Content Meets Consciousness

Lynne Rudder Baker
University of Massachusetts, Amherst

In *Consciousness Explained*, Daniel C. Dennett investigates consciousness from an empirical, third-personal point of view. The facts of consciousness are to be validated, not by a subject's own introspective authority, but by neurophysiology. Citing neurophysiological research, Dennett discredits one model of the brain ("the Cartesian Theater") and replaces that model with a more empirically adequate model of the brain ("the Multiple Drafts model"). Like most other investigators of consciousness, Dennett focuses on sensory phenomena that are putatively conscious. However, since much of what we report as conscious is intentional ("I just can't stop thinking about you"), and since Dennett's theory of intentionality makes no appeal to brain processes, the question arises: How can Dennett's neurophysiological method accommodate (putatively) conscious episodes with intentional content? How is Dennett's theory of consciousness, which identifies conscious states with particular brain events, related to his theory of content, which does not identify "contentful" states with brain events? Speaking of his overall project, Dennett says:

My fundamental strategy has always been the same: first, to develop an account of content that is independent of and more fundamental than consciousness—an account that treats equally of all unconscious content-fixation (in brains, in computers, in
Dennett goes on to comment that “this apparently shallow and instrumentalistic criterion of belief puts a severe constraint on the internal constitution of a genuine believer.” The severe constraint is that only systems with complex internal states turn out to satisfy the criterion, not that there is a one-to-one matchup between particular brain states and beliefs. Dennett thus disagrees with Fodor, whom Dennett characterizes as holding that “the pattern of belief must in the end be a pattern of structures in the brain, formulae written in the language of thought.”

As I see it, the difference between Dennett and Fodor is this: According to Fodor, whether a person believes that \( p \) is fixed by whether the person has a particular brain state that plays a particular causal role and that means that \( p \); whereas, according to Dennett, whether a person believes that \( p \) is determined wholly by whether or not a belief that \( p \) is predictively attributable to the person. The fact that only systems with complex internal states turn out to have beliefs is an interesting fact, but what makes a belief attribution true, on Dennett’s view, concerns only patterns of behavior, not any particular internal state. It is the absence of any attempt to identify particular beliefs with particular internal states that distinguishes Dennett’s intentional-stance theory from Fodor’s “industrial-strength realism.” So, according to Dennett’s theory of content, if we want to understand the contents of a person’s mental states, we see what the person does and says; we do not turn to neurophysiology.

The intentional stance, from the perspective of which a person has “contentful” states, is thus contrasted with the physical stance. From the physical stance, “if you want to predict the behavior of a system, determine its physical constitution (perhaps all the way down to the microphysical level) and the physical nature of the impingements upon it, and use your knowledge of the laws of physics to predict the outcome for any input.” Properties attributed from the physical stance are what I have called elsewhere ‘stance independent.’ That is, from the physical stance, we attribute properties that are instantiated independently of (the possibility of) anyone’s taking any particular stance toward them. On the other hand, from the intentional stance, we discern patterns that “are not out there entirely independent of us, since they are patterns composed partly of our own ‘subjective’ reactions to what is out there; they are the patterns made to order for our narcissistic concerns.”

As we have seen, according to intentional-stance theory, an entity has a “contentful” state only in virtue of someone’s predictive strategies. This is true not only of a person’s having a belief, as we have just seen, but of any entity’s having any “contentful” or intentional state. Since intentional-stance theory is Dennett’s only account of content—one that applies equally to brains, to computers, and to Mother Nature herself—an entity’s feature of having content is not a physical or stance-independent feature. Having
content is not an intrinsic property (and hence is unlike, say, the property of being constituted by H₂O); it is not a relational physical property (and hence is unlike, say, being a planet). Rather, having content is a stance-dependent feature—a feature that a system can have only in virtue of its (possibly) being the object of the intentional stance.¹⁷

Now turn to Dennett's theory of consciousness. Dennett aims to show how the various conscious phenomena "are all physical effects of the brain's activities."¹⁸ That is, he aims to give a physical-stance theory of consciousness. To this end, he uses neurophysiological research to impugn the model of consciousness as a "Cartesian Theater," where "a light-and-sound show is presented to a solitary but powerful audience, the Ego or Central Executive."¹⁹ Brain research suggests a better model, the "Multiple Drafts" model of consciousness, according to which consciousness is not a single narrative, with an author of record, but rather the gappy product of many processes of interpretation in the brain.²⁰

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 of the investigation, from which a transcript is prepared. The transcriber distinguishes text from noise (e.g., a hiccup) and makes appropriate changes: If a phonetic transcription would read, say, 'from right to light,' the transcriber would interpret the speaker to mean 'from left to right.' Then the theorist interprets the transcript as a record of speech acts—"not mere pronunciations or recitations but assertions, questions, answers, promises, comments, requests for clarification, out loud musings, self-admonitions."²¹

Both transitions—from tape to "direct-quotation" transcript (my term) and from "direct-quotation" transcript to interpreted text—require the theorist to adopt the intentional stance, as Dennett points out: "we must 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."²²

The text thus interpreted is the subject's "heterophenomenological world," a third-person 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 London or the world according to Garp."²³ 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.²⁴ 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." As Dennett puts it:

My suggestion, then, is that if we were to find real goings-on in people’s 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 declare that people were just mistaken in the beliefs they expressed, in spite of their sincerity.²⁵

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 that a self-attribution of conscious belief may be shown to be mistaken on the basis of what neuroscientists discover about the brain.

To investigate consciousness empirically, then, Dennett subjects heterophenomenological items to what I shall call the brain-mapping test: What a person says about her conscious life is to be tested, item by item, by what goes on in the brain. Neurophysiological discoveries confirm, disconfirm, or leave indeterminate the truth of the subject's reports of her conscious experience. In this way, the theorist of consciousness seeks to discover "how heterophenomenological worlds map onto events in the brain."²⁶

CONSCIOUS BELIEF

What sorts of things are the heterophenomenological items in question? The heterophenomenological items relevant to studying consciousness will include qualia, pains, and mental images, certainly—but also more. We are also putatively conscious, at least sometimes, of our beliefs. There is one pretheoretical sense in which all heterophenomenological reports are reports of beliefs that the subject is conscious of having (barring self-deception).²² 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 pretheoretical 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 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.
If Dennett wanted to maintain ties with this pretheoretical sense of 'consciousness' without jeopardizing the brain-mapping test, he might regard a reported belief as putatively conscious, and thus to be subjected to the brain-mapping test. However, I do not think that this is what Dennett has in mind either. For a heterophenomenology yields a whole world from the subject’s point of view. A heterophenomenology, like the text of The World According to Garp, contains much more than reports of introspective deliverances; it contains judgments about the world, not just reports of "inner life." If each heterophenomenological item were subjected to the brain-mapping test, then intentional-stance theory would be gratuitous. Thus, we should not suppose that a belief is putatively conscious (and hence susceptible to the brain-mapping test) simply because a subject reports it. Whether or not a heterophenomenological report such as "Winters are long in Vermont" is true is to be checked against the weather in Vermont, not the processes in the subject's brain; whether or not the subject believes that it is cold in Vermont is to be checked, from the intentional stance, against the subject's behavior, not against the subject's brain states.

So, I think that Dennett's object of study is neither consciousness in the sense that if S reports that p, then she is conscious of a belief that p, nor is it consciousness in the sense that if S reports a belief, then she is putatively conscious of a belief that p. What, then, are the heterophenomenological items relevant to the study of consciousness in Dennett's sense?

The answer, I think, is that the putative items of consciousness—the ones that are to undergo the brain-mapping test—are those reported in heterophenomenological texts of mental episodes of which the subject claims to be 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. Hence, a theory whose domain excluded such reports as even putatively conscious would seem to me to fail seriously short of an adequate theory of consciousness. 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 time. So, I assume that the heterophenomenological items relevant for studying consciousness include reports of conscious belief, conscious thinking, conscious judgments. These are the items that raise the question of the relation between intentional-stance theory and the theory of consciousness, and on these I shall focus.

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, which should be subject to the brain-mapping test. In the case of such putatively conscious belief, the object of consciousness is a belief. But, according to Dennett, whether or not a person has that belief is determined by whether the belief is predictively attributable on the basis of her (intentionally described) behavior. So, in order to bring intentional-stance theory in line with the brain-mapping test, there must be a brain event onto which the putative object of consciousness (the belief) can plausibly be mapped. But if the existence of conscious beliefs depends on whether beliefs can be mapped onto brain states, then Dennett's theory begins to look like Fodor's.

Given Dennett's two halves, the difference between a belief that is conscious and a belief with the same content that is not conscious is that the former can be mapped onto an event in the brain, whereas the latter may resist such mapping. But what makes any particular brain event a candidate for such a mapping? Dennett says that "many of the brain events bear a striking resemblance to denizens of the heterophenomenological worlds of the subjects." But as far as I see, Dennett has no parameter of similarity between brain events and attributions of belief that could constrain any such mapping.

An example of putatively conscious belief will further illustrate the tension between the theory of content and the theory of consciousness. Suppose that Jane's heterophenomenological text contained this: "At that moment, I realized that I believed 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 would put it on the tape that provided her heterophenomenological text) that she believed that Hal was trying to embarrass her, Jane's behavior changed. Until then, she had had the belief (the theorist noted that she had been avoiding Hal altogether) but had been unaware of it. 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 the intentional stance (the only stance from which beliefs are discernible in the first place), it is Jane's becoming aware of her belief, not just her having the belief, that accounts for her change of behavior.

Now suppose that the theorist, seeking to confirm or disconfirm her report, looked for a brain state or process with which to identify the heterophenomenological item (reported by Jane as becoming conscious of her belief that Hal was trying to embarrass her). In the absence of any parameter of similarity between brain states and attributions of belief, how would the theorist know whether she had found the "right" brain state? 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. In that case, I think that Dennett would have two alternatives: (i) stick with
the real patterns discerned from the intentional stance and give up the neurophysiological approach to consciousness; or (ii) stick with the neurophysiological approach to consciousness and deny that Jane became conscious of her belief that Hal was trying to embarrass her—despite the “real patterns” discerned from the intentional stance. Either alternative would leave half of Dennett’s project in the lurch.

CONTENT-FIXING IN THE BRAIN

So far, in discussing how the theory of content and the theory of consciousness do (or do not) fit together, I have considered intentional-stance theory in the context in which Dennett developed it: One considers human beings as rational agents and attributes attitudes that are predictive of their behavior. But in the context of Dennett’s study of consciousness, the theory of content must be put to another use as well.

Suppose that Eve’s heterophenomenological text contains this: “I was suddenly conscious of the fact that I was not alone in the house.” In so saying, she reported an episode with the propositional content that she was not alone in the house. Indeed, the propositional content—that she was not alone in the house—was essential to the episode; a (putatively) conscious episode with any other propositional content, or with no propositional content, would not have been the same episode. The brain-mapping test would check to see whether or not the reported episode could plausibly be mapped onto a brain event or process. If not, then, according to the theory of consciousness, it is false that she was suddenly conscious of the fact that she was not alone.

There are two possibilities: Either there are never any relevant brain events or processes onto which such heterophenomenological items can be mapped; or there are sometimes relevant brain events. If there are never any relevant brain events onto which such heterophenomenological items can be mapped, then, according to the theory of consciousness, no one is ever conscious of anything with propositional content; no one is ever conscious of any state of affairs. In such an eventuality, one could either give up consciousness as a phenomenon or give up the theory that has the unfortunate consequence that no one is ever conscious of any state of affairs. [Although I would not hesitate to give up the theory before giving up the (putative) phenomenon, I shall not press that point here.] So, more optimistically from the point of view of Dennett’s theory of consciousness, let us consider the possibility that there are brain events or processes onto which heterophenomenological items with propositional content can be mapped.

Now onto what kind of brain process is Eve’s consciousness that she was not alone to be mapped? Here is one suggestion: “a process that serves, over time, to insure a good fit between an entity’s internal information-bearing events and the entity’s capacity to express (some of) the information in those events in speech.” On this suggestion, the heterophenomenological item—Eve’s consciousness that she was not in the house alone—is to be mapped onto a brain process whose elements are information-bearing events.

What sort of information do these brain events bear? Dennett sometimes speaks of a different kind of content in the brain. The brain may have an “internal semantics”—referring to memory addresses, internal operations, other states of the system, and so forth, not to things and events in the outer world.” How does this kind of content get fixed? Is it fixed from the intentional stance or is it fixed physically, apart from anyone’s possible predictive strategies? At some points, Dennett sounds as if he is a staunch realist about this kind of content; but if he is, he needs an additional theory of content—presumably, a physical-stance theory of content for brain events. On the other hand, since intentional-stance theory is supposed to be a completely general theory of content that applies to brains as much as to persons, perhaps content is fixed from the intentional stance by discerning “real patterns” in the brain. In either case, questions arise: How would the internal semantics of brain states be related to ordinary propositional content (e.g., that Eve was not alone in the house)? And how does this internal content get fixed in the brain?

Dennett at least seems to be committed to there being brain processes onto which propositional content can be mapped. For he says: “What there is, really, is just various events of content-fixation occurring in various places at various times in the brain. These are nobody’s speech acts, and hence they don’t have to be in a language, but they are rather like speech acts; they have content, and they do have the effect of informing various processes with this content.” Perhaps there is a clue to content-fixation in the brain in Dennett’s discussion of the phi phenomenon—the phenomenon of subjects’ reliably reporting seeing a moving dot that changes color when presented with two stationery dots, one red and one green. Dennett says that “retrospectively the brain creates the content (the judgment) that there was intervening motion, and this content is then available to govern activity and leave its mark on memory.” (Sometimes Dennett contrasts what we judge to be the case with what we are conscious of. I am not concerned with such a contrast but with cases in which the subject is putatively conscious of her own judgment.) Dennett’s remark again suggests that brain events and processes have propositional content (that there is a moving dot) and that this content has causal powers (the content can govern activity and leave its mark on memory).

So, the content that Dennett attributes to the brain seems to be full-fledged propositional content, not just, say, reference to addresses of other brain states. This at least suggests that the theorist should seek to map Eve’s putatively conscious episode onto a brain process that has the content, “I am not alone in the house.” 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 heterogeneous phenomenological worlds of the subjects." However, it is difficult to see how any brain events could bear a "striking resemblance" to Eve's consciousness of not being alone in the house. What neural properties could even count as bearing a striking resemblance to what Eve reported when she said, "I was suddenly conscious of the fact that I was not alone in the house"?

The question, then, is this: How does content get fixed in the brain? Officially, Dennett's only basis for attributing content is his intentional-stance theory. For Dennett takes intentional-stance theory to be "an account of content that treats equally of all unconscious content-fixation (in brains, in computers, in evolution's 'recognition' of properties of selected designs)." On intentional-stance theory, a brain process has content only in relation to someone's (possible) predictive strategies. Officially, having content is not a physical or stance-independent property of anything. So, if we understand content-fixing in the brain from the point of view of intentional-stance theory, then not even the neurophysiological side of Dennett's theory of consciousness has left the intentional stance behind; like the Cartesian he criticizes, Dennett himself has not avoided "the lazy extrapolation of the intentional stance all the way in." Thus, he has not shown how hetero-phenomenological items even could map onto purely physical events in the brain; for the brain events onto which the hetero-phenomenological items map are themselves identifiable only from the intentional stance.

On the one hand, if Dennett retains his unified account of content in terms of intentional-stance theory, then the relevant brain events are the content-fixing events that are only relative to someone's (possible) predictive strategies. And if content-fixation is itself stance dependent, then either content-fixation is no part of neurophysiology or neurophysiology is itself an intentional-stance theory. Given the intentional-stance theory of content, events of content-fixation in the brain cannot be part of a wholly physicalistic theory.

On the other hand, if content-fixation in the brain is stance independent, then Dennett violates his intentional-stance theory of content; at the least, Dennett needs a separate account of content-fixation in the brain from his general theory of content. But even with a physicalistic theory of content-fixation in the brain, we would need to know how physically-fixed content is related to propositional content attributed from the intentional stance. We cannot turn to the would-be mapping for an answer; for the question is about the possibility of such a mapping in the first place: Which of the indefinitely many possible mappings onto brain states of hetero-phenomenological items with intentional content would be a correct mapping? In the absence of answers to such questions, the brain-mapping test is not an adequate test for conscious events with intentional content.

So, here is a dilemma: Either the neurophysiological theory of consciousness is itself a theory that essentially depends on the intentional-stance theory of content or it is not. If it is, then even "when we aspire to a science of the mind," we do not leave the intentional stance behind. If it is not, then we need a physicalistic account of content-fixation in the brain—an account that would push Dennett toward Fodor. Why has this dilemma gone unnoticed? Let me hazard a guess: Dennett takes intentional-stance theory to give him all the content he ever wants—for free, so to speak. So, he is not worried about the propriety, from a physicalistic point of view, of speaking of events of content-fixation in the brain. However, I do not think that things are so easy. For if intentional talk about the brain is just a manner of speaking, then how can content "leave a mark on memory," if memory is understood neurophysiologically? And if content does any work in neurophysiology, then we need a physicalistic theory of content—on pain of making neurophysiological features themselves stance dependent.

Let me conclude this section by pointing out another place where intentional-stance theory and the neurophysiological theory of consciousness tug in opposite directions. The Multiple Drafts model of the brain—according to which there is just continual revising and editing in the brain with no sharp demarcation between what is conscious and what is not—itself is formulable only from the intentional stance, not from the physical stance at all. Since editing and revising are intentional processes, it is only from the intentional stance that there is editing and revising in the brain. In responding to the charge that replacement of the Central Meander (of the Cartesian Theater model) by a Pandemonium of Homunculi (of the Multiple Drafts model) is simply replacement of one set of metaphors by another, Dennett says that "metaphors are not 'just' metaphors; metaphors are the tools of thought." I agree. But the metaphors in question are intentional-stance metaphors. Can Dennett's preferred model of the brain be expressed except from the intentional stance?

THE METAPHOR OF "DEPTH"

We have examined two places where the theory of content meets the theory of consciousness—putatively conscious belief and putative content-fixation in the brain—and have seen that Dennett's two halves do not sit comfortably on the same bench. In this section, I want to show that Dennett's physicalism, with its metaphor of "depth," precludes any rapprochement of the two theories.

Like many physicalists, Dennett is committed to a metaphor of depth. According to the depth metaphor, there are distinct levels of reality. In an inversion of the idea of the Great Chain of Being, what is genuinely real resides at the bottom level. And the bottom level of reality is physical; its
properties are stance independent. The intentional stance is introduced in contrast to the physical stance, from which one predicts and explains an entity's behavior on the basis of "its physical constitution (perhaps all the way down to the microphysical level) and the physical nature of the impingements upon it." Indeed, "if one wants to predict and explain the 'actual, empirical' behavior of believers, one must cease talking of belief, and descend to the design stance or physical stance for one's account." Physical-stance theories are thus deeper than intentional-stance theories. The intentional stance is only a resting place on the way to the 'lower,' more mechanistic stances from which stance-independent theories of genuine reality are formulated.

With this understanding of depth in mind, recall that Dennett's strategy is "first, to develop an account of content that is independent of and more fundamental than consciousness...and second, to build an account of consciousness on that foundation." Content is more fundamental than consciousness in that it is much more widespread: Entities can have contentful states without having consciousness, but (presumably) no entities can have consciousness without also having contentful states. Moreover, in the passage just quoted, Dennett aims to build an account of consciousness on the foundation of the theory of content.

Given the depth metaphor, how can content as Dennett understands it be a suitable foundation for a physicalistic account of consciousness? What the theory of content appeals to—patterns discernible only from the intentional stance—are, from a physicalistic point of view, less fundamental than what the theory of consciousness appeals to—neurophysiological processes that are presumably stance independent. So, given the metaphor of depth, it is difficult to see how Dennett's theory of content could be a foundation for the theory of consciousness—at least if the theory of consciousness is to be a physicalistic theory—when the features countenanced by Dennett's theory of content are much less fundamental than those countenanced by the theory of consciousness.

Let me try to be more explicit. Dennett may be committed to two different ways in which one theory may be more fundamental than another:

(i) Theory A is more fundamental than theory B if and only if theory B presupposes (in some sense) theory A and theory A does not presuppose theory B.

(ii) Theory A is more fundamental than theory B if and only if the properties, states, and entities that theory A refers to are more fundamental (are closer to the physical ground-level of reality) than those that theory B refers to.

We have seen ways in which Dennett's theory of consciousness presupposes his theory of content. Content has a crucial role not only in the construction of heterophenomenologies but also in the identification of brain events as events of content-fixation. That is, even the "neurophysiological side" of his theory presupposes the theory of content. So, according to (i), the theory of content is more fundamental than the theory of consciousness. Also, Dennett says that his theory of content is more fundamental than his theory of consciousness, and in sense (i), it is. However, there is another sense in which physicalists take one theory to be more fundamental than another: Physics is more fundamental than chemistry because the entities countenanced by physics (subatomic particles) are more fundamental than the entities countenanced by chemistry (molecules). In this sense—sense (ii)—Dennett's theory of consciousness is more fundamental than his theory of content: The properties, states, and entities countenanced by his theory of consciousness (brain states, etc.) are closer to the physical ground-level of reality than those countenanced by his theory of content (intentional patterns of behavior). That is, the two ways in which one theory may be more fundamental than another come apart for Dennett.

The problem is not only that (i) and (ii) come apart in Dennett's two theories. The further problem is that the depth metaphor of physicalism commits Dennett to giving precedence to (ii), according to which the theory of consciousness is really the more fundamental theory. According to the depth metaphor, stance-dependent features are always less fundamental than stance-independent features. If the neurophysiological theory of consciousness is stance independent, then it should be explicated in a way that does not presuppose the stance-dependent features of intentional-stance theory. But if the neurophysiological theory of consciousness is not stance independent, then it is not a purely physicalistic theory. For a theory cannot legitimately claim to be physicalistic (and stance independent) if it appeals to stance-dependent features that are irreducible as far as we know.

Dennett cashes out talk of Mother Nature's intentions in terms of the theory of natural selection, which does not presuppose unreduced intentional-stance features. Similarly, the theory of consciousness—if it is to be physicalistic—should provide a way to cash out talk of conscious states (including those with intentional content) without presupposing unreduced intentional-stance features. Dennett may respond that the beauty of intentional-stance theory is that it invokes no features that need to be reduced; it simply affords a convenient way to predict phenomena for which we have no strictly physical account. But in that case, a ground-level physicalistic theory must be formulable without appealing to features discernible only from the intentional stance. To the extent to which Dennett is less a realist about what is discerned from the intentional stance than about what is discerned from the physical stance, a genuinely physicalistic theory (of anything) cannot appeal to features discernible only from the intentional stance.

To sum up the difficulty, as I see it: According to the depth metaphor, Dennett's theory of consciousness and theory of content have domains at different levels of reality. The theory of consciousness is more physicalistic.
and thus (by (ii) above) the more fundamental theory. Yet, the theory of
consciousness presupposes the (less fundamental) theory of content. The
depth metaphor precludes an easy reconciliation of the two theories. For
what is “deeper” cannot presuppose what is “shallower.” So, the problem is
this: According to the depth metaphor, the theory of consciousness is more
fundamental than the theory of content; but the theory of consciousness
presupposes the theory of content. Therefore, according to the depth metaphor,
the more fundamental theory presupposes the less fundamental theory.

This problem would dry up if we dropped the physicalistic metaphor of
depth. For there would be nothing untoward about a neurophysiological
theory of consciousness that presupposes a theory of content—and hence is
the more fundamental theory in sense (i)—if one did not also hold the neuro-
physiological theory to be “deeper” than the theory of content that it presup-
poses. So, the metaphor of depth precludes reconciliation of Dennett’s theory
of content with his theory of consciousness: The deeper theory cannot
presuppose the shallower one, and Dennett’s theory of content cannot be a
foundation for a physicalistic account of consciousness. But if Dennett gave
up the depth metaphor, he would lose the motivation for the brain-mapping
test to determine what a subject is conscious of.

REAL PATTERNS

My proposal is that Dennett take the same approach to consciousness that he
does to content. Test heterophenomenological items, not against brain
states, but against their predictive attributability. That proposal rejects the
dichotomy according to which the facts of consciousness are established
either by introspective reports, taken to be authoritative, or by neurophysi-
ology; yet, it retains the empirical, third-personal approach to conscious-
ness. In this section, I want to suggest that his physicalism alone does not
commit Dennett to a neurophysiological approach to consciousness and that
Dennett’s discussion of “real patterns” gives him the resources to approach
consciousness in the same way that he approaches content.

In “Real Patterns,” Dennett discusses, among other things, the relation
between what is discerned from the intentional stance and what is discerned
from the physical stance. As a physicalist, Dennett takes what is discerned
from the physical stance to be what is “really there;” but Dennett also is a
‘mild realist’ about what is discerned from the intentional stance. Beliefs,
says Dennett, are as real as centers of gravity. This “mild realism is the
doctrine that makes the most sense when we are talking about real
patterns, such as the real patterns discernible from the intentional stance.”
Since we lack the time and often the means to make predictions of behavior
in terms of fundamental physics (which Dennett takes to be the real locus of
reality), we adopt the intentional stance as a basis for prediction. In the
interest of efficiency we “trade[ed] off reliability and accuracy of prediction
against computational tractability.” We know that we have discerned a “real
pattern” (and not just noise) when it is the basis for reliable predictions. And
since beliefs and desires are reliable predictors of intentional action, we
know that there are real intentional patterns. And, as I have emphasized, we
know this quite apart from any neurophysiological investigation: Beliefs and
desires would still be reliable predictors no matter how our brains turned out
to be organized. This is a central message of intentional-stance theory.

It is important to see that the “real patterns” of behavior discerned from
the intentional stance need not mirror any detectable patterns of physical
states. 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 does not “correspond to” any pattern—either of bodily
motions or of brain states—discernible from the physical stance. Indeed,
according to “Evolution, Error, and Intentionality,” transferring Beth to a
different environment could change her intentional states without changing
her physical states. But absence of correspondence between physical and
intentional patterns does not impugn our claim to have found a real pattern—
albeit a real intentional pattern.

In general, real patterns of a person’s intentional behavior do not mirror
real physical patterns of the person’s brain states or bodily motions. What
makes something the intentional action that it is, is often determined by
context: The same bodily motion (produced by the same type of brain state)
may be a vote in a faculty meeting, a request to be excused, or an attempt to
distract the speaker. As intentional-stance theory suggests, there is no more
reason to think that intentional behavioral patterns mirror neural patterns
that cause the relevant bodily motions than to think that the patterns of play
that win the U.S. Open in tennis mirror muscular motion in a player’s arms
and legs.

This important point may be obscured by Dennett’s emphasis on
examples like “bar code” and Life World as examples of “real patterns.”
Looking at these examples, one may suppose that a believer or agent is to her
internal physical states as “gliders” in the Life World are to arrays of pixels
that constitute them. Such an analogy would misfire: for the glider’s behav-
ioral patterns do mirror the physical (geometrical) patterns of the pixels that
constitute the glider together with the pixels in adjoining cells. But Beth’s
intentional behavioral patterns (voting for Republicans, say) do not mirror
physical patterns of her bodily movements (nor of her bodily movements
in conjunction with the physical motions of nearby objects). Indeed, we know of
no physical patterns whatever that are even candidates for mirroring patterns
of one’s voting behavior.
A second way in which I think that the Life World is misleading is that it suggests that understanding a bit-map (and the law that governs the "bits") gives a deeper understanding of the glider. But consider this: Bill Gates of Microsoft Corporation has bought up digital rights to many art works all over the world. Does a bit map afford deeper understanding of The Birth of Venus? Would we suppose that we can now understand the Mona Lisa's enigmatic smile because we can reproduce it digitally? Suppose that we found no bit-map-level patterns that mirrored differences between smiles and leers; would we use the bit map to deny the reality of the smile or to say that there was no fact of the matter about whether a painting represented a smile or a leer? No doubt neurophysiology can show us necessary conditions for consciousness; but it is a non sequitur to suppose that whatever "real patterns" there are in consciousness must be mirrored by "real patterns" in the brain—just as it would be a non sequitur to suppose that whatever "real patterns" there are in an artwork must be mirrored by "real patterns" in a bit map of it.

In his "Appendix A (for Philosophers)," in Consciousness Explained, Dennett responds to the charge that there is "a tension—if not an outright contradiction—between the two halves of [his] theory" by saying this:

The shock-absorbers that deal with the tension are the strained identifications of heterophenomenological items (as conceived under the traditional perspective [from which we treat people as single-minded agents]) with events of content-fixation in the brain (as conceived under the new perspective [from which we break the single-minded agent down into miniagents and microagents with no single Boss]).

Far from being shock-absorbers, "the strained identifications of heterophenomenological items ... with events of content-fixation in the brain" seem to me to expose the tension that they are supposed to overcome. From the perspective of neurophysiology, there are no unified agents. However, from the intentional stance, of course there are unified agents; intentional-stance theory was designed specifically to accommodate the "real patterns" discernible when we think of each other as unified agents. From the intentional stance, the question of whether there is a Boss neuron is simply irrelevant to whether a system is a unified agent. The fact that there are no unified agents discernible from the physical stance is wholly unsurprising: neither are beliefs, desires, or plans discernible from the physical stance. Why should the existence of unified agents any more depend on neurophysiological facts than the existence of beliefs? Yet, Dennett is a "mild realist" about beliefs. I suggest that he assume the same position about agents. Indeed, from the intentional stance, agents and believers are treated in exactly the same way, as they should be.

Here is my suggestion about consciousness: Look at consciousness from the intentional stance; as we have seen, intentionality infects Dennett's theory of consciousness anyway. Suppose that Jack reports becoming conscious of a belief that has been induced by advertising. The transcript for his heterophenomenological text contains this: "I suddenly saw that I had believed that Wisk got clothes cleaner than Tide. When I realized how I had been taken in, I stopped buying Wisk and went back to Tide." Here is a report of a conscious belief, and one the consciousness of which changed Jack's behavior. The "real patterns" here are real intentional patterns. The stability of these patterns is independent of whether or not they map onto brain patterns. If such patterns are reliable bases for prediction, then they are real (mildly real?) regardless of the outcome of the brain-mapping test.

Before publication of Consciousness Explained, Dennett's major insight (in my opinion) was that intentional patterns are "real patterns" and that real intentional patterns do not mirror patterns of brain processes. Why does he turn his back on his own insight when he considers consciousness? Perhaps Dennett would take the answer to be obvious: In the absence of immaterial souls, an account of consciousness must be in terms of brain processes. Early on, Dennett remarks that his heterophenomenological approach "permits theorists to agree in detail about just what a subject's heterophenomenological world is, while offering entirely different accounts of how heterophenomenological worlds map onto events in the brain (or the soul, for that matter)." This at least suggests that Dennett sees only two possible loci for studying consciousness: brain or soul. But this dichotomy is a false one. Events of consciousness need not map onto events in the brain or soul.

Now Dennett's particular brand of materialism may lead him to approach the study of consciousness via neurophysiology—even when such an approach conflicts with his own intentional-stance theory. But a neurophysiological approach to consciousness is not the only approach consonant with materialism. According to intentional-stance theory, an intentional attribution (as many attributions of conscious episodes are) is true if and only if it is predictive; and it is predictive if and only if the attributer has discerned a "real pattern." Now a perfectly respectable materialist may suppose that there are real intentional patterns in Dennett's sense, without supposing that these intentional patterns map onto any known or knowable physical patterns.

For example, when investigators suspect someone of insider trading of stock in a company which is on the verge of announcing a take-over, they look for patterns of buying and selling the particular stock in the days before the announcement. Sometimes they find such patterns; sometimes juries are sufficiently convinced of the reality of the patterns that they are willing to send people to prison on the basis of them. I agree with juries that there are such patterns: Patterns of buying and selling a stock are real intentional patterns, and such patterns stand on their own whether we ever find any physical-pattern correlates for them or not. Indeed, it would be highly unlikely that the real patterns of trading the stock were mirrored in real
On the one hand, in the case of putatively conscious phenomena with intentional contents, we get a distinction between conscious belief and nonconscious belief from the intentional stance alone: From the intentional stance, the theorist confirms Jane's report that she was conscious of believing that Had was trying to embarrass her at a particular time by noting the change in Jane's intentional behavior at that time. Brain mapping is irrelevant from the intentional stance. On the other hand, it is difficult to see how to go about applying the brain-mapping test to putatively conscious beliefs anyway. If the object of (putative) consciousness has intentional content, application of the brain-mapping test requires that we be able to assess the similarity between Jane's brain states and what is attributed to her by, for example, a belief that Had is trying to embarrass her. What properties of brain states could be similar to what Jane attributed to herself when she said, "At that moment, I realized that I believed that Had was trying to embarrass me"? What would count as similarity in this case? So, it seems to me that Dennett neither needs nor has logical space for the brain-mapping test in the case of putatively conscious phenomena with intentional contents.

Finally, a look at real intentional patterns, about which Dennett is a "mild realist," suggests that the brain-mapping test is not required for an empirical, third-personal account of consciousness. A claim to have found an intentional pattern would not be impugned by our failure to find a physical pattern onto which to map the intentional pattern; this is plain in nonpsychological cases. (It is no defense against allegation of income-tax fraud to complain that the prosecutors have not come up with a physical pattern that the culpability-producing intentional pattern mirrors.)

Dennett himself says that one can adopt the intentional stance to construct a heterophenomenological text "without giving up science." Since I am dubious that we will ever have a theory of consciousness, or of the mind generally, that is free of intentional presuppositions, I find it heartening that Dennett does not think that intentionality imperils science. But if we can adopt the intentional stance without giving up science, then we need not worry about mapping heterophenomenological items onto brain events. So, my suggestion for someone who wants an empirical, third-personal account of consciousness is to take the science and leave Dennett’s brand of physicalism with its metaphor of depth behind.

CONCLUSION

Dennett’s theory of consciousness is not, and cannot be, independent of his theory of content. Yet, as we saw in examining his brain-mapping test for determining what we are and are not conscious of, the two halves of Dennett’s project do not fit comfortably together. Moreover, I argued, the metaphor of depth precludes reconciliation of Dennett’s theory of consciousness with his intentional-stance theory of content. Although I would be more of a realist than Dennett about what is discerned from the intentional stance, I suggested that intentional-stance theory offers a better way to study putatively conscious phenomena with intentional contents than does the brain-mapping test.

NOTES

1. I am not claiming that thoughts of a person are wholly intentional, always devoid of any sensory aspect, nor that sensory phenomena (reported as, e.g., "I see a moving red dot") are wholly nonintentional. My point is that intentionality intrudes on any investigation of the full range of putatively conscious phenomena.


6. Ibid.


12. This aspect of Dennett’s work is responsible for his being taken to be an instrumentalist about belief. In “Real Patterns,” he resists the simple ‘instrumentalist/realist’ dichotomy. However, his discussion on pages 48f., with continued endorsement of Quine’s indeterminacy thesis, reaffirms the idea that to have an intentional feature is to be the object of a particular interpretive stance. Dennett allows that at least on occasion, “the choice of a pattern would indeed be up to the observer” (49). Here I am not concerned with the argument about realism but with the relations between the intentional stance and the physical stance.


16. *Consciousness Explained*, 76. In “How to Change Your Mind,” in *Brainstorms*, Dennett distinguishes between beliefs that beings without language can have and what he called ‘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” (78).

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

18. Ibid., 81.

19. Ibid., 407.

20. Ibid., 85; emphasis his.

21. Ibid., 81.

22. I am simply pointing out a way to isolate putatively conscious beliefs from all others—including tacit beliefs and beliefs that are subconscious in a psychoanalytic sense.


24. I would take this result to be tantamount to saying that no one is ever conscious of anything. Whatever consciousness would remain after removing consciousness of states of affairs (only pure sensations?) would seem to me negligible. For what it is worth, pure sensations seem at most to be a patrty part of one’s conscious life.


29. Ibid., 128.

30. The mathematical idea of ‘information’ in the Shannon–Weaver sense is a syntactic notion that does not suffice for propositional content. Syntax underdetermines propositional content.

31. Dennett emphasizes the indeterminacy of content in the brain and takes his opponent to claim that a thinker “begins with a determinate thought to be expressed” (*Consciousness Explained*, 241). I differ from both Dennett and his fictional opponent in that I hold that it is a mistake to try to locate propositional content in particular brain processes at all.


33. Ibid., 457.

34. Ibid., 458; emphasis his.

35. A physicalistic account of content-fixation would not collapse Dennett into Fodor; for on Fodor’s view, but not Dennett’s, brain events have syntactic structure.

36. I have no objection to helping oneself to intentionality, but then I am not a physicalist.

37. *Consciousness Explained*, 455.

38. Indeed, it would seem that only from the intentional stance is anything a metaphor in the first place.

39. “True Believers,” 16. The intentional stance is also contrasted with the design stance, from which one predicts and explains an entity’s behavior on the basis of its function or normal operation.


41. Compare: “Of course, if some version of mechanistic physicalism is true (as I believe), we will never need absolutely to ascribe any intentions to anything” ("Conditions of Personhood," in *Brainstorms*, 273).

42. *Consciousness Explained*, 457; emphasis his.

43. Although I agree that physics is more fundamental than chemistry, I reject the idea that all the sciences are neatly ordered according to depth.

44. Unlike Dennett, I would be as much a realist about features discerned from the intentional stance as about features discerned from the physical stance. But that major difference between Dennett and me is irrelevant to my proposal that Dennett treats consciousness as he treats content.


46. Ibid., 36.

47. A physicalist who holds that intentional patterns mirror undetectable physical patterns is in no position to use physical patterns as a constraint on intentional patterns.


49. If the thesis of global supervenience is true, then the intentional pattern may correspond to some physical pattern, but not one localized in space and time. Thus, it would be folly to try to find patterns of an agent’s physical properties that correspond to the agent’s intentional patterns. One does not have to be a dualist to see that brain states and bodily motions do not match up with beliefs and actions.

50. For a description of the Life World, see “Real Patterns.”


52. There may be cases of indeterminacy here; but there are many clear cases in which an intentional pattern is a smile and not a leer.


55. I would treat agents and believers in the same way as well; but, giving up the metaphor of depth, I would be more than a “mild realist” about both.

56. See "Evolution, Error, and Intentionality."

57. *Consciousness Explained*, 81.
What Is Dennett's Theory a Theory of?

Ned Block
MIT

In Consciousness Explained and some papers written before and since, Dan Dennett expounds what he says is a theory of consciousness. But there is a real puzzle as to what the theory is about. There are a number of distinct phenomena that ‘consciousness’ is used by Dennett and others to denote. If the theory is about some of them, it is false; if it is about others, it is banal.

A convenient locus of discussion is provided by Dennett’s claim that consciousness is a cultural construction. He theorizes that “human consciousness (1) is too recent an innovation to be hard-wired into the innate machinery, (2) is largely the product of cultural evolution that gets imparted to brains in early training.” Often, Dennett puts the point in terms of memes. Memes are ideas such as the idea of the wheel or the calendar or the alphabet; but not all ideas are memes. Memes are cultural units, the smallest cultural units that replicate themselves reliably. In these terms then, Dennett’s claim is that “Human consciousness is itself a huge complex of memes.” The claim is sometimes qualified (as in the “largely” above). I think the idea is that consciousness is the software that runs on genetically determined hardware. The software is the product of cultural evolution, but it would not run without the hardware that is the product of biological evolution.

I claim that consciousness is a mongrel notion, one that picks out a conglomeration of very different sorts of mental properties. Dennett gives us little clue as to which one or ones, which of the “consciousnesses” is