

With Science in Mind

In his *Confessions*, Augustine lamented, “What, then, is time? If no one asks me, I know: if I wish to explain it to one who asks, I know not.” In this respect, consciousness is like time. If no one asks me what consciousness is, I know. To pay attention to something is to become conscious of it. Indeed, everything with which I can be familiar -- from the sound of your footsteps to my own daydreams -- can be an object of my consciousness. Yet, if I wish to explain consciousness to one who asks, I know not. I (and, I suspect, we) have little theoretical grip on the nature of consciousness. Not for nothing did Schopenhauer call the problem of consciousness ‘the world knot.’ What, then, is consciousness?

Rather than approaching this daunting question directly, I want to consider some slightly more tractable questions: Will we ever develop a scientific theory that could reasonably be called a ‘theory of consciousness?’ If not, why not? If so, what kind of theory would it be? Of course, whether or not there will be a scientific theory of consciousness depends in part on what should be counted as a scientific theory. This is a complicated issue in the philosophy of science, which I cannot hope to settle. I assume that, at a minimum, a scientific theory of consciousness must be “naturalistic” in the sense of not invoking immaterial souls or supernatural beings. But I do not assume that all science is physical science, nor that the term ‘science’ is an honorific for whatever is true; nothing is scientific unless it is validated by a relevant disciplinary community, but much is true without such validation. Without worrying too much about what ought to count as ‘scientific,’ let us consider the kinds of answers that have been given to the question--Will we develop a scientific theory of consciousness? I shall divide the positions on the issue of a scientific theory of consciousness into two groups: the pessimistic ones that deny that there will be a scientific theory of consciousness, and the optimistic ones that are sanguine about the prospects for a scientific theory of consciousness. Let’s begin with a sample of the pessimists, whose range of reasons for

skepticism about a theory of consciousness could not be more diverse.

1. The first pessimist is Thomas Nagel. There can be no theory of consciousness, says Thomas Nagel; theories study phenomena objectively, and consciousness is inherently subjective, forever out of reach of any objective science. If *x* is conscious, there is something that it is like to be *x*, and what it is like to be *x* is accessible only from *x*'s point of view. The subjective character of experience is "fully comprehensible from only one point of view." 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."¹ We could know everything there is to know about the strange sensory modality called 'echolocation,' everything about the physiology of bats, everything about how a bat will behave in various conditions, everything objective about bats and still not know what it is like to be a bat. As Nagel famously said, "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."² Physicalism may yet be true, but "we do not at present have any conception of how it might be true."³

2. The second pessimist is Colin McGinn. It is not that there is no naturalistic theory of consciousness in principle; rather such a theory would be beyond our grasp. Consciousness has a physical basis in the brain. There is some neural property (or set of properties) *P* that gives rise to consciousness, but it is beyond our comprehension what exactly *P* is and how *P* could make consciousness intelligible. The "nature of the psychophysical connection has a full and non-mysterious explanation in a certain science, but...this science is inaccessible to us as a matter of principle."⁴ McGinn offers two kinds of arguments. First, we could not identify *P*: neither perception of the brain, nor coherent methods of concept introduction, could lead us to *P*. *P* could not be discovered by

¹ Thomas Nagel, "What is it Like to be a Bat?" in *Mortal Questions* (Cambridge: Cambridge University Press, 1979): 174.

² Nagel, 165-6.

³ Nagel, 176.

⁴ Colin McGinn, *The Problem of Consciousness: Essays Towards a Resolution* (Oxford: Basil Blackwell, 1991): 17.

empirical studies of the brain: Since we do not need consciousness to explain the observed physical data provided by the brain, “we do not need the property [P] that explains consciousness.”⁵ P is cognitively closed to us--as the property of being an electron is cognitively closed to a monkey.

Second, even if P were not cognitively closed to us, we could not understand the theory that invoked P. A theory of consciousness would link certain kinds of experience (say, a bat’s experience) with some explanatory physiological property P. But knowledge of such a correlation would not help us grasp the subjective character of the bat’s experiences. In a sense, we could not even understand the theory; for the theory would assert a correlation between neurological properties P with bat-experience E. But we do not understand the correlation unless we understand the correlates, one of which is bat-experience E; and we do not understand E unless we grasp its subjective character.⁶ McGinn comments: “[O]ur concepts of consciousness just *are* inherently constrained by our own form of consciousness, so that any theory the understanding of which required us to transcend these constraints would *ipso facto* be inaccessible to us.”⁷

Despite his pessimism about our ability to understand how “technicolour phenomenology [can] arise from soggy grey matter,” however, I should add that McGinn is optimistic about “our hopes for removing the philosophical perplexity.”⁸ There is no *metaphysical* mystery about how the brain generates consciousness. “Consciousness, in short, must be a natural phenomenon, naturally arising from certain organizations of matter.”⁹ The appearance of mystery is “a mere artifact of our cognitive limitations.”¹⁰

3. A final reason for the conclusion that there will be no theory of consciousness is taken by Kathleen Wilkes. The phenomena that we call ‘conscious’ are simply too heterogeneous to form the domain of a theory. “[C]onsciousness does not pick out a natural kind, does not refer to the sort of thing that has a ‘nature’ appropriate for scientific

⁵ McGinn, 13.

⁶ This way of putting McGinn’s argument avoids Owen Flanagan’s charge that McGinn uses ‘grasps’ in two senses. See *Consciousness Reconsidered* (Cambridge MA: MIT/Bradford, 1992): 94-95.

⁷ McGinn, 9.

⁸ McGinn, 1, 16.

⁹ McGinn, 6, 18.

¹⁰ McGinn, 17.

study, or which can constitute a ‘joint’ into which nature is to be carved by the sciences.” “Nor,” she adds parenthetically, “do ‘carpet’ or ‘calendar.’”¹¹ There will be no theory of consciousness because “there is no ‘thing’ which is consciousness--no unitary or special capacity or state of mind.”¹² So, according to Wilkes, there will be no theory of consciousness, not because the property of being conscious is mysterious, but because there is no such property: phenomena that we call conscious just do not have anything theoretically interesting in common.

So, the reasons that the pessimists have for being skeptical about the prospects of a science of consciousness are quite varied. First, there is the claim that there is something in the nature of consciousness that is inherently inaccessible to scientific method. Second, there is the claim that there is something in the nature of physical properties that determine conscious experience that is beyond our ability to grasp. Finally, there is the claim that there is no nature of consciousness to be discovered; the phenomena that we call ‘conscious’ are a hodge-podge, too superficial and heterogeneous even to be explananda for a scientific theory. To put it metaphorically, some pessimists about a science of consciousness think that consciousness is too “deep” for science; others that it is too “shallow” for science.

Before turning to those who propose various scientific theories of consciousness, let me discuss a position that is difficult to locate definitively on either side of the line. The position is the one taken by David Ray Griffin, the convener of this conference, in his manuscript, *Unsnarling the World Knot*. Although he asks a series of questions about consciousness, he argues that (what has been understood to be) the *basic* problem of consciousness is conceptual, and that its solution is philosophical. One of his purposes is “to remove from the back of scientists a false problem with which they have been saddled by bad philosophy.” He goes on: “[M]ost scientists [studying consciousness] have been trying, among other things, to

¹¹ Kathleen V. Wilkes, *Real People: Personal Identity Without Thought Experiments* (Oxford: Clarendon Press,, 1993): 192-93.

¹² Wilkes, 173.

answer a question that is impossible in principle to answer. No amount of empirical research, no matter how brilliant, can answer a question based on conceptual confusion.”¹³ Griffin takes the question that has interested investigators of consciousness to be this: ‘How does conscious experience arise from insentient neurons?’ This question, he argues, has a false presupposition--namely that “consciousness has arisen out of insentient neurons” in the first place. If we don’t begin by positing a great divide in nature--with insentient neurons on one side and conscious experiences on the other--we won’t have an impossible ontological gap to bridge.

Griffin himself argues for what he calls ‘panexperientialism,’ according to which “the two basic features that we associate with mind--experience and spontaneity--[belong] to all the units of nature.”¹⁴ The arguments for panexperientialism rest on the apparent requirements of reason, not on the deliverances of science. Panexperientialism postulates essential properties, undetected by present-day physics, of the entities in the domain of physics; and it postulates these properties (e.g., sentience) not because of any anomalies discovered by research in physics, but because such properties are rationally satisfying. This indicates that panexperientialism is grounded in a metaphysics that is conceptually independent of scientific theory.

Although panexperientialism is not itself a scientific theory, its proponents take panexperientialism to clear the way for a scientific theory of consciousness. Having dismantled the question ‘How does conscious experience arise from insentient neurons?’ the panexperientialist may pose the question ‘How does conscious experience arise from sentient (but nonconscious) neurons?’ A panexperientialism may believe that scientific theory can be brought to bear on this latter question. However, I do not think that the question ‘How does conscious experience arise from sentient (but nonconscious)

¹³ David Ray Griffin, *Unsnarling the World Knot: Consciousness, Freedom, and the Mind/Body Problem*, typescript: 5-6.

¹⁴ Griffin, 70ff.

neurons?’ would admit of a scientific answer either--unless the entire edifice of the physical sciences were overthrown and replaced by new theories of the physical world which included the property of being sentient in their taxonomies. Since I think that the prospects of this are negligible, I think that the prospects of there being a scientific theory of consciousness that has anything to do with panexperientialism are equally negligible. This is no criticism of panexperientialism inasmuch as I do not think that all knowledge is scientific knowledge (in any significant sense of ‘scientific’) anyway. In fairness, I should say that proponents of panexperientialism see the matter in another way. They are optimistic about a scientific theory of consciousness, because their understanding of science differs from the one that I am assuming here. There is much more to be said on this issue, but it will have to wait for the discussion period.

Now turn to the optimists--those who look forward to a scientific theory of consciousness. The optimists display an equally diverse range of views.

5. At one extreme is the view that the problems of understanding consciousness are not philosophical at all, but are wholly neurophysiological. I take this, at least implicitly, to be the view of Francis Crick, author of *The Astonishing Hypothesis*. The astonishing hypothesis “is that ‘You,’ your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules.”¹⁵ Crick suggests that speaking of “a vivid internal picture of the external world” is “merely another way of talking about the behavior of neurons.”¹⁶

Crick is wary of those cognitive scientists who “construct models that use general ideas based on our commonsense understanding of the mind, which they express in engineering or computing terms.”¹⁷ Such high-level functionalist terms allow one to investigate “the information processed by the brain, and the computational processes the

¹⁵ Francis Crick, *The Astonishing Hypothesis* (New York: Charles Scribner’s Sons: 1994): 3.

¹⁶ Crick, 9.

¹⁷ Crick, 16.

brain performs on this information, without considering the neurological implementation of these processes.”¹⁸ This, Crick advises, is a mistake.

“Why not,” he asks, “look inside the black box and observe how its components behave?...When, eventually, we know in some detail how the brain works, then a high-level description (which is what functionalism is) may be a useful way to think about its overall behavior. Such ideas can always be checked for accuracy by using detailed information from lower levels, such as the cellular or molecular levels.”

Not only does neurophysiology trump higher-level considerations about the nature of consciousness, but also considerations at the engineering or information-processing level have no theoretical significance. “Our provisional high-level descriptions should be thought of as rough guides to help us unravel the intricate operations of the brain.”¹⁹ To ‘conceptualize consciousness,’ Crick says, it is “essential to think in terms of neurons, both their internal components and the intricate and unexpected ways they interact together.”²⁰ In his lack of concern about qualia (e.g., the redness of red) or about the nature of meaning or about how neurons encode meaning, Crick may seem to overlook what has puzzled others working toward a scientific theory of the mind. For on Crick’s view, consciousness is not even a high-level feature of the brain; rather the whole truth about consciousness resides at the level of the neuron and neuronal interaction.

6. A second advocate of a neurophysiological approach to consciousness is Daniel Dennett. Dennett is an extremely sophisticated philosopher, who thinks that the ultimate truth about consciousness will be revealed wholly by neurophysiology, but who, unlike Crick, at least develops a method for linking neurophysiology to questions about consciousness posed at higher levels--e.g., at the level of a subject’s reports of her own experience. I shall return to Dennett in greater detail shortly.

¹⁸ Crick, 18.

¹⁹ Crick, 18.

²⁰ Crick, 256.

7. While Dennett and Crick (and Paul and Patricia Churchland) look to neurophysiology as a theory of consciousness, Owen Flanagan takes a more ecumenical approach. According to Flanagan's view, there will be a science of consciousness, which will feature neuroscientific results prominently, but not exclusively. Flanagan seeks to bring the results of phenomenology, psychology and neuroscience into reflective equilibrium. Although Flanagan agrees with Wilkes that conscious phenomena are heterogeneous, "there is something that binds all the various concepts of consciousness and that cannot be given up. Conscious experience names the class of mental states or events that involve awareness. A conscious experience is a state such that there is something it is like to be in it."²¹

Flanagan sees the term 'consciousness' as a superordinate category term--like 'memory'--with subcategories that, taken together or individually, show no signs of being a hodge-podge. A theory of declarative or explicit long-term memory (memory for facts and events), a theory of nondeclarative or implicit long-term memory (memory of acquired skills and habits, classical conditioning, subliminal learning, and so on), and a theory of short-term memory all look quite different. Different types of memory have different neural substrates and are described by different psychological generalizations. But there is a sense in which there is a theory of memory that includes a classification scheme as well as these different psychological and neurophysiological generalization about the events classified. "Like the superordinate categories of perception, memory, and learning," Flanagan says, "the category of consciousness might play a coherent, nonarbitrary role in the science of the mind."²²

Perhaps all conscious experiences are subserved by brain states that share some micro property like 40-hertz oscillation patterns. Perhaps there is some neurophysiological property, or set of neurophysiological properties that are necessary and sufficient for all conscious experiences. "If different sets of brain properties subserve different types of conscious experiences, then there exist generalizations linking the

²¹ Owen Flanagan, *Consciousness Reconsidered* (Cambridge MA: MIT/Bradford Press, 1992): 31.

²² Flanagan, 214.

neural with the phenomenological.”²³ Indeed, we already have true counterfactual-supporting generalizations about consciousness: “Persons with damage to the speech centers, and in particular to the left brain interpreter, will have trouble generating a narrative model of the self;” “Persons with certain kinds of frontal-lobe damage will have trouble formulating plans and intentions; other kinds of frontal-lobe damage will obstruct links between consciously formulated action plans and actually carrying out the intentions.” And we may expect true, counterfactual-supporting generalizations at the purely psychological level as well: “[P]ersons with qualia of kind q do x in circumstances c, but persons without qualia q (who are otherwise identical) fail to do x in c.”²⁴

Flanagan concludes, “There must be truths about consciousness, since consciousness exists, is a natural phenomenon, and is in need of explanation. so there can be a theory of consciousness. What sort of unity the theory will possess and what interrelations it will have to other theories within the overall science of the mind we do not yet know. The best strategy is to get on with the hard work of providing the right fine-grained analysis of conscious mental life and see where it leads.”²⁵

Despite the diversity among the pessimists and the diversity among the optimists about a theory of consciousness, the members of each group share common characteristics that differ from those shared by members of the other group. I want to call attention to two closely related types of division between the pessimists and optimists about a science of the mind: the first concerns fundamental methodological commitments; the second concerns conditions of adequacy for a theory of consciousness. First, the pessimists and optimists differ in their methodological commitments: The pessimists about a science of consciousness think that we can know enough about what consciousness pre-theoretically, prior to scientific investigation, to conclude that there will be no science of consciousness. The optimists divide into the proponents of a neurophysiological theory of consciousness and the proponent of an interdisciplinary

²³ Flanagan, 216.

²⁴ Flanagan, 218.

²⁵ Flanagan, 220.

theory. The proponents of a neurophysiological theory hold that consciousness is whatever the appropriate science says it is, and that neurophysiology is the appropriate science; the advocates of an interdisciplinary theory of consciousness hold that it is obvious that consciousness exists as a natural phenomenon, and there can be some kind of theory of anything that exists as a natural phenomenon.

Second, with one exception, the pessimists and optimists differ in what they require of an adequate theory of consciousness. The second way in which the optimists and pessimists differ is not unrelated to the first. With the exception of Wilkes, who thinks that what we call 'conscious' is too heterogeneous to comprise a domain of a science, the pessimists demand that a theory of consciousness, in effect, "save the phenomena." (Wilkes is unimpressed with the putative phenomena to be saved.) If to be in a conscious state is for there to be something it is like to be in that state, then a theory of conscious states will have to capture (in some way) what it is like to be in that state. That is, for the typical pessimist, an adequate theory of consciousness would "permit us to understand how the subjective aspects of experience depend upon the brain--which is really the problem we are trying to solve."²⁶ The optimists about a theory of consciousness require no such thing. Provision of necessary and sufficient neurophysiological conditions for conscious states would surely suffice for a fully explanatory and predictive theory of consciousness. So, in the main, it is not the case that the optimists affirm what the pessimists deny.

From this apparent impasse, we have a choice: join the pessimists and simply give up on the notion of a theory of consciousness, or join the optimists and not require a theory of consciousness to convey phenomenal feels--what it feels like to be a bat, say--to those incapable of experiencing them. Rather than joining the pessimists at this stage of inquiry, however, it seems to me prudent to lower expectations for a theory of consciousness and to keep on plugging to see just what we come up with. Provisionally assuming that there can be a theory of consciousness in some sense, then, what would the theory look like?

²⁶ McGinn, 10.

As we have seen, one prominent idea is that a theory of consciousness will be a neurophysiological theory, where by ‘a neurophysiological theory,’ I mean a theory whose domain is events, states and processes in the brain and whose vocabulary is continuous with that of current neurophysiology. No one here, I suspect, disputes that neurophysiology will contribute to an understanding of consciousness, and of every other mental function. But the question is whether neurophysiology holds the key to understanding conscious phenomena. Will conscious phenomena be explained and predicted wholly by a neurophysiological theory? Using Dennett as an example, I shall argue that the answer is no: Even if we put aside the phenomenal feels and the subjective character of conscious experience generally, I do not think that neurophysiology can provide an adequate theory of consciousness.

I shall discuss Dennett rather than Crick, because Dennett is more explicit about how he thinks traditional questions about consciousness are connected to neurophysiology. Crick dismisses Dennett as “a philosopher who knows some psychology and also a little about the brain,” but who relies too heavily on psychological experiments, without indicating that “experimental confirmation by the methods of neuroscience is essential.”²⁷ I think that Crick here misses the boat. One may wonder just what about consciousness Crick thinks the methods of neuroscience would be confirming in the absence of psychological results. Moreover, since Dennett is explicit that the “design” features of anything at all are physically implemented, he surely supposes that the functional and engineering features of the brain are neurologically implemented.

To investigate consciousness empirically, Dennett proposes an ingenious method. The theorist begins with a sound tape of a subject, perhaps accompanied by an electroencephalograph; the taped sounds provide the raw data, from which a transcript is prepared. The transcriber distinguishes text from noise (e.g., the transcriber does not record a hiccup), and makes appropriate changes: if a phonetic transcription would read, say, ‘the ball moved from left to light,’ the transcriber would interpret the speaker to mean, ‘the ball moved from left to right.’ Then, the theorist interprets the transcript as a

²⁷ Crick, 282-3.

record of speech acts--"not mere pronunciations or recitations but assertions, questions, answers, promises, comments, requests for clarification, out-loud musings, self-admonitions."²⁸ Both transitions--from tape to "direct-quotation" transcript, and from "direct-quotation" transcript to interpreted text--require that we "treat the noise-emitter as an agent, indeed a rational agent, who harbors beliefs and desires and other mental states that exhibit intentionality or 'aboutness,' and whose actions can be explained (or predicted) on the basis of the content of these states."²⁹

The text thus interpreted is the subject's "heterophenomenological world," a third-personal description of the world as it seems to the subject. The subject's heterophenomenological world is "a stable, intersubjectively confirmable theoretical posit, having the same metaphysical status as, say, Sherlock Holmes's Lond or the world according to Garp."³⁰ Initially remaining agnostic about the existence of the heterophenomenological items in the subject's world, the theorist attempts to "relate" the heterophenomenological items to states and processes of the subject's brain.³¹ Whether or not the subject's sincere reports about conscious experience are true is to be determined by the "real goings-on in people's brains." Dennett elaborates:

My suggestion, then, is that if we were to find real goings-on in peoples brains that had enough of the 'defining' properties of the items that populate their heterophenomenological worlds, we could reasonably propose that we had discovered what they were really talking about -- even if they initially resisted the identifications. And if we discovered that the real goings-on bore only a minor resemblance to the heterophenomenological items, we could reasonably

²⁸ Daniel C. Dennett, *Consciousness Explained*, 76. In "How to Change Your Mind," in *Brainstorms: Philosophical Essays on Mind and Psychology* (Montgomery VT: Bradford Books, 1978), Dennett distinguishes between beliefs that beings without language can have and what he calls 'opinions,' which are more language-infected states. In *Consciousness Explained*, Dennett comments on the distinction: "While I will not presuppose familiarity with that distinction here, I do intend my claims to apply to both categories." p. 78.

²⁹ Dennett, *Consciousness Explained*, 76. Since Dennett's intentional-stance theory is independent of his theory of consciousness, I am not complaining about the theorist's exploitation of the resources of intentional-stance theory in investigating consciousness.

³⁰ Dennett, *Consciousness Explained*, 81.

³¹ Dennett, *Consciousness Explained*, 407.

declare that people were just mistaken in the beliefs they expressed, in spite of their sincerity.³²

Dennett is not just noting that people have false beliefs (everyone knows that), nor is he just pointing out that people are sometimes mistaken about what they believe. (There is ample evidence that one may believe that she opened the window because it was hot; while, in fact, she opened the window under hypnotic suggestion.) The point in this passage is a methodological one: whether or not one really had an experience that she reports as conscious is determined by what neuroscientists discover about the brain.

To investigate consciousness empirically, then, Dennett subjects heterophenomenological items to what I shall call **the brain-mapping test**: The accuracy of what a person says about her conscious life is to be tested, item by item, by what goes on in the brain. Neurophysiological observations confirm, disconfirm, or leave indeterminate the truth of the subject's reports of her conscious experience. In this way, according to Dennett, the theorist of consciousness seeks to discover "how heterophenomenological worlds map onto events in the brain..."³³ That is, the heterophenomenological texts contain descriptions of what the subject reports she was conscious of; but what she was really conscious of is determined by the neuroscientist.

What neurophysiologists will discover about the brain seems to me clearly an empirical matter, not to be decided by philosophers. However, as I understand the state of neurophysiology (and as Dennett argues), many of the items in our heterophenomenological worlds (that is, much that we report being conscious of) will not be verified by the brain-mapping test. But, methodologically, on Dennett's view, the brain-mapping test (which reveals the "real goings-on in people's brains") is the final arbiter of consciousness.

Just what sorts of things are the heterophenomenological items that are to be subject to the brain-mapping test? They include reported qualia, pains, mental images, certainly--but also more. We are putatively conscious, at least sometimes, of our beliefs.

³² Dennett, *Consciousness Explained*, 85.

³³ Dennett, *Consciousness Explained*, 81.

(There is one pre-theoretical sense in which *all* heterophenomenological reports are reports of beliefs that the subject is conscious of having (barring self-deception). This is the sense in which conscious beliefs are contrasted with tacit beliefs and with subconscious beliefs. In this sense, “Winters are long in Vermont,” if contained in a heterophenomenological text, would report a belief of which the subject is conscious, merely in virtue of the fact that it was reported. However, this pre-theoretical sense of ‘consciousness’ is clearly not Dennett’s target. For any heterophenomenological text is filled with reported beliefs; but Dennett wants to be agnostic about whether or not anything is conscious--until the relevant heterophenomenological items pass the brain-mapping test. Moreover, if reporting a belief sufficed for a subject to be conscious of that belief, then the brain-mapping test would be otiose.

What, then, are the heterophenomenological items that are to undergo the brain-mapping test? The answer, I think, is that they are mental episodes of which the subject claims to have been conscious at a particular time. So, heterophenomenological items relevant to studying consciousness will include reports of conscious episodes with intentional contents--e.g., “I suddenly realized that I left the oven on.” Since work on consciousness, including Dennett’s, tends to focus on qualia, pains and the like, perhaps Dennett’s theory is not supposed to apply to all putatively conscious phenomena, but only to those phenomena that do not have intentional contents. However, much of what one takes to be one’s conscious life consists of episodes with intentional content--she calls to mind her college graduation; she “relives” the events of the day; she weighs the pros and cons of moving; she does mental arithmetic to calculate the tip for the waitress. Hence, a theory whose domain excluded such episodes as even putatively conscious would seem to me to fall seriously short of an adequate theory of consciousness. So, the heterophenomenological items relevant to the study of consciousness, and hence subject to the brain-mapping test, should include all those of which the subject reports that she is conscious at some particular time.

Sometimes we report being conscious of our beliefs. So, a belief that the subject reports having been conscious of at a particular time is a putatively conscious belief that

is subject to the brain-mapping test. Suppose that Jane, selected as an experimental subject for Dennett's theory of consciousness, consents to wear a (fantastic) brain monitor that records neuronal activity. Suppose that Jane also undergoes extensive psychological testing and interviewing, so that her psychotherapist has a clear psychological profile of Jane. Jane's heterophenomenological text contained this: "At that moment, I realized that I believe that Hal was trying to embarrass me--although he later convinced me that I had been wrong." Suppose that at the moment at which it dawned on her (as Jane might put it) that she believed that Hal was trying to embarrass her, Jane's behavior changed. Until then, she had had the belief that Hal was trying to embarrass (the psychotherapist noted that she had been avoiding Hal altogether), but she had been unaware of having that belief. But, since Jane gets aggressive when she thinks that someone is trying to embarrass her, she turned confrontational when she became conscious of the belief (as she put it). So, looked at from a behavioral point of view, there was a moment at which Jane became aware of her belief, and that awareness accounted for her change in behavior toward Hal.

Now suppose that the neuroscientist, seeking to confirm or disconfirm her report, checked the record of the brain monitor for a brain state or process with which to identify the heterophenomenological item (reported by Jane as becoming aware that Hal was trying to embarrass her). There is an initial difficulty with even applying the brain-mapping test. Dennett suggests that "[i]n some regards, you could say that my theory identifies conscious experiences with information-bearing events in the brain--since that's all that's going on, and many of the brain events bear a striking resemblance to denizens of the heterophenomenological worlds of the subjects."³⁴ However, it is difficult to see how any brain events could bear a "striking resemblance" to Jane's awareness of her belief that Hal was trying to embarrass her. What neural states could even count as bearing a striking resemblance to what Jane reported when she said, "At that moment, I realized that I believed that Hal was trying to embarrass me"? Since Dennett provides no parameter of similarity between brain states and reports of beliefs, how would the theorist know what to look for? How would she know whether she had found one? But even if

³⁴ Dennett, *Consciousness Explained*, 459.

this problem about applying the brain-mapping test were solved, we would still be faced with the following prospect: For suppose that we had a standard of similarity, but that the neuroscientist could find no brain state or process with which to identify the putatively conscious belief.

On Dennett's approach, the neurophysiological evidence is taken to be decisive with respect to what is really conscious and what is not. So, when Jane's reported consciousness of her belief fails the brain-mapping test--there are no "real goings-on" in her brain "that had enough of the 'defining' properties" of the belief--Dennett's approach would dictate the conclusion that Jane did not really become conscious of the belief that Hal was trying to embarrass her. This would be an instance in which, in Dennett's words, "we could reasonably declare that people were just mistaken in the beliefs they expressed, in spite of their sincerity." However, in the case imagined, it is not just that Jane reported becoming conscious of her belief. The psychologist, who knows Jane well and was observing her change of behavior toward Hal, confirms Jane's report. Jane's behavior exhibits patterns the stability of which is independent of whether or not they map onto patterns detectable from a neurophysiological point of view. That is, all the behavioral evidence points one way, and the neurophysiological evidence the other. In such a case (which, I would predict, would be rather frequent), it would be a methodological mistake simply to let the neurophysiological evidence trump all other evidence. For if neurophysiological evidence trumped all other evidence, then we would have to conclude that, Jane's report and the massive behavioral evidence to the contrary, Jane was not aware of her belief that Hal was trying to embarrass her. But if the neurophysiological evidence did not trump all other evidence, then neurophysiology would not provide the fundamental evidential base of the theory of consciousness; evidence from a "higher-level" (i.e., less fundamental) source would supersede neurophysiological evidence--at least sometimes. In that case, neurophysiology alone would not stand as a theory of consciousness.

Perhaps someone will object that this example is too fine-grained. We may expect neuroscience to discover, perhaps, that brain events with 40-hertz oscillation patterns are

necessary for conscious events, but, the objector may insist, it is implausible to suppose that for each actual conscious event there is a discoverable brain event onto which the conscious event can plausibly be mapped, and with which the conscious event can ultimately be identified. Such an objection, I think, holds no water. For Dennett's brain-mapping test is, as we have seen, just this fine-grained.

Perhaps another neurophysiological theory of consciousness would not be so fine-grained as Dennett's. OK, but the less fine-grained a theory is, the less likely it is to be an adequate theory of consciousness. For conscious phenomena are extremely fine-grained; indeed, they are individuated in a fine-grained way: the difference between Jane's reflecting on her trip to Vienna and Jane's reflecting on her trip to Paris is marked, in consciousness, by the difference in content of her thoughts. In conscious experience we distinguish between seeing a movie and watching a video; we distinguish between becoming aware that we are getting a little grayer and becoming aware that we are getting a lot grayer. Perhaps we should not expect a neurophysiological theory to make as fine-grained distinctions as we do in conscious awareness. But to the extent that it does not, it seems to me to be less a theory of consciousness than a theory of the physical basis of consciousness. A theory of the physical basis of consciousness would be no mean feat, but it would be disappointing as a theory of consciousness if it could not account for the features of conscious phenomena that make one conscious episode different from another.

Of course, with 100 billion neurons, and with a larger number of possible neuronal connections than there are particles in the universe, there are enough physical differences in the brain, quantitatively speaking, to match any differences in conscious experience. That is not the point. The point concerns neurophysiological theory: Can we look forward to a neurophysiological theory that finds neurophysiologically significant differences that can account for--predict and explain--the apparent differences in conscious experience? Dennett thinks that we cannot; but, with his neurophysiological bent, as we have seen, he is willing to discount reports of conscious experience if not borne out by neurophysiology. I agree with Dennett that it is unlikely that neurophysiological theory will discover relevant significant differences that can account

for all the apparent differences in conscious experience. But I draw a rather different conclusion--namely, that neurophysiology alone is insufficient as a theory of consciousness.

Of course, it is an empirical matter just what commonalities are found in various states and processes in the brain, but my guess is this: *At best*, a neurophysiological theory will provide a necessary and sufficient condition for a conscious event or episode of a given modality. So, perhaps, there will be laws like this: Subject S has a conscious visual experience at t iff neurophysiological property P is instantiated by S at t, where neurophysiological property P is some property or combination of properties expressible in the vocabulary of neurophysiology. Perhaps neurophysiology could ascertain how pleasurable or painful the conscious experience was. But I doubt--and I'm willing to bet on this--that it will not tell us that the subject is "reliving" her dissertation defense. If I am right in my guess about the future of neurophysiology and if the contents of intentional conscious episodes are crucial to their being the episodes that they are, then neurophysiology will not provide a theory of consciousness.

Eliminativists about consciousness may still insist (with Dennett) that there is no more to consciousness than what neurophysiology reveals: If neurophysiologists discovered that "the real goings-on [in people's brains] bore only a minor resemblance to the heterophenomenological items," as Dennett said, "we could reasonably declare that people were just mistaken in the beliefs they expressed, in spite of their sincerity."³⁵ In effect, neurophysiology could be construed as providing an "error theory" of reports of consciousness. In order for a report of conscious experience to be true, it must have a "truth-maker"-- a fact about the brain in which its truth consists. Reports of conscious experience, on this view, are reports about the brain; if there is no fact about the brain for a report of consciousness to report, then the report is false. On this view, if Sally reports that she is conscious of a moving red dot at t, and there is no appropriate neurological correlate that could reasonably be described as the conscious experience of a moving red dot at t, then we should say that it is false that Sally is conscious of a moving red dot at t.

³⁵ Dennett, *Consciousness Explained*, 85.

Now this view of neurophysiology as an “error theory” of discourse about consciousness is, in my opinion, implausible for two reasons: First, it is mistaken about language. It is not simply a linguistic fact that reports of particular conscious experiences are reports of particular brain states. (On being told that there is no neurological correlate that “mirrors” her report of being conscious of a moving red dot at t , Sally--a competent speaker and a rational person--may still insist that she was conscious of a moving red dot at t all the same; and her insistence would not call into question her rationality or her competence as a speaker.) To say that reports of conscious experience are reports of brain states should be a *result* of a theory of consciousness, not a *constraint* on such a theory. If brain states that “mirror” reports of conscious experience are not to be found, we should deny that talk about conscious experience is talk about brain states, rather than to deny (as an “error theory” would) that reports of conscious experience are true.

The second reason that it is implausible to take neurophysiology as an “error theory” of discourse about consciousness is this: If neurophysiology is to reveal the whole truth about consciousness (and that is the premise for taking neurophysiology to falsify reports of conscious experience), then neurophysiology must give an account of the error that Sally makes when she sincerely and competently reports, wrongly, that she is conscious of a moving red dot at time t . There must be some neurophysiological account of her mistaken report. Is Sally conscious of something that she mistakes for an experience of a moving red dot at t ? If so, then neurophysiology owes us an account of the real character of what she is conscious of, and also an account of how something with that real character can be mistaken for an experience of a moving red dot. On the other hand, if it is not the case that Sally is conscious of something that she mistakes for an experience of a moving red dot, then neurophysiology owes us an account of how a brain state that is not something that she mistakes for an experience of a moving red dot can cause Sally sincerely to report (albeit erroneously) that she is consciously experiencing a moving red dot.

To put the point colloquially, an error theorist might say that it seems to Sally that she is consciously experiencing a moving red dot at *t*, but that neurophysiology proves her wrong. My question is how, from a neurophysiological point of view, can it even seem to Sally that she is consciously experiencing a moving red dot at *t*? What does the seeming consist in? From a neurophysiological point of view, there must be a neurophysiological correlate for the seeming; but if we cannot find a neurophysiological correlate for Sally's consciously experiencing a moving red dot at *t*, it is highly unlikely that we shall find a neurophysiological correlate for its seeming to Sally that she is consciously experiencing a moving red dot at *t*. But if neurophysiology is taken to be an "error theory" of discourse about consciousness, then it must either give some neurophysiological account of how it could *seem* to Sally that she is consciously experiencing a moving red dot at *t* when she is not, or deny that it even seems to Sally that she is consciously experiencing a moving red dot at *t*. Although each of these alternatives deserves further exploration, neither is ultimately credible to me. So, neurophysiology--that science whose domain is brain states described nonintentionally--is not going to be adequate as a theory of consciousness.

But there may yet be a theory of consciousness, and neurophysiology may provide a theory of the physical basis of consciousness in the form of general necessary conditions for conscious experience. I think that there is a tendency to suppose that if you distinguish between consciousness and its physical basis, you must be an ontological dualist. Not so. We distinguish between a famine and its physical basis, without fear of dualism. But a theory of its physical basis (which would include reference to drought, erosion, crops rotting in the fields, etc.) is manifestly not a theory of famine (which would also include reference to distribution of food, and social and political conditions). Or again: we distinguish international currency transfers from their physical basis. No one would look to a theory of the operation of computer chips as a theory of international currency transfer. This is an old story, but there is no pernicious dualism. Rather, a theory of the physical basis of a kind of intentional phenomenon generally is not a theory

of the phenomenon itself. And that includes conscious intentional phenomena. So, there is no risk of mind/body dualism in claiming that neurophysiology will not be a theory of consciousness. Psychology, and perhaps the social sciences as well, will be needed to explain conscious phenomena.

So, here are my predictions: (1) If there is a theory of consciousness, it will be an interdisciplinary science of the sort that Flanagan proposes. Since, in such a theory, neurophysiological considerations will not always override psychological and other considerations, a theory of consciousness will not be reductionistic. Nor will it be dualistic. (I take Cartesian dualism to be a scientific nonstarter.)

(2) If there is a theory of consciousness, it will be a theory in a fairly weak sense. It will have counterfactual-supporting generalizations; it will be predictive. But it will not display the kind of unity that we expect from the physical sciences. In particular, we will not know how all the relevant generalizations of the theory fit together. This is so, because some of the relevant generalizations will be intentional, and others neurophysiological, and the “real patterns” (to use Dennett’s term) captured by intentional generalizations do not map onto any discernible “real patterns” captured by neurophysiological generalizations.

We can see this in the case of intentional behavior. Suppose that there is a real pattern in Beth’s behavior: she votes Republican in national elections; she goes to school board meetings to protest increases in the budget; she sends money to a group whose goal is to privatize prisons and abolish welfare--all out of a belief that taxes are too high. This real pattern has a stability and predictability all its own, quite apart from whether or not it corresponds to any physical pattern--either of bodily motions or of brain states. That is, the reality of this real pattern does not depend on our finding any real patterns in Beth’s brain that mirror it in any way. The patterns of consciousness are just a special case of such real intentional patterns.

So, at this point, I am on the side of the optimists: there will be a theory (of sorts) of consciousness. But we began with the broader question: What, then, is consciousness? I can’t help thinking that there is something that it is like to have conscious experiences (in

contrast to whatever mental processing goes on when you are under general anesthesia) and this “what it is like” will elude linguistic expression, and hence will elude expression in the language of science. So, I am not closing the door on the pessimists either. Any theory of consciousness, no matter how predictive and explanatory, will leave out something that we want to understand. So, even if we have a theory of consciousness that predicts and explains until the cows come home, we may still ponder the question with which we began: What, then, is consciousness?

Lynne Rudder Baker
University of Massachusetts at Amherst
October 5, 1994