DENNETT'S *CONSCIOUSNESS EXPLAINED*: ITS CRITICS, AND THE CONTROVERSY OVER THE "TRUE NATURE" OF CONSCIOUSNESS

BENJAMIN NEWMAN
Swarthmore College

> Are zombies possible? They're not just possible, they're actual. We're all zombies. (Dennett *Consciousness Explained*, 406)

> It would be an act of desperate intellectual dishonesty to quote this assertion out of context! (footnote to the above*).*

IN *CONSCIOUSNESS EXPLAINED*, Daniel Dennett sets forth an extended sketch of what may, or may not, be an explanation of consciousness. The controversy regarding what, exactly, Dennett *does* explain, and what, if anything, he omits, arises both from profound and genuine disagreement with his views on the part of many contemporary philosophers and from ambiguities in the definitions of some key terms (e.g. "consciousness" and "explanation"). Between the fundamentally conflicting worldviews and behind the conceptual ambiguities, is there a real philosophical controversy to be found, or just a lot of empty rhetoric? I believe that there is a real and meaningful question lurking in the interplay between Dennett and his critics, and it will be my goal in this paper to bring that question to light and to frame it in such a way that it is clear what is at stake.

What is at stake? Dennett's approach to the problem of consciousness is firmly reductive and materialistic. He believes that the physical facts about the brain (and, to a lesser extent, about the context in which it finds itself) are sufficient to account for all of the facts about the mind, consciousness included. Some of Dennett's critics would have ended that last sentence with the words "consciousness excluded," and they do have a point. Dennett's account could be considered an eliminative one, a case of explaining consciousness *away.* He himself is somewhat equivocal on this point. If we understand "consciousness" to mean some special, ineffable, all-or-nothing property, we will find nothing like *that* in Dennett's explanation. If, on the other hand, we take "consciousness" to mean a sophisticated sensitivity to the world around us, a sense of self, and the having of a train of thought, Dennett's account does include *these*

things, and offers compelling answers to the questions of how they might be realized in the brain.

So, which is it to be? There is a fundamental disagreement between Dennett and his critics regarding what it is that must be explained. It will not do, Dennett ("Living on the Edge", 147) reminds us, to appeal to our shared intuition of what we mean when we say that one or another entity is or is not conscious. It may be uncontroversial to interpret such an attribution as meaning simply that the entity in question has whatever it is we have when we are conscious, but it is also obviously circular—Dennett (ibid.) argues that the question of whether or not something is conscious has "no *clear* pre-theoretical meaning," and that therefore we ought to develop a theory *first*, and see afterwards whether or not the theory itself sheds any light on such questions of definition. This is exactly what Dennett has done, drawing his conclusions on what consciousness is from his theory, rather than plugging them in beforehand. It is thus not surprising that his conclusions differ strikingly from the preconceived notions of what consciousness must be on which most philosophers rely.

The key difference between Dennett and his critics, as far as can be discerned in this preliminary analysis, seems to be a matter of definitions. It is not that Dennett holds to one particular definition of consciousness while his critics hold to one or several others. Rather, Dennett is willing to postpone the question of what consciousness is until the question of how consciousness works has been settled. It is to this latter question that *Consciousness Explained* really offers an answer, and it is to this answer that we now turn in greater detail.

Dennett's Explanation of Consciousness: Discoursing on Methods

Since the key disagreement highlighted by this paper is methodological, it is fitting that Dennett begins his explanation with a clear statement of his methodological choices. His first such choice is an absolute rejection of dualism. How is this decision a question of methodology? Dennett notes that most celebrated dualists do not attempt any explanation of how the nonphysical mind (or nonphysical mental properties of the brain) is supposed to work—they do not seem compelled to give an account of the mind's essential nature in a way that sheds any light on its function. While dualism does leave open the possibility of detailed empirical study of the mind, few avowed dualists are engaged in such research, and no dualist has a research program on the question of how the mind relates to the body—most declare quite openly that they think the question cannot be answered. While it might turn out that the dualists are right on this score, Dennett is perfectly justified in adopting the working assumption that his project is

possible. As far as he is concerned, "accepting dualism is giving up" (*Consciousness Explained* 37). Thus, in choosing to reject dualism, Dennett is making the methodological decision not to quit with an easy (but vacuous) answer whenever he runs into difficulties.

It is for essentially the same reason that Dennett also rejects *ab initio* the behaviorist tactic of "feigning anesthesia"—denying that states of conscious awareness exist. To say that consciousness does not exist is no more an explanation than to say that its real nature is introspectively obvious and consider the matter closed (as most dualists do). Thus, Dennett commits himself to the premise that there really are conscious states. "If I wish to deny the existence of some controversial feature of consciousness," he writes, "the burden falls on me to show that it is somehow illusory" (40). Since Dennett intends to thoroughly dismantle the commonsense view of consciousness to make room for his own, he will be taking up this particular burden of proof quite frequently. The omission of any manifest feature of our phenomenological world is a move that requires justification. On the other hand, Dennett reminds us firmly that some such features must be left out of a final explanation—an explanation of pain (for example) that ultimately appeals to pain's "intrinsic awfulness" is circular and question-begging.

How are we to put any trust in the data of introspection while rejecting such appeals to special properties? Dennett's answer to this question is the core of his methodological position. We cannot simply take subjects' introspective reports to be authoritative, nor can we make the "first-person-plural presumption" and assume that everyone agrees on the meanings of such reports. Introspection, Dennett claims, is never a matter of pure observation, untainted by theory. Thus, in order for us to approach introspective data as objectively as possible, a certain amount of interpretation (or de-interpretation) will be required. The form of interpretation Dennett suggests is *heterophenomenology.* In applying this method, we treat the subject's introspective report as a text, granting uncritically that what the subject says is what the subject really meant to say—without presuming that everything the subject says is true. In effect, the subject is given complete authority to create a fictional world, defined by the text. Just as it can be an objective fact that Sherlock Holmes resided at 221B Baker St., despite the great detective's fictitious nature, we can treat the subject's introspective report as establishing a set of objective facts about a fictional world. The contents of this world make up the subject's heterophenomenology, and they are objective down to the last detail.

Here, then, is a methodological answer to the question of what it is that must be explained. The heterophenomenological data are what

must be explained, and an explanation of them would take the form of a "drafts" theory of the mind. Our sense of a unified stream of consciousness, he claims, is dependent on the time scale at which we view ourselves. At a higher level of magnification, that stream breaks up into many parallel streams of content-fixing events, each carried out by one of the many specialized subsystems that make up the brain. Taken together, these content fixations form something like a narrative, except that this narrative has no canonical form—there are always multiple working drafts, all subject to constant editorial revision, stretching into the unwritten future. The only way a piece of content can become part of any "canonical" narrative is in response to a *probe*—a direct question to the effect of "What are you thinking of right now?"—and the particular content that emerges in response to such a probe is extremely sensitive to the probe's nature and timing. What we catch in the "stream of consciousness" depends a great deal on when we go fishing and how we bait the hook.

Since the contents of consciousness are, in the sense just articulated, indeterminate if we never bother to check them, the existence of anything like a unified stream of consciousness grows out of the habit of constantly checking, or self-probing—actually several habits, including, but not limited to, such familiar tricks as talking-to-oneself and its silent descendant, private deliberative thought. How do we acquire such habits? Some of them are built in, such as our insatiable curiosity, an innate endowment of most primates, while others are learned, but the greatest contributions, including self-talk, come from language and culture. Not only is language itself one of our most important habits of mind, but it makes possible the learning of such habits by deliberate training and cultural transmission. The ancient Greeks had amazing powers of memory—their storytellers had to be able to recite epics like the Odyssey from memory, something we would find very difficult. Their brains were no different from ours, so why are our memories so different? They used an advanced system of learned mnemonics, which we no longer teach each other because we no longer need them—when we want to remember something as long as a book, we just write it down. Thus, changes in culture can lead to changes in the mental landscape, by affecting both the necessity and the availability of mental habits.

At the level of these higher-level, self-probing activities, there finally is one, canonical inner narrative, although it lapses occasionally (whenever we forget to check), and it is meaningless to ask it for fine temporal distinctions shorter than the time scale of the activities which generate it. Thus, a serial, one-thing-at-a-time information process emerges from the activities of a massively parallel information process,

effectively creating a "virtual machine"—computer science lingo for a computer of one type implemented as a program on a computer of another type. This virtual machine greatly extends the capabilities of the brain by providing the organization required to embark on long-term projects (like philosophy term-papers), but it is not very efficient—that's why deliberate pondering, the kind of thought that resembles talking-to-oneself, is so slow compared to more hardwired mental activities like deciding to swat a mosquito. Besides being inefficient, the "Joycean machine" is also unreliable—easily distracted and sidetracked, often falling into ruts or repeating itself, and even repeating itself. If these pitfalls are to be avoided, the machine must have some way of regulating its own behavior. One solution is to make use of the already existing perceptual equipment, which has proven so very effective at keeping tabs on complex, unpredictable features of the environment. The system will just have one more complex, unpredictable, vitally important thing to keep track of itself.

In such a system, most of the mechanisms will be *ad hoc* jury-rigs, but most (including many of the jury-rigs) will also be wonderfully sophisticated. Over time, regulatory systems will gradually adapt to the quirks of the systems they regulate, ultimately giving the virtual machine as a whole an effortless familiarity with its own idiosyncrasies. Explicitly keeping track of its own internal states as well as the world around it, organizing and directing the activities of the multiple parallel processes out of which it emerges, and intimately familiar with the quirks of its own activities in a private way that no similar but distinct system could ever be, this virtual machine could account for all of the activities of a human mind. "But is it conscious?" we ask. "YES," Dennett declares boldly, "my theory is a theory of consciousness. Anyone or anything that has such a virtual machine as its control system is conscious in the fullest sense ... *because* it has such a virtual machine" (281). This is the essence of Dennett's explanation, which he further develops by using it to work through some of the thornier philosophical problems that have surrounded consciousness. Although it is on his answers to these problems that Dennett has so often been taken to task by his critics, I will not be organizing my critique of the opposition around the problems themselves, which include zombies, qualia, and the self. This is because the problems overlap, and many of the same (or deeply similar) arguments are used across all of them. For this reason, the next section is organized around the forms of argument used by Dennett's critics, and it is within this context that we will encounter the problems themselves.

The Other Side: What's Wrong with Dualism?

John Foster ("Dennett's Rejection of Dualism") directly attacks Dennett's admittedly *a priori* rejection of dualism. His characterization of Dennett's own reasons for this move is quite apt—on the one hand, to accept dualism is to give up on the search for explanations, while on the other, any meaningful dualistic view violates the closure of the physical universe. Foster's answer to the first point is that even explanations in terms of physical laws must "bottom out" and remain in terms of some entities which are taken to be fundamental. Why should we ask more of the dualist? Dennett answers ("Living on the Edge" 141) that he is not asking more—merely asking for *any* explanation of a mental phenomenon in dualist terms that is not vacuous or question-begging. It is one thing to say that fundamental physical particles just are the way they are, since "the way that they are" is simple and easily described. It is quite another thing to say "It just is that way" of such a complex and subtly variable entity as the human mind, especially since the dualist rarely goes to any great lengths in describing this "way".

Foster's answer to the second point is more troubling—he simply rejects that the physical universe is a closed system (25). Why, he asks, should we simply assume that brains obey the same physical laws as the rest of the universe, until we have reconfirmed all of those laws with living brain tissue? He agrees that, *ceteris paribus*, this is a fine working assumption, but he denies the c. p. condition—other things are *not* equal, for we have all of the traditional antimaterialist arguments weighing in on the side of dualism. Until these arguments are roundly refuted, we do have good cause to reject the hypothesis that brains obey the laws of physics. My answer to Foster's argument here, which Dennett notes only obliquely (*Consciousness Explained* ch. 7), is that such a difference would be a physical difference—everything in the universe except for human brains obeys the laws of physics. If this is to be the case, then correctly stated physical laws must take this into account, rather in the manner of "$E = mc^2$, except inside living human brains." It then becomes valid to ask how a physical particle would *know* that it were inside a human brain and should therefore violate the laws of physics, given the fact that a brain is a complicated structure which is not trivial to recognize. It is clear that Foster does not intend to furnish us with an answer to this question, or even to investigate it, thus reinforcing Dennett's first objection to dualism—the absence of a research program.

It's Just Obvious!

While Searle denies any affinity with antimaterialism, and Sprigge ("A Disillusioned Zimbo") hints at preferring idealism to dualism, the

dualist flag will be waved, if not taken up outright, by many of Dennett's other critics. However, most of them do not aim directly at Dennett's rejection of dualism, as Foster does. While claiming that there is something inherent in the nature of mind which makes it invulnerable to materialist explanations like Dennett's, they seem to feel that it is not necessary resorting to argument in establishing these claims about the nature of the mind. Rather, they are content to let the matter rest upon our shared intuitions about the mind, as if all of the important facts were simply obvious, and everyone could simply agree on them. Dennett draws the reader's attention to this in the introduction to his chapter on methodology (67), calling it the "first-person plural presumption", and objecting to it on the grounds that there is still a great deal of controversy, even among those who claim that we can all agree on the nature of consciousness as approached through introspection. His answer to this, as mentioned before, is to deny that introspection is ever simply a matter of looking inside and seeing what's there—all introspection is construction, dependent on theory and subject to categorical as well as particular errors. His critics, of course, will have none of this.

John Searle says that it is not difficult to give a common-sense definition of consciousness, one that, not being analytic, will not "aim to analyze the underlying essence of [the] phenomenon," and to his credit, he does give such a definition: simply put, consciousness is what we have when we are not, in the vernacular sense, unconscious. As Searle admits, this definition does not cut to the heart of the matter, but from it he draws the following conclusions, entirely without argument: consciousness must be all-or-nothing, present or absent. It is an "inner, first-person, qualitative phenomenon." Searle offers neither evidence nor argument for any of these claimed features of consciousness, nor does he carefully define any of these adjectives. As he freely admits, he has not provided a definition which is pertinent to the analysis of the phenomenon in question, yet he proceeds to render a great deal of analysis from it. On what does he base his analysis? On what grounds does he claim to know what he is talking about when he uses the word "consciousness"? Searle seems to be claiming that we can uncritically pull important philosophical results directly from our shared, commonsense intuitions—in other words, that all of the important features of consciousness are simply obvious, to everyone.

Searle's conviction that the nature of the mind is obvious is demonstrated admirably in his most celebrated thought experiment, the Chinese Room ("Minds, Brains, and Programs"). Searle grants to his opponents that we have a computer program that passes the Turing Test (in Chinese). He then implements this program as a set of lookup

tables and ledgers, whose operations he will carry out by hand. He refers to the program itself as a "set of instructions" which he is "given", perhaps suggesting a relatively modest packet of papers with easily followed step-by-step instructions in English. As claimed, the program passes the Turing Test just as surely under Searle's implementation as before, so we might be tempted to say that it understands Chinese; but Searle protests. He himself doesn't understand Chinese—he never did—and there is *obviously* nowhere else in the system as he has described it for the understanding to reside.

Lest we claim that somehow it is the whole system which understands, Searle proposes to swallow the system. Simply put, he will memorize the program. It is at this point where Dennett and I both cry foul. Searle has pulled a fast one on his readers, for the program which he asked them to imagine is in all likelihood quite complicated, and now he asks them to imagine that he has got it all memorized. If we really go to all the trouble of imagining a program that passes the Turing Test, implemented on paper, we should probably imagine a large library, not a few small paper packets. Yet, despite this cheat of imagination, Searle claims that it is still *obvious* that the system does not understand, allowing a number of controversial premises to slip by without argument. While careful to separate such notions as understanding, intentionality, and consciousness, this pattern is typical of his reasoning on all three topics.

Another important philosopher who relies on the "obviousness" of his key premises is Thomas Nagel. In "What Is it Like to Be a Bat?" he claims that for an organism to be conscious, there must be "something it is like to *be* that organism—something it is like for the organism." He does not offer any argument for this position, nor does he bother to explain what he means by it in any detail. He does, however, go to some lengths to show that no materialist theory can accommodate this feature of consciousness—thus roundly refuting all (or all current) materialist theories, if we all agree on what he claims without argument or evidence, presumably on the basis that it is just obvious. After all, Nagel is quite willing to "assume we all believe that bats have experience" (393) before using them as a case study. Apparently he also assumes we all agree with his definition of what it means to be conscious. John Foster, whom we have already encountered in defense of dualism, makes a similar claim in his response to Dennett's demand that a dualist give some account of the nature of the nonphysical mind. "I do not myself see," he writes, "why the dualist needs to admit that there is anything more to the nature of the mind than what introspection can reveal" (1993, p.29). Dennett's answer has already been noted—that simply positing a mind with all

the properties "revealed" by introspection doesn't really explain anything, and leaves us with no recourse to any better-established theory when intuitions clash. This is, after all, the basic problem with accepting anything about the mind as "obvious"—unless we truly agree on what is obvious about the mind, we cannot expect to agree on any conclusions based on our "obvious" intuitions.

The Pangs of Regress

Another argument often used by Dennett's critics is that of the need for mental properties to have their origin in some source that embodies them intrinsically—if any narrative produced by a system is to have "real meaning", there must be a fount of intrinsic meaning somewhere in the system, for obviously no mere text, mere syntactic operation, or mere physical interaction could provide meaning (or understanding, or consciousness). A version of this view is expressed by Searle, as detailed above, and by Fellows & O'Hear ("Consciousness Avoided"). They write (in defense of Searle), "computer programs need minds to read them, hence minds cannot be computer programs merely" (82). Later, responding to Dennett's account of the self as a posited center of narrative gravity, they ask, "who it is who is assigning meaning to words and texts," insisting (and I would agree on this point) that "words and texts do not have meanings intrinsically." But they then go on to demand that, "on pain of circularity, Dennett cannot say that it is just another text ... which gives texts meaning. At least some texts [must] have the meaning conferring properties of selves and agents" (87). In other words, some entity must be the first one which has meaning, and thus it must possess its meaning intrinsically.

In Dennett's view (and mine), this is one of the most popular category mistakes in the field of philosophy. Granted, it doesn't make sense in ordinary English to speak of an entity as "almost" or "a little bit" conscious, or as "almost" having genuine semantic content, but this is not a sure guide to the nature of the phenomena in question—merely a sure guide to the limitations of our vernacular. As an escape from the dilemma of infinite regress or foundation in some intrinsic source, Dennett ("Living on the Edge" 144) proposes a finite, fuzzy regress, with the following example: A mammal's mother is a mammal, yet the history of mammals is finite. The solution to the seeming paradox is to recognize that the genealogy which begins with non-mammals and ends with mammals "can only be partitioned arbitrarily." Replace "mammal" with "genuinely meaning-bearing content-vehicle" in the argument, and Dennett's view on how meaning and consciousness can arise out of a meaningless and unconscious substrate becomes clear. I

would make the additional point in response to Fellows & O'Hear that circularity is to be hoped for in identifying the self as the interpreter of its own narrative, for if a self is a narrative, then it must, on pain of nonexistence, be circularly self-interpreting.

You Can't Tell by Looking

In this section I will address the related problems of qualia and zombies. It's often difficult to tell which of these problems an argument addresses, and this is not surprising, since their gist is the same—the traditional philosopher proposes that there are features of consciousness over and above the functional interplay of mental contents; thus, these features might be changed, or removed entirely, without affecting this functional interplay, and thus without affecting an agent's outward behavior *in any way.* Dennett and I believe that this notion is incoherent, but many of his critics still cling to it in one or both of its most popular forms—qualia and zombies.

Among the features of consciousness which materialists are most often accused of omitting, the raw, qualitative aspects, or *qualia*, of our states of awareness take pride of place. Almost every one of Dennett's critics accuses him of failing to do justice to qualia, and, to credit those critics, Dennett says outright that there are no such qualities (*Consciousness Explained* 372). By this he means that mental states do not have intrinsic, ineffable properties distinct from the functional roles they play, proposing instead that these functional dispositions are sufficient to account for all of the claimed facts about qualia, *including* their supposed ineffability. Many philosophers also claim that they can imagine "zombies", creatures that possess all of the functional attributes Dennett imagines—to the point of being physically or at least outwardly indistinguishable from conscious human beings—but without genuine consciousness. That the two claims are closely related is best indicated by the fact that qualia head the list of what zombies don't have. To Dennett's suggestion that *we* are zombies, and our consciousness as distinct from our functional underpinnings is an illusion, they ask who, if not some conscious self, is deceived by this illusion, echoing the arguments already noted regarding regress. However, my response to the issue of qualia and zombies will be different. It turns on the notion of evidence.

What does it mean to have evidence for something? Presumably, by "evidence for P" we mean information that, *ceteris paribus*, makes the truth of P more likely—thus, information whose availability to us is causally related to the truth of P. Also, when we say that we make some statement based on evidence for P, we mean that our utterance is caused by the availability of this information. Our definition of

evidence supports the following counterfactual: *ceteris paribus*, the removal of P causes the removal of evidence for P in at least some circumstances. Otherwise, the relation between the evidence for P and the probability of P does not hold. Thus, if we have made some utterance based on evidence for P, we cannot escape the counterfactual "if not for P, I would not have made that utterance." To deny the counterfactual is to deny that any of the causes of the utterance can truly be called evidence for P. Now, let the proposition P be "We are really conscious, and not just zombies who act exactly like conscious people in *every way.*" The counterfactual is clearly false in this case, for *by hypothesis* a zombie-philosopher will behave just like his conscious counterpart, right down to assertions about being conscious and not just a zombie. Thus, it is not the case that replacing a philosopher with a zombie could ever cause him to express different opinions about his own consciousness. Whatever is the cause of philosophers' views regarding the existence of consciousness in any sense that would make zombies coherently possible, it *cannot* meaningfully be called evidence. The same argument can be applied to statements about qualia.

The above argument is related to Richard Rorty's "Holism, Intrinsicality, and the Ambition of Transcendence" that makes an even stronger case against the proponents of intrinsic properties, asking whether the notion of intrinsic, non-relational properties is coherent at all. He suggests that not just selves but all objects are, so far as they figure in our discourse, nothing over and above centers of narrative (or at least descriptive) gravity. In other words, there is nothing essential to an object X beyond its being that of which most X-sentences are true. How could we assert otherwise? To do so would simply be to add another X-sentence to the list. If Dennett took this metaphysical thesis (attributed to Wittgenstein) as far as Rorty does, he wouldn't even have to call his heterophenomenological entities "mere fictions," for if *everything* is just a center of descriptive gravity, the whole fact-fiction, real-intentional distinction dissolves entirely. Unfortunately (in Rorty's view), Dennett does not go this far—he still clings to the possibility of ontology as distinct from epistemology. However, Dennett does go far enough to be able to claim that non-relational properties make no sense. His critics, Rorty says, disagree with him on this deep, metaphilosophical level—they (Rorty singles out Searle and Nagel) are all believers in intrinsic properties. Certainly, Searle and Nagel say as much themselves. However, Rorty also chides Dennett for claiming heterophenomenology as neutral ground in this debate, for there is no neutral ground—to claim that the aspects of a phenomenon that can be talked about are all of the important aspects is to beg the question

against the possibility of intrinsic properties that might lie beyond the scope of language. The gap separating Dennett from the likes of Searle and Nagel is one that can only be crossed by a leap of faith, not by argument. Still, Rorty has secured one important point for Dennett—the essentialists cannot prevail by force of argument either, and thus, while there might be reasons for believing (or not believing) in qualia, there can be no *evidence* either way.

All of this is emphatically not to deny that there are conscious states or that they have what could be called qualities—only to deny that these features are distinct from the role that these states play in the operation of the organism. One confusion in Dennett's view seems to draw out the qualia-boosters in particular force. Dennett argues that there is no such thing as a real seeming apart from *judging* that something is the case, and that all mental contents are *propositional* contents. Many of his critics misunderstand his use of these terms. Sprigge, for example, imagines that Dennett's phenomenal world must consist of "brilliant verbal dialectics" (49) rather than of sensory qualities, and Lockwood objects that seeing cannot be like believing because "judgements ... are too anaemic, too high-level, too *intellectual* a duty for the *substance* of sensation and perception" ("Dennett's Mind" 69). Dennett is quick to respond by convicting these critics of yet another category mistake. We are, he reminds them, speaking of *propositions.* There is no limit to the amount of content one proposition—for example, that the view before my eyes looks just exactly like {representation-of-a-visual-scene} is a perfectly good proposition. We are not speaking of *sentences, so* it will not do to object that "seeing is like pictures and believing is like sentences, so since a picture is worth a thousand words, seeing cannot be believing" (49-150). Perhaps there is a principled reason to doubt that perception is more than preconscious judgements made in various content-streams, but that judgements are not *rich* enough, or cannot carry *enough* content, is not such a reason.

What's Really Going On

In my introduction, I suggested that the real division between Dennett and his critics is one of definitions—Dennett is content to find consciousness in his theory after the fact, while his critics begin with preconsidered definitions of consciousness. At first, this might seem to be a sensible way to go about explaining something. After all, how can you explain something if you can't even say what it is? In the preceding sections, I offered some vague hints to the effect that the problem with taking this approach to consciousness is really a problem with language. I will now clarify this notion. It makes sense to

describe non-mental phenomena before attempting to explain them because we describe them and explain them from the same, third-person point of view, and usually in much the same terms. With mental phenomena this is not the case, since the terms which our language gives us for describing consciousness carry with them many presuppositions, embodied both in their definitions and in their usage. That it seems *linguistically* awkward to speak of an entity as "almost conscious" is one example. To open our investigation of consciousness by describing the phenomenon to be explained in such loaded terms is to beg many of the most important questions, perhaps without even realizing it. For example, in admitting a "commonsense" definition of consciousness, Searle begs the question of whether or not consciousness is an all-or-nothing phenomenon. He does so by adopting a language in which the sentence "This entity is a little bit conscious" sounds wrong because it is *ungrammatical.* Of course, question-begging is still possible even if such linguistic traps are avoided—Dennett admits to begging the question against dualism, after all—but beginning one's investigation by defining consciousness in any terms whatsoever, particularly in vernacular terms, is likely to result in at least some accidental question-begging. Most philosophers will agree that this is sloppy at best, and at worst forecloses many options and leads to ill-founded conclusions. If we are going to beg important questions, let us at least do so on purpose.

Of course, Dennett cannot help starting with a few metaphilosophical assumptions, and thus begging a few important questions (as Rorty notes), but he does his best to avoid this by not starting with any definition of consciousness. Instead, he focuses on the intersubjectively verifiable phenomena we normally associate with consciousness, developing a coherent theory that can explain these phenomena, and only when the model is complete does he see if it contains any hooks on which we can hang our prior notions about consciousness. Here are the hooks he chooses: a mental content is *conscious* if and only if it is available to be used in subsequent processing leading to its expression in overt (especially verbal) behavior. Since this criterion is not often well-defined at the time when a particular content is fixed, the fact of any content's status as conscious or not is indeed contingent on future cognition—there is no fact of the matter about which mental contents are conscious at any given moment. The qualities of a mental content, or its qualia, are exactly the sum total of all reactive dispositions that are inherent in the organism as a result of the presence of that particular content, and the pleasant or aversive nature of these qualities is due to the presence among these reactive dispositions of very ancient, primordial pursuit

and avoidance reactions. The *self* is an explanatory posit, a center of narrative gravity created by an organism (in this case, a human one) in the ongoing process of narrative-spinning that is higher cognition. The self is not, however, an illusion, for there are true facts that are most easily stated in terms of selves—for example, the very fact that some particular set of narratives, besides having all issued as speech sounds from the same orifice of the same human body, also cohere logically as autobiographical of a single entity.

As Dennett argues throughout *Consciousness Explained*, these identifications are not at odds with any facts about consciousness that can be supported by argument from intersubjectively uncontroversial premises. In this paper, I have tried to identify and call into question some of the premises underlying the overwhelmingly negative critical response to *Consciousness Explained* from within the philosophical community. These premises range from the viability of dualism to the existence of intrinsic properties, but by far the most common and pervasive is the simple claim that we *already know* what consciousness is, that we already possess a description of it, and that the job of an explanation is to use the data of empirical science to justify *this* preexisting description. It is thus enlightening that so many of Dennett's critics gleefully despair of ever finding such an explanation. In Dennett's view, and my own, that a particular description of a phenomenon guarantees that it will never be explained *disqualifies* that description, unless it can be argued conclusively that all descriptions guarantee this. Given the absence of such an argument (for none of Dennett's critics consider alternative *descriptions* of consciousness), Dennett is justified in proposing that an explanation will require a completely new description, but to adopt such a new description means letting go of some of our most cherished intuitions. It is this step that Dennett is willing to take, in defiance of his critics. Some critics claim that it is more noble, more heroic, and truer to human nature to hold out for some special essence that makes us conscious and human. Dennett asks, is it not more heroic to let go of our preconceived notions, and to dare to accept the consequences of our honest inquiry into the nature of consciousness? And I answer, Yes, it is.

References

"Symposium: Daniel C. Dennett's *Consciousness Explained.*" *Inquiry* 36 (1993): 3-160.

Dahlbom, B. Ed. *Dennett and His Critics.* Cambridge: Blackwell Publishers, 1993.

Dennett, D. *Consciousness Explained.* Boston: Little, Brown & Co., 1991.

-----. "Living on the Edge." *Inquiry* 36 (1993).

Fellows, R. and A. O'Hear. "Consciousness Avoided." *Inquiry* 36 (1993).

Foster, J. "Dennett's Rejection of Dualism." *Inquiry* 36 (1993).

Hofstadter, D. and D. Dennett. *The Mind's I: Fantasies and Reflections on Self and Soul*. New York: Bantam Books, 1981.

Lockwood, M. "Dennett's Mind." *Inquiry* 36 (1993).

Nagel, T. *Other Minds: Critical Essays 1969-1994*. Oxford: Oxford University Press, 1995.

-----. "What Is It Like to Be a Bat?" in Hofstadter, D. and D. Dennett. *The Mind's I* (1974).

Rorty, R. "Holism, Intrinsicality, and the Ambition of Transcendence." in Dahlbom, B. Ed. *Dennett and His Critics* (1993).

Searle, J. "Minds, Brains, and Programs." in Hofstadter, D. and D. Dennett. *The Mind's I* (1980).

Sprigge, T. "A Disillusioned Zimbo?" *Inquiry* 36 (1993).