

Some recent work in cognitive social psychology (see Nisbett and deCamp Wilson, 1977) indicates that many people are Cartesians in an even deeper way than Descartes. They believe that they occupy an epistemically privileged position with respect to the external causes of their mental states and behavior. The data indicate, however, that across a wide array of simple and emotionally unloaded tasks, individuals are mediocre at assessing the causes of their behavior and beliefs. The data indicate that the fact that these causes happen *to us* in no way licenses the conclusion that we are sensitive to these causes *as causes*, or indeed that we are sensitive to them at all. Most people, for example, show a decided preference for objects toward the right of center in an array of identical objects; but almost no one is sensitive to the causally efficacious role of right-most position.

The experimental data suggest a parallel conclusion with regard to Cartesianism about mental states, contents, and processes. The fact that these things *happen inside us* in no way licenses the conclusion that we have reliable access to them, indeed that we have access to them at all. Consider mental processes first. The very existence of an experimental cognitive psychology is predicated on the assumption that our intuitions are unreliable in this domain. The interesting thing from an epistemological point of view is that people are often extraordinarily confident that they do things, at the processing level, in ways they obviously do not do them. Consider this experiment by Sternberg (1966). Sternberg was interested in the processes whereby we retrieve recently encoded memories. Sternberg's procedure involved having subjects memorize lists of up

to six numbers. On each trial the subject saw a randomly generated list, for example, 2, 3, 7, 9. The list was visually displayed for just over one second. After a two second delay a test digit appeared, say, 3. The subject was to pull lever A if the test digit was on the memorized list, lever B if it was not. The dependent variable was the subjects' reaction times.

When asked how they solve, or would solve, this sort of problem, most people have powerful intuitions along either of the following two lines. Some subjects claim to visualize the entire list and just "look to see" whether the test digit is on it. Others claim with equal confidence that they move down the list serially from left to right until they find a match, or in cases where there is no match, all the way to the end of the list. If the first intuition is correct one would expect reaction times to be invariant no matter what the location of the digit on the list and no matter what the length of the list (so long as it was not too long to fit into the visual field of the "mind's eye"). If the second intuition is correct, reaction times should vary linearly with the location of the digit on the list.

Sternberg found that neither was the case. Instead he found that reaction time varied linearly with the length of the list but had no relation to the location of the test digit on the list. His conclusion was that we solve this problem by doing an exhaustive serial search, that is, we scan serially from left to right, but we do not stop the search as soon as we make a match. Instead, and completely counterintuitively, we note the match but then always proceed to the end of the list.

I do not at all mean to suggest that our intuitions about mental processing are always so far off. There are cases where we seem pretty much on target. The well-known work of Shepard and Metzler (1971) indicates that subjects solve some geometrical problems more or less as they say they do. Subjects claim to test for congruence between two figures at different angular orientations by mentally rotating the leftmost figure in the direction of the rightmost figure and testing for a match. The data seem to bear these intuitions out. Reaction times are a linear function of the degree of rotation. That is, figures rotated 180 degrees take twice as long to match as figures rotated 90 degrees.

There is also a third class of cases, namely, cases where there is obviously cognitive processing going on, for example, in linguistic understanding, but where we have no intuitions one way or another as to how we are doing what we are doing. This third class is undoubtedly the largest of all.

The doctrine that we are omniscient with regard to occurrent mental states and their associated contents also runs afoul of the data. Consider first that despite feelings of infallibility regarding such states, hindsight sometimes makes such powerful convictions yield. What was initially experienced and reported as a *fear that p* sometimes turns out in retrospect to look as if it was actually a *desire that p*. Such overriding, of course, only makes sense if we are sometimes wrong about what mental state we are in. Actually the word "in" here is ambiguous. We need not think that such overriding, when it occurs, involves

overriding the initial awareness *qua* awareness or the initial report *qua* autophenomenological report. The naturalist can allow incorrigibility about what is *in awareness*. What he or she denies is that what is in awareness tells us which among the many states we are in is most salient from a psychological point of view. That is, what the naturalist denies is that the state in awareness is the state we are *really-in*, where "*really-in*" refers to the state which is most plausibly seen as driving the whole system at the time the awareness occurs or the report is made.

I take it that one of the main functions of voluntarily self-monitoring our mental states as well as performing autophenomenological speech acts is to keep track of and apprise others of what states are currently prominent causally, that is, what states are figuring centrally in our overall psychological economy. We override when we seem to be mistaken about what state we are in in this rich causal sense. The popular idea that the mind, far from being a simple unity is, as they say, "a massively parallel processor" helps explain how we could be mistaken in the relevant sense. Relatedly, Dennett's (1978) distinction between "computational access" and "personal access" calls our attention to the fact that major portions of the cognitive system can be operating on, and in that sense have access to some mental state without the whole person also having access to that state.

The point I am trying to make can be brought out more clearly if we focus again on some experimental findings. Many people, it turns out, are fast shadowers; they can repeat what they are hearing almost as fast as it is uttered, with a latency of only 250 milliseconds. Fast shadowers, however, can almost never consciously summarize what they were talking about while shadowing. Indeed they deny doing any comprehending during shadowing. Nonetheless, fast shadowers show good comprehension on imaginative tests designed to detect it (Marslen-Wilson, 1973). It looks as if semantic processing is first of all, not itself necessarily a conscious process, second, not directly linked to consciousness, and third, sometimes extremely resistant to conscious attempts to get at its products. To be sure, the subjects here believe they have comprehended nothing; that is the belief that springs to mind when queried; it is what is *in awareness*; but it does not accurately reflect what state the system is *really-in*.

Or consider this related experiment by Lackner and Garrett (1973). Subjects were divided into two groups. Both groups were instructed to attend to just one channel in a set of earphones. In the attended channel both groups of subjects heard the ambiguous sentence "The officer *put out* the lantern to signal the attack." In the unattended channel the first group heard sentences which, if understood, would help provide an unambiguous interpretation of the target sentence, for example, "The officer extinguished the lantern," while the second group heard irrelevant sentences, such as, "Spring is beautiful in South Carolina." Both groups could report with great accuracy what they heard in the attended channel but, as expected, neither group could report what they heard in the

unattended channel. Both groups of subjects were then given a test which required them to interpret the meaning of the target sentence. Members of the group which heard the semantically unrelated sentence in the unattended channel divided over the interpretations that the officer *put the lighted lantern outdoors* to signal the attack and the interpretation that he *snuffed it out* to signal the attack. Members of the group which had heard the semantically related sentence overwhelmingly preferred the latter interpretation.

The irresistible conclusion is that the sentence that occurred in the unattended channel, and that the subjects claimed not to know about, was not only acoustically processed, but was semantically processed as well. The noise in the unattended channel was processed as a meaningful mental content and it was causally relevant *qua* meaningful content to the interpretation the subjects provided on the posttest.

A more compartmentalized model of the mind provides insight into how an agent could fail to know about certain causally efficacious mental states he or she is in. In this model, the mind is not one undivided whole through which a pontifical and omniscient consciousness effortlessly courses. Instead the mind has some sort of modular structure; its various parts and activities are differentially penetrable by consciousness. In addition to accommodating such facts as are revealed in the experiments I have mentioned, this model also fits much better with evolutionary theory than it does with the theistic background theory that framed Descartes's vision. Evolutionary forces often operate more like Rube Goldberg than like an omniscient, omnipotent, and all loving God, producing workable but not necessarily optimally designed devices. Given what we know about the ways of nature there is every possibility that the mind is really—as Marvin Minsky puts it—something of a “kludge.”

Cognitive Science and the Multiplicity of Awarenesses

Let me now turn to the second strategy for hunting down this elusive phenomenon—or better, to the second phase of the hunt. Whereas the first phase is meant to provide a general understanding of consciousness, its place in the biological world, as well as some sense of its scope and limits, it has the disadvantage of all very global analyses, namely its generality. This second strategy is intended to solve the problem in a more fine-grained way. Beyond this virtue of greater specificity this second strategy also has the virtue of being less sanguine about the ultimate utility of concepts like “mind” and “consciousness” than the first.

Several philosophers of psychology have urged recently that the concept of consciousness is both too simplistic and too general to be of service in framing the generalizations of a mature science of the mind—either as explanans or explanandum (see Churchland, 1983; Dennett, 1969, 1983; Rey 1983). To be sure, the thing that makes talk about consciousness coherent in daily life is that

we have a reasonably good sense of the very general functional feature that binds events belonging to the category, namely, they are *experienced by persons*, or better, so as not to exclude other species, they are experienced by whole organisms. But to someone who wants a deep understanding of consciousness, an understanding which goes beyond folk psychology, citation of the shared functional feature will sound simplistic.

The way to move beyond the simplistic account is, I suggest, by directing one's attention to the more fine-grained, subordinate categories that make-up the vague superordinate category of "mind." The reason is simple: it is within the overall economy of mental life that conscious mental events figure and it is only within a theory of mental life as a whole that such events can be understood.

So, what are these subordinate categories which constitute "mind"? Well, in one important sense, we will not know for sure until the science of the mind is more fully developed than it is at present. The best—indeed the only—strategy is to let actual scientific practice dictate the subordinate categories. Traditional epistemology as well as scientific psychology tell us that for a complete theory of mind we will need among other things an account of each sensory modality—of vision, taste, audition, olfaction, and touch, of proprioception, of language, of the emotions, of reasoning, of volition, and of whatever else turns up. As a complete and detailed picture of each domain and of the relations among them emerges the roles and kinds of conscious awareness will also become clearer.

Ideally one wants to accomplish several tasks for each domain. First, one wants to give a functional account—a homuncular-computational account—of how the domain works. Second, one wants to provide a realization theory, a theory of how each capacity is realized in the nervous system. Third, one wants to say something about when, where, and how, if at all, awareness plays a role in each domain, and if it does play a role, one will want to say something about the kind of awareness it is. Fourth, one will want to talk about the ways all the different domains interact, about how information, both conscious and unconscious, is passed around the cognitive system.

Fodor makes clear the substantive assumption behind this second strategy when he recommends,

the view that many fundamentally different kinds of psychological mechanisms must be postulated in order to explain the facts of mental life. Faculty psychology takes seriously the apparent heterogeneity of the mental and is impressed by such *prima facie* differences as between, say, sensation and perception, volition and cognition, learning and remembering, or language and thought. Since, according to faculty psychologists, the mental causation of behavior typically involves the simultaneous activity of a variety of distinct psychological mechanisms, the best research strategy would seem to be divide and conquer: first study the intrinsic characteristics of each of the presumed faculties, then study the ways in which they interact. Viewed from the faculty psychologist's perspective, overt, observable behavior is an interaction effect *par excellence*. (1983, p. 1)

Consciousness itself is notable in its absence from both Fodor's list and mine—this despite the intuitive pull to think of consciousness as a faculty or as a

unified kind. Thinking that way is, however, a temptation to avoid. As we have seen, the main mistake of Cartesian inspired epistemology is to think that all mental activity is in fact experienced, that all mental activity is under constant conscious surveillance. An alternative model is one that relegates awarenesses of many different kinds to many different, often quite limited, functions, in many different kinds of systems, in many different cognitive domains.

Marr's (1982) important work on vision is a model of how the sort of fine-grained, domain by domain, analysis of mind might profitably proceed. Our retinas consist of about 160 million light receptors onto which an image is cast and then dispersed into a two-dimensional array of dots of varying intensities. A common assumption was that the only way to get from this impoverished two-dimensional dot array to what we actually see is if we bring acquired knowledge to the image and, as it were, enrich it. Indeed, this assumption that seeing is epistemically driven, dominated vision research until Marr came along. Although the idea of the retinal image being enriched is on target, the enrichment process is not nearly as epistemically orchestrated—not nearly as driven by acquired knowledge or attention—as previously believed.

According to Marr and his colleagues, visual processing consists of three main stages, with awareness occurring at the very end. First, the brain takes the retinal image and derives a two-dimensional "primal sketch" by computing where the light intensity changes from one set of dots to the next. This computation results in a rough delineation of the edges and contours of the external objects in the visual field. Second, the brain analyzes various saliencies including shading and motion and computes what Marr calls the $2\frac{1}{2}$ -D sketch. The $2\frac{1}{2}$ -D sketch makes "explicit the orientation and rough depth of the visible surfaces, and contours of discontinuities in a viewer centered coordinate frame" (Marr 1982, p. 37). The $2\frac{1}{2}$ -D sketch enriches the "primal sketch" but it gives no information about the object's appearance from any perspective other than the viewer's. Third, and finally, the brain performs the most exotic computation of all. It takes the $2\frac{1}{2}$ -D sketch and determines whether there is any line which, when drawn through it, establishes its "basic pattern of symmetry." Humans as well as all our mammalian relations have one principle line of symmetry running through the center of the face and body, as well as subsidiary lines running through the limbs. Amazingly, the brain takes the lines of symmetry it divines from the $2\frac{1}{2}$ -D sketch and transposes the contours it has discerned in the $2\frac{1}{2}$ -D sketch onto them, thus drawing the 3-D image we see—the image we are aware of. According to Marr, no previously acquired knowledge affects the enrichment of the retinal image on its way to becoming the 3-D image we see. Our first awareness is of an image entirely computed by a system all of whose powers are built-in by nature.

Marr's analysis is a good example of one which satisfies the desiderata I recommended above. First, it rigorously specifies the computational processes operating in the visual system. Second, the alleged computational principles mesh well with known facts about the eye and the visual cortex. Third, the

analysis tells us where awareness figures in the visual system considered in isolation, namely, it occurs at the point the 3-D image is seen — at the end of the whole process.

Whereas the visual system's job stops with this simple kind of seeing, this simple kind of awareness — call it "Marr-seeing" — the information contained in what we Marr-see rarely stays contained in the visual system. This brings me to the fourth desideratum. Eventually, we want an account of how the output from one system becomes input for other systems; and of how, if at all, awareness figures in these information transfers, and if it does figure we will want to know the kind of awareness it is and what role it plays.

With respect to vision this means we want to know the fate of the information contained in what we Marr-see. At present we know much less about the systems with which vision interacts than we do about vision itself, so we cannot say much that is rigorous about what happens once Marr-seeing occurs. But here are some familiar things that sometimes seem to happen once we see something. First we often recognize an object we see as an object of a certain kind. Such recognition comes in at least two forms. Imagine yourself reading while sitting in front of a bowl of apples, glancing up, and "mindlessly" reaching for an apple and taking a bite. Here you have done more than Marr-see, you have recognized what you saw as an apple, as more than a mere round-red-thing. Still it is well-known that one can recognize what one has seen in this sense without really noticing what has gone on. Other times we recognize what we see in some stronger, more vivid, sense; in a sense associated with being able to name what we see. The evidence suggests that both kinds of recognizing require an additional step beyond Marr-seeing. They require bringing acquired knowledge to bear and thus performing some kind of memory search. It is possible that the phenomenological differences in the two kinds of recognizing are rooted in differences in the kinds of memory search executed. That is, it may be that the memory search required for "mindless" recognition that what we see is an apple is not routed through lexical memory whereas "mindful" recognition is. But how exactly memory is organized, and whether or not it is unitary, is not yet known.

A second and still different fate for the output of the visual system can be brought out by further reflection on the apple-case. Not only did I recognize that what I saw was an apple as I reached for it, but some information about shape, roughly the information contained in what I Marr-saw, was passed onto the tactile-kinaesthetic system as I reached for it. How visual information is made available for motor activity is something we have no first-person access to, but it is of the greatest importance to the way sighted people move about the world.

Yet a third fate for the output of the visual system occurs when the recognized object is also a desired object but hard to get. Suppose the apple is suspended ten feet overhead rather than sitting in a bowl within reach. Here the visual

information will not simply be sent to get the fingers moving. It will need to go upstairs for some active working over, for some quite vivid problem-solving. No one has a clear idea of how such unencapsulated activities as problem-solving work. But there are active research programs (see Anderson, 1983) investigating such higher level central processes. The important point for now is simply this: As we follow information through the cognitive system we will discover that it can be routed in a variety of directions, serve a variety of functions, and be accompanied by a variety of kinds of awareness depending on how it is routed and the function it serves.

My hunch is that two things will be revealed as the second phase of the project—the phase of analyzing each cognitive domain and the relations among them—proceeds. First, although conscious mental events will appear again and again in the overall account of mind, their ratio relative to unconscious mental events will be extraordinarily low. Second, there will simply be too many different kinds of conscious mental events, with too many different causal roles, to think of them as tokens of a unified kind, let alone as comprising a single faculty.

Nagel and the Intractability of Consciousness

Finally, I want to return to the worries of Thomas Nagel with which I began. Nagel's argument for the intractability of consciousness proceeds in several steps. First, he claims that "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* that organism" (1979, p. 166). Second, he claims that if the naturalist's program is to succeed, the phenomenological features of mental life—"the something that it is like to *be*" features—must be given a naturalistic account. Third, he denies this possibility. Nagel tells us that the "reason is that every subjective phenomena is essentially connected with a single point of view, and it seems inevitable that an objective, physical theory will abandon that point of view" (1979, p. 167). Finally, he informs us that,

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. (1979, p. 172)

Although Nagel is right that it is partly constitutive of having a conscious mental life that there be something it is like to have it, something it is like *for* the organism itself, the conclusion which he draws from this fact is much too strong. Nagel is led astray it seems to me because of two implicit, misguided, and interconnected assumptions. The first source of trouble is Nagel's "intuitive realism," as Rorty (1982b) calls it—his naive tendency to think that his intuitions about the nature of mind have implications for the nature of mind. The other,

and I think equally important, source of trouble originates ironically in a certain positivistic picture Nagel has of the nature of science and by implication of the nature of philosophical analyses that take science seriously. This attitude comes out in Nagel's persistent worry that any naturalistic account will be unable to make room for the autophenomenological facts—for "the true character of experiences" (1979, p. 172).

Actually Nagel seems to have two different concerns about a naturalistic analysis. The first is that it will necessarily fail to *exhaustively analyze* consciousness, the second that it will fail to *capture* what consciousness is like from a particular point of view. The first claim is implausible, the second true, but inconsequential.

As I have been framing the naturalistic project there are a series of questions that will need to be answered and then brought into equilibrium with one another. One important part of the inquiry has to do with getting clear on whether there are any shared autophenomenological features of conscious mental life, whether, that is, there is anything it is like to be a member of our species. In trying to frame an answer to this question we will be somewhat less interested in what exactly things seem like for any particular individual than in the overlap among individuals. But this greater interest in the type than in the unique features of the tokens in no way implies that the naturalist doubts that there is something it is like to be each particular one of us. The issue here is simply one of interest-relativity. For obvious reasons, you, your loved-ones, and your therapist, will be much more interested in the fine-grained details of how exactly your inner life seems than will the framer of a general theory of mind.

In any case, once we get a fairly good picture of how conscious mental life seems we will want to see if the seeming features can be interpreted realistically. That is, we will want to see how the way things seem from the first-person point of view fit with data from other sources—from evolutionary theory, cognitive psychology, and neuroscience in particular. Nagel often suggests that this move will involve abandoning the subjective point of view and he claims that "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" (1979, p. 174).

But Nagel is playing a trick here, actually two tricks. First, there is nothing in the naturalist's approach which requires abandoning the subjective point of view as the source of the autophenomenology, nor as a rich source of data for hypothesis generation about what is, in fact, going on. Indeed part of the overall strategy I have been recommending involves a fine-grained attention to autophenomenological detail, partly on the supposition that there are many different kinds of awareness with many different causal roles—some but not all of which (the causal roles, that is) we have reliable access to.

Second, and relatedly, the claim that moving to the objective point of view "does not take us nearer to the real nature of the phenomenon; it takes us

farther away” — is deceptively ambiguous between two senses of “real nature.” If “real nature” is, as it appears, meant to refer to the way things seem to some particular person, then, of course, it is true that going more “objective” will take us farther away from the phenomenon. But the reason for this — for why it is a bad idea to have third parties report on how things seem for others — is something the naturalist can easily explain. It is because persons are uniquely causally well-connected to their own experiences. They, after all, *have them*. Furthermore, there is no deep mystery as to why this special causal relation obtains. It is simply that the organismic integrity of individuals grounds their special relation to how things seem to them. John Dewey put it best: “Given that consciousness exists at all, there is no mystery in its being connected with what it is connected with” (1922/1957, p. 62).

If, on the other hand, by “real nature” Nagel means what is really going-on in the cognitive system as a whole including whether conscious mental events are actually playing the causal role they seem to the agent to be playing, or whether they are physically realized or not, then the claim that going heterophenomenological will necessarily lead us astray is quite incredible; its only conceivable warrant is allegiance to the bankrupt version of the privileged access doctrine.

The important point is this: there is absolutely no reason why a naturalist cannot both acknowledge the existence of subjectivity and view getting an accurate description of it as part of the overall project of understanding human nature. Once this much is granted it is hard to see how a utopian naturalistic theory of mind could fail to provide an exhaustive analysis of consciousness. It will provide a rich autophenomenology, a theory of how the autophenomenology connects up with actual goings-on, a theory about how conscious mental events — taxonomized into many different classes of awareness — figure in the overall economy of mental life, a theory of how mental life evolved and thereby a theory of which features of mind are the result of direct selection and which features are free-riders, and finally it will provide a neuroscientific realization theory — a theory about how all the different kinds of mental events, conscious and unconscious, are realized in the nervous system.¹ It is hard to see how the analysis could be more exhaustive.

Here one might expect Nagel to shift ground and claim that no matter how well analyzed the phenomenon of consciousness is by such a (yet-to-be-developed) theory, the theory will fail fully to *capture* consciousness. Now there are several ways in which a theory that provides an exhaustive analysis of consciousness might nonetheless be said to fail to capture something important about consciousness. First, it might be charged that such a theory fails to capture what exactly conscious mental life is like for each individual person. Nagel’s continu-

¹Of course, it is not to be expected that the neurophysiological realization theory will map onto our psychological theory in anything like the perfect way envisaged by the type identity theory; and thus there is no expectation that the psychological account will reduce to or be replaced by the realization theory.

al mention of the way consciousness attaches essentially to a "single point of view" indicates that this bothers him. But here there is an easy response. Theorizing of the sort I have been recommending is not intended to capture what it is like to be each token person, but only to capture, in the sense of providing an analysis, of the type (or types): conscious mental life. Although the general analysis is not intended to do so, you of course are entitled to capture — indeed it is unavoidable — what it is like to be you.

There is a second and more perplexing sort of "failure to capture" charge. Recall that one part of a general naturalistic theory will be a theory of how the human nervous system works, a theory of how mental life is realized in us. The core assumption is that although mental states are relational states involving complex causal connections with the natural and social environment as well as with other mental states, they are, in the final analysis, tokened in the brain. But Nagel, and he is not alone, finds it unimaginable that a neurophysiological realization theory could reveal "the true character of experiences" (1979, p. 172), where by the latter phrase Nagel means how these experiences feel. Indeed it is this worry in particular which leads Nagel to the view that we at present have no conception how physicalism *could* be true.

Here there are several possible lines of response. On one reading Nagel makes the same sort of logical error as Descartes makes in arguing for mind-body dualism (see Flanagan, 1984, pp. 12-13). On this interpretation Nagel is to be read as arguing from the fact that he *recognizes* his experiences as elements of his true conscious self, but does not *recognize* his brain states as elements of his true conscious self to the conclusions that his experiences are not brain states, in which case he is committing an *intensionalist fallacy* (Churchland, 1985). Alternatively, Nagel can be read as offering a modal, extensionalist argument of the following form:

1. My mental states are knowable by me introspectively.
2. My brain states are not.
3. Therefore, my mental states are not my brain states.

Nagel claims in correspondence with Paul Churchland to prefer this argument to the fallacious intensionalist one. But here, as Churchland shows, Nagel begs the question in the second premise. If physicalism is true, that is, if "mental states are indeed identical with brain states, then it is really brain states that we have been introspecting all along" (1985, p. 21). Substitute "temperature" in the first premise and "mean molecular kinetic energy" in the second to get a feel for this possibility.

This is not to say that those of us who believe that all mental events, conscious and unconscious, are tokened in the brain believe that the theory that eventually explains *how* they are tokened will capture "the true character of the experiences" *qua* experiences. The whole idea that the qualitative feel of some experience should reveal itself in a theoretical description of how that experi-

ence is realized fails to acknowledge the abstract relation between any theory and the phenomena it accounts for. Even the autophenomenological part of the project I have recommended which, unlike the realization theory, is directly concerned with how things seem to the subject, is itself at one remove (namely, a linguistic remove) from the experiences themselves. But the naturalist is the first to accept that a particular realization will only be the experience for the agent who is causally connected to the realization in the right sort of way. Once again the biological integrity of the human body can account straightforwardly for the happy fact that we have our *own* experiences. But this can hardly be much comfort to Nagel in so far as it shows the coherence of a naturalistic account of the unique way experiences are "captured" by the subject of them.

In the final analysis, your experiences are yours alone, only you are in the right causal position to know exactly what they are like. Nothing could be more important with respect to how your life seems, and to how things go for you overall; but nothing could be less consequential with respect to the overall fate of the naturalistic picture of things.

References

- Anderson, J. (1983). *The architecture of cognition*. Cambridge: Harvard University Press.
- Churchland, P.M. (1985). Reduction, qualia, and the direct introspection of brain states. *Journal of Philosophy*, 82, 8-28.
- Churchland, P.S. (1983). Consciousness: The transmutation of a concept. *Pacific Philosophical Quarterly*, 64, 80-95.
- Dennett, D.C. (1969). *Contents and consciousness*. London: Routledge and Kegan Paul.
- Dennett, D.C. (1978). Towards a cognitive theory of consciousness. In *Brainstorms*. Montgometry, Vt: Bradford.
- Dennett, D.C. (1982). How to study consciousness empirically: Or nothing comes to mind. *Synthese*, 53, 159-180.
- Dennett, D.C. (1983). Intentional systems in cognitive ethology. *The Behavioral and Brain Sciences*, 6, 343-390.
- Descartes, R. (1975). *Discourse on method*. In E. Haldane and G.R.T. Ross (Eds.), *The collected philosophical works of Descartes* (Vol. 1). London: Cambridge University Press. (Original work published 1637)
- Dewey, J. (1957). *Human nature and conduct*. New York: Modern Library. (Original work published 1922)
- Flanagan, O. (1984). *The science of the mind*. Cambridge: Bradford/MIT Press.
- Fodor, J. (1983). *The modularity of mind*. Cambridge: Bradford/MIT Press.
- James, W. (1976). *Principles of psychology*. (3 Vols.). Cambridge: Harvard. (Original work published 1890)
- Jaynes, J. (1976). *The origin of consciousness in the breakdown of the bicameral mind*. Boston: Houghton Mifflin.
- Lackner, J. and Garrett, M. (1973). Resolving ambiguity: Effects of biasing context in the unattended ear. *Cognition*, 1, 359-372.
- Marr, D. (1982). *Vision*. San Francisco: Freeman.
- Marslen-Wilson, W. (1973). *Speech shadowing and speech perception*. Unpublished doctoral dissertation. Cambridge, Massachusetts: Institute of Technology.
- Nagel, T. (1979). What it is like to be a bat. In *Mortal Questions*, (pp. 165-180). New York: Cambridge.

- Nisbett, T. and deCamp Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 321-359.
- Rey, G. (1983). A reason for doubting the existence of consciousness. In R.J. Davidson, G.E. Schwartz, and D. Shapiro (Eds.), *Consciousness and self-regulation*, (Vol. 3, pp. 1-38). New York: Plenum.
- Rorty, R. (1982a). Contemporary philosophy of mind. *Synthese*, 53, 323-348.
- Rorty, R. (1982b). *Consequences of pragmatism*. Minneapolis: University of Minnesota Press.
- Shepard, R. and Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701-703.
- Sternberg, S. (1966). High speed scanning in human memory. *Science*, 153, 652-654.