

Explaining Consciousness

David M. Rosenthal

Among mental phenomena, none seems so thoroughly to resist informative explanation as does consciousness. Part of the difficulty is due to our using the term 'conscious' and its cognates to cover several distinct phenomena, whose connections with one another are not always clear. And that often leads us to run these distinct phenomena together. Any attempt to explain consciousness, therefore, must begin by distinguishing the various things we call consciousness.

One such phenomenon is closely related to simply being awake. We describe people, and other creatures, as being conscious when they are awake and their sensory systems are receptive in the way normal for a waking state. I call this phenomenon *creature consciousness*. Consciousness in this sense is a biological matter, consisting in a creature's not being unconscious—that is, roughly, in its not being asleep or knocked out.

But we also use the term 'consciousness' for other phenomena that seem a lot less tractable to understanding and explanation. Not only do we distinguish between conscious and unconscious creatures; we also distinguish between mental states that are conscious and those which are not. I'll call this second property *state consciousness*. It's widely recognized that not all mental states are conscious. Intentional states such as beliefs and desires plainly occur without being conscious.¹ And, despite some division of opinion on the matter, I shall argue that the same is true of sensory states, such as pains and sensations of color. Such states not only can, but often do occur nonconsciously.²

Though creature consciousness and state consciousness are distinct properties, they are very likely related in various ways. Perhaps, for example, creatures must themselves be conscious for any of their mental states to be con-

scious, though if ordinary dreams are ever conscious states they are counterexamples to this generalization.³

Whatever the case about that, the property of creature consciousness is relatively unproblematic. We can see this by considering creatures mentally less well-endowed than we are whose mental states are never conscious, even when they are awake.⁴ Their mental states are all like the nonconscious mental states we are in when we are awake. Doubtless some creatures are actually like this, say, frogs or turtles. And it's plain that when none of a creature's mental states is conscious, there is nothing puzzling about what it is for the creature to be conscious. Some theorists might deny that such a case is possible, urging that no creature counts as conscious unless some of the mental states it is in are conscious states. But this seems little more than an unwarranted extrapolation of the normal human waking state to the case of all creatures. Even if their view were correct, moreover, it would be state consciousness that introduces the apparent mystery.

What is puzzling about consciousness must therefore be a matter not of creature consciousness, but of the consciousness of a creature's mental states. Because creature consciousness involves being responsive to sensory stimuli, if sensory states were all conscious, every conscious creature would perforce be in some conscious states. But it would still, then, be the consciousness of the states, not of the creature, which seems to induce some mystery.

For this reason, I shall focus here on state consciousness. After laying some groundwork in section I, I go on in section II to develop a hypothesis about what it is for a mental state to be conscious. On this hypothesis, a mental state is conscious if it is accompanied by a specific type

of thought. This is so whether the state that is conscious is itself an intentional state or a sensory state. Section III, then, supports this hypothesis with an argument that appeals to the ability creatures like ourselves have to report noninferentially about their own conscious states.

Sections IV and V take up the special case of conscious qualitative states. I argue in section IV that such sensory consciousness is just a special case of state consciousness and poses no additional problems of its own. And section V gives reasons for thinking that an accompanying intentional state can actually result in there being something it's like for one to be in a conscious sensory state. Section VI, finally, considers two general questions about state consciousness: What function it might have and whether consciousness can misrepresent what mental states we are in.

I. State Consciousness and Transitive Consciousness

Whatever else we may discover about consciousness, it's clear that, if one is totally unaware of some mental state, that state is not a conscious state. A state may of course be conscious without one's paying conscious attention to it and, indeed, even without one's being conscious of every mental aspect of the state. But if one is not at all aware of a state, that state is not a conscious state. This observation provides a useful start toward a theory of state consciousness. Because it is sufficient for a state not to be conscious that one be completely unaware of it, being aware of a state is perforce a necessary condition for that state to be a conscious state.

Being aware of a mental state, however, is not also a sufficient condition for the state to be conscious. There are ways we can be aware of our mental states even when those states are not conscious states. So, if we can rule out those ways, we'll be left with the particular way in which we are aware of our mental states when those states are conscious states. And this would give us a condition that's both necessary and sufficient for a mental state to be conscious.

For present purposes, I'll speak interchangeably of being aware of something and being conscious of that thing. So my strategy is to explain a state's being a conscious state in terms of our being conscious of that state in some particular way. No circle is involved here, since we are ex-

plaining one phenomenon in terms of another. It is one thing for us to be conscious *of* something—what we may call *transitive consciousness*—and another for a state to be a conscious state—what I'm calling *state consciousness*. And we understand transitive consciousness—our being conscious *of* things—independently of understanding what it is for mental states to be conscious states. We are transitively conscious of something by virtue of being either in an intentional or a sensory state whose content is directed upon that thing. And a state's having a certain content is a distinct property from that of a state's being conscious.⁵

It seems relatively uncontroversial that a state of which one is in no way transitively conscious could not be a conscious state. Even Descartes' usage, which still strongly influences our own, conforms to this commonsense observation, since he invariably describes the states we call conscious as states we are immediately conscious of. Nonetheless, Fred Dretske has recently challenged the observation that we are conscious of all our conscious states. According to Dretske, a state's being conscious does not consist in one's being conscious of the state; rather, a state is conscious if, in virtue of being in that state, one is conscious of something or conscious that something is the case. But every mental state satisfies this condition; so Dretske must hold that all mental states are conscious states. Accordingly, he urges that alleged cases of nonconscious mental states are unconvincing. Thus it is often said that a long-distance driver whose attention lapses perceives the road unconsciously,⁶ but Dretske rightly notes that perceiving can be inattentive without failing to be conscious.⁷

Many other examples of nonconscious mental states, however, are far more decisive. We often consciously puzzle over a question about what to do or how to solve a problem, only to have the answer occur to us later, without the matter having in the meantime been in any way consciously before our mind. Though it doesn't seem, from a first-person point of view, that we were thinking about the issue, it's clear that we must have been. And unlike the case of the long-distance driver, here no shift of attention would change things. Also we often take in sensory information without being at all aware of doing so, again no matter what we're paying attention to. Since, from a first-person perspective, we seem not to be in any relevant sensory states, those states are not conscious states.

Dretske also argues, however, that there are actual counterexamples to the idea that we are transitively conscious of all our conscious states. To adapt his argument slightly, consider two scenes, one of ten trees and the other just like it, but with one tree missing. And suppose that I consciously see both scenes, and indeed that I consciously see all the trees in each scene. But suppose, finally, that despite all this I do not notice any difference between the two scenes.

Dretske sensibly assumes that in this case I have conscious experiences of both scenes, including all the trees in each. Moreover, there is some part of the conscious experience of ten trees that is not part of the conscious experience of nine trees. That part is itself a conscious experience—a conscious experience of a tree. But, because I am not transitively conscious of the difference between the two scenes, Dretske concludes that I will not be transitively conscious of the experience of that extra tree. If so, the experience of the extra tree is a conscious experience of which I am not transitively conscious.⁸

This sort of thing is hardly an esoteric occurrence. Indeed, it happens all the time; let one scene be a slightly later version of the other, such that the later scene is altered in some small, unnoticed way. So, if Dretske's argument is sound, we often fail to be conscious of our conscious experiences.⁹

But the argument isn't sound. One can be conscious of an experience in one respect while not being conscious of it in another. For example, one may be conscious of a visual experience as an experience of a blurry patch, but not as an experience of a particular kind of object. Similarly, one could be conscious of the experience of the extra tree as an experience of a tree, or even just as part of one's overall experience, without being at all conscious of it as the thing that makes the difference between the experiences of the two scenes. Presumably this is just what happens in the case Dretske constructs. Dretske has not described a conscious state of which one is not transitively conscious.

There is a complication in Dretske's discussion that is worth noting. Dretske insists that being conscious of a difference, unlike being conscious of concrete objects and events, always amounts to being conscious "that such a difference exists."¹⁰ So he might urge that being conscious of a difference is always being conscious of it as a difference. But this won't help. The experience of the extra tree is that in virtue of which the two overall experiences differ.

Still, one can be conscious of the thing in virtue of which they happen to differ without being conscious that they do differ. As Dretske would put it, one can be conscious of that in virtue of which they differ but not of the difference between them;¹¹ indeed, he explicitly acknowledges that this very thing can happen.¹² Dretske's argument does not, therefore, undermine the commonsense observation that we are transitively conscious of all our conscious states.¹³

II. The Hypothesis

Let us turn, then, to the question of what it is that is special about the way we are transitively conscious of our mental states when those states are conscious states. Perhaps the most obvious thing is that, when a state is conscious, we are conscious of it in a way that seems immediate. Descartes emphasized this intuitive immediacy,¹⁴ which many have thought points toward a Cartesian theory of mind, on which a mental state's being conscious is an intrinsic property of that state. If nothing mediates between a state and one's being transitively conscious of it, perhaps that transitive consciousness is something internal to the state itself.

But the intuition about immediacy does not show that a mental state's being conscious is internal to the state. It does seem, from a first-person point of view, that nothing mediates between the conscious states we are conscious of and our transitive consciousness of them. But all that shows is that, if anything does mediate between a conscious state and our transitive consciousness of it, the mediating factor is not one we are conscious of. And the absence of conscious mediation is no reason to think that nonconscious mediation does not occur.¹⁵ Failure to appreciate this has led some to hold that we are conscious of our conscious states in a way wholly unlike the way we are conscious of everything else.

Even when something mediates between a conscious mental state and our being conscious of it, we can be conscious of the mediating factor; we just cannot be conscious of it as mediating. Compare what happens in perceiving. When we consciously perceive things, our conscious sensory states mediate between our perceptions and the objects we perceive, and since those states are conscious, we are conscious of them. Still, nothing in these cases seems intuitively to mediate. That's because we aren't con-

scious of anything as mediating. And the best explanation of that, in turn, is that the conscious sensory states that do in fact mediate do not figure in any conscious inference on which our perceiving is based. Similarly with the way we are conscious of our conscious mental states. Our being conscious of them seems unmediated because we are conscious of them in a way that relies on no conscious inference, no inference, that is, of which we are aware.¹⁶

Consider a case. I am annoyed, but unaware of it. Though my annoyance is not conscious, you observe my annoyed behavior and tell me I am annoyed. There are two ways I might react. I might accept what you tell me, but still feel no conscious annoyance. My belief that I'm annoyed would be the result of a conscious inference based on your remark, and possibly also a conscious inference from my coming to notice my own relevant behavior.¹⁷ But there is another possibility; your remark might cause me to become conscious of my annoyance independently of any such conscious inference. In that case my annoyance would have become a conscious state.

A state's being conscious involves one's being noninferentially conscious of that state. Can we pin down any further the way we are transitively conscious of our conscious states? There are two broad ways of being transitively conscious of things. We are conscious of something when we see it or hear it, or perceive it in some other way. And we are conscious of something when we have a thought about it. Which kind of transitive consciousness is relevant here? When our mental states are conscious, do we somehow sense those states or do we have thoughts about them?

The perceptual model may seem inviting. When we perceive things, we seem intuitively to be directly conscious of them; nothing seems to mediate between our perceptions and the objects we perceive.¹⁸ So perhaps the perceptual model can explain the apparent immediacy of the way we are conscious of our conscious states. But this advantage of the perceptual model won't help us decide between that model and the alternative view that we are conscious of our conscious states by having thoughts about them. Even though our thoughts do often rely on conscious inferences involving perceptions or other thoughts, they often don't.

There is, however, another consideration that seems to favor the perceptual model. A theory of consciousness must explain the qualitative

dimension of our conscious sensory states. And sensing always involves some sensory quality. So if we are conscious of our conscious states by sensing them, perhaps we can explain the qualitative dimension of consciousness as due to that higher-order sensing. Such an explanation, however, would at best just put off the problem, since the qualitative aspect of this higher-order perceiving would itself need to be explained in turn.

Not only do the considerations favoring the perceptual model fail to hold up; there is also reason to reject the model. Higher-order sensing would have to exhibit characteristic mental qualities; what qualities might those be? One possibility is that the higher-order perception and the state we perceive would both exhibit the same sensory quality. But this is theoretically unmotivated. When we perceive something, the quality of our perceptual state is distinct from any property of the object we perceive. When we see a tomato, for example, the redness of our sensation is not the same property as the redness of the tomato.¹⁹ So we have no reason to think that the higher-order qualities would be the same as those of our lower-order states.

If the higher- and lower-order qualities were distinct, however, it's a mystery what those higher-order qualities could be. What mental qualities are there in our mental lives other than those which characterize our first-order sensory states? And if the higher-order qualities are neither the same as nor distinct from our first-order qualities, the higher-order states in virtue of which we are conscious of our conscious states cannot have qualities at all. But if those higher-order states have no qualitative properties, they can only be higher-order intentional states of some sort.²⁰

We must therefore reject the perceptual model of how we are transitively conscious of our conscious states. The only alternative is that we are conscious of our conscious states by virtue of having thoughts about them. Since these thoughts are about other mental states, I shall refer to them as *higher-order thoughts* (HOTs).

This narrows down somewhat the way we are transitively conscious of our mental states when those states are conscious. But we can narrow things down even more. When a mental state is conscious, we are conscious of being in that state; so the content of our HOT must be, roughly, that one is in that very state.²¹ And, since merely being disposed to have a thought about something does not make one conscious of that

thing, the HOT must be an occurrent thought, rather than just a disposition to think that one is in the target state. Moreover, when we are conscious of something by being in an intentional state that's about that thing, the intentional state is normally assertoric. Indeed, it's likely that being in an intentional state whose mental attitude is not assertoric does not result in one's being conscious of the thing the intentional state is about.²² So we should require that the HOT has an assertoric mental attitude.²³ Finally, to capture the intuition about immediacy, we have seen that our HOTs must be independent of any inference of which we are aware. Our hypothesis, therefore, is that a mental state is conscious just in case it is accompanied by a noninferential, nondispositional, assertoric thought to the effect that one is in that very state.²⁴

One problem that seems to face this hypothesis is that, even when we are in many conscious states, we are typically unaware of having any such HOTs. But this is not a difficulty; we are conscious of our HOTs only when those thoughts are themselves conscious, and it's rare that they are. Moreover, the hypothesis readily explains why this should be so. The HOTs it posits are conscious thoughts only when they are accompanied, in turn, by yet higher-order thoughts about them, and that seldom happens. Not having conscious HOTs, moreover, does nothing at all to show that we do not have HOTs that fail to be conscious.

There is another reason it's useful to distinguish cases in which HOTs are conscious from cases in which they are not. The way we are ordinarily conscious of our conscious states differs from the way we are conscious of mental states of which we are introspectively conscious. Being introspectively conscious of a mental state involves, roughly, our deliberately focusing on that state, and very few of our conscious states are the subjects of any such introspective scrutiny. If being conscious of a mental state were the same as being introspectively conscious of it, it would be rare that we are conscious of our conscious states, and we would be unable to explain state consciousness in terms of transitive consciousness. Not distinguishing the two, moreover, would lead one mistakenly to see the HOT hypothesis as providing a theory only of introspective consciousness, and not of state consciousness generally.²⁵ But the present hypothesis actually allows us to explain what is distinctive about introspective consciousness. A state is introspectively conscious when the ac-

companying HOT is a conscious thought. Ordinary, nonintrospective state consciousness, by contrast, occurs instead when the HOT is not itself conscious.

The HOT model is a hypothesis about the nature of state consciousness, not an analysis of that concept. So it doesn't count against the hypothesis simply that one can imagine its not holding; one can always imagine things being different from the way they are.

There is an especially interesting argument that supports the appeal to HOTs. When a mental state is conscious, one can noninferentially report being in that state, whereas one cannot report one's nonconscious mental states. Every speech act, moreover, expresses an intentional state with the same content as that of the speech act and a mental attitude that corresponds to its illocutionary force. So a noninferential report that one is in a mental state will express a noninferential thought that one is in that state, that is, a HOT about the state. We can best explain this ability noninferentially to report our conscious states by supposing that the relevant HOT is there to be expressed. Correspondingly, the best explanation of our inability to report nonconscious states is that no HOTs accompany them.²⁶

One might reply that the ability to report conscious states shows only that there is a disposition for these states to be accompanied by HOTs, not that any HOTs actually accompany them.²⁷ Indeed, Peter Carruthers has extensively developed and supported the view that conscious states are simply those disposed to be accompanied by HOTs, and no actual HOT need occur. This, he argues, avoids having to posit the overwhelming computational capacity and cognitive space required for actual HOTs.²⁸

But this concern is not compelling. Neural implementation is not a problem, since ample cortical resources exist to accommodate actual HOTs. And, though introspection seems to suggest that the mind cannot accommodate very many actual HOTs at a time, that worry is also groundless. Introspection can tell us only about our conscious states, and by hypothesis HOTs are seldom conscious.

In any case, the dispositional model cannot explain what it is for states to be conscious. A mental state's being conscious consists in one's being conscious of that state in some suitable way, and simply being disposed to have a thought about something cannot make one conscious of it. Carruthers urges that we can get

around this difficulty if we understand a state's intentional content in terms of what other intentional states it is disposed to cause. A state's being disposed to cause a HOT might then confer suitable higher-order content on that state itself. But, if a state's being disposed to cause a HOT were a function of its intentional content, we could no longer explain how a state with some particular content is sometimes conscious and sometimes not.

III. Sensory Consciousness

On this argument, sensory consciousness is simply a special case of state consciousness—the special case in which the state that's conscious is a sensory state. Sensory states are states with sensory quality. So sensory consciousness occurs when a mental state has two properties: sensory quality and the property of state consciousness.

Moreover, these two properties are distinct and can occur independently of one another. State consciousness can of course occur without sensory quality, since nonsensory, intentional states are often conscious. But the converse is possible as well; sensory qualities can occur without state consciousness. Sensory qualities are just whatever properties sensory states have on the basis of which we distinguish among them and sort them into types. Since state consciousness consists in our being conscious of a mental state in some suitable way, these properties are independent of state consciousness. We would need some special reason to think that the properties on the basis of which we distinguish among sensations cannot occur except when we're conscious of the states that have those properties. It's hard to see what special reason there could be.

This conclusion conflicts with the familiar contention that sensory quality cannot occur nonconsciously. On that view, state consciousness is intrinsic, or essential, to sensory quality. But it's far from clear that this view is correct. Subliminal perception and peripheral vision both involve perceptual sensations of which we're wholly unaware, and the same is very likely true of such dissociative phenomena as blindsight.²⁹ Bodily sensations such as pains can also occur without being conscious. For example, we often have a headache or other pain throughout an extended period even when distractions intermittently make us wholly unaware of the pain.

One could of course simply dig in one's heels and insist that these phenomena are mere physiological occurrences that instantiate no sensory quality, and therefore that they are not mental phenomena at all. But without independent argument, that move amounts simply to saving a view by verbal fiat.

In any case there is good reason to resist that claim. The relevant nonconscious phenomena occur as essential parts of distinctively mental processes, and that suggests that they are themselves mental phenomena.³⁰ More specifically, conscious sensory states play the same roles in mental processing when their sensory qualities are the same, and correspondingly different roles when the qualities differ. And the nonconscious states in subliminal perception, peripheral vision, and blindsight play roles that in some respects at least parallel the roles played by conscious sensory states.

When bodily and perceptual sensations occur consciously, we taxonomize them by way of the sensory qualities we are conscious of. What is it, then, in virtue of which we taxonomize the nonconscious states that occur in these cases? Since many of the same qualitative distinctions figure in the nonconscious cases as figure in conscious sensing, we must assume that the nonconscious cases have the very same qualitative properties.³¹ Sensory qualities are the distinguishing properties of sensory states, the properties in virtue of which we classify those states. We use the properties we are conscious of to taxonomize sensory states generally, whether they are conscious or not. It's just that in the nonconscious cases we are not conscious of those properties. And, since there is nothing problematic about these distinguishing properties when the states that have them are not conscious, there can be no reason to find those properties puzzling when we are conscious of them. Sensory qualities will seem mysterious only if we assume that they cannot occur without being conscious. These considerations make the claim that sensory quality must be conscious seem less like a compelling commonsense intuition than a question-begging theoretical doctrine.

There is, of course, nothing it's like to have a pain or a sensation of red unless the sensation in question is conscious. And some have argued from this to the conclusion that sensory quality simply cannot exist unless there's something it's like to have it.³² But what it's like for one to have a pain, in the relevant sense of that idiom,

is simply what it's like for one to be conscious of having that pain. So there won't be anything it's like to have a pain unless the pain is conscious. Of course, if nonconscious pains were impossible, there would be no difference between a pain's existing and its being conscious, and its sensory quality would then exist only when there is something it's like to have it. But it begs the question simply to assume that pains, or other sensations, cannot exist nonconsciously. Moreover, the intuition that sensory states cannot exist nonconsciously gets whatever force it has from our first-person point of view. And it's unreasonable to rely on consciousness to tell us whether some phenomenon can exist outside of consciousness.

In a useful series of papers, Ned Block has urged that there are two distinct properties of mental states, both of which we call consciousness. One is captured by the notion of there being something it's like for one to be in a particular mental state; Block calls this property *phenomenal consciousness*. A state has the other property when its content is "poised to be used as a premise in reasoning, . . . [and] for [the] *rational* control of action and . . . speech."³³ This second property Block calls *access consciousness*. And he maintains that the two properties are, conceptually at least, independent. If Block is right, there is no single property of state consciousness, and the kind of consciousness that is characteristic of sensory states is, conceptually at least, distinct from the kind exhibited by many nonsensory states.

The idea behind Block's account of access consciousness is that a state's playing various executive, inferential, and reporting roles involves one's having access to that state, and having access to a state makes it conscious. But that's not always the case. States often play executive, inferential, and even certain reporting roles³⁴ without being conscious in any intuitive sense whatever. So, for a state to be access conscious, one must have access to that state, presumably by being transitively conscious of it in an intuitively immediate way.³⁵

Block's appeal to states' playing these roles doubtless reflects a desire to account for this kind of consciousness in computationally inspired functional terms, by providing a kind of flow chart that charts the connections a state has with various relevant systems. But for any such attempt to succeed, it must reflect an initial account of such consciousness in ordinary folk-psychological terms. Going straight to a subper-

sonal account is unlikely to give even an extensionally adequate account.

Block is doubtless right that access consciousness often occurs without phenomenal consciousness. We frequently have access to our mental states in the relevant way without there being anything it's like for us to be in them. Indeed, that's typically how it is with our thoughts and other intentional states. But the converse is far less clear. A state is access conscious only if one is transitively conscious of it. And if one is in no way transitively conscious of a mental state, there is nothing it's like for one to be in that state. It's not enough for the state just to have the distinguishing properties characteristic of some type of sensory state; for there to be something it's like for one to be in a state, one must be conscious of those distinguishing properties. So phenomenal consciousness cannot occur without access consciousness. Block's distinction does not, after all, show that sensory states are conscious in a way distinct from other types of mental state, nor that sensory states are in some special way invariably conscious.³⁶

IV. HOTs and What It's Like

Nonetheless, there does seem to be a serious problem about what it is for sensory states to be conscious. When a sensory state is conscious, there is something it's like for us to be in that state. When it's not conscious, we do not consciously experience any of its qualitative properties; so then there is nothing it's like for us to be in that state. How can we explain this difference? A sensory state's being conscious means that we are transitively conscious of that state in some suitable way. So being transitively conscious of a sensory state, in that particular way, must result in there being something it's like to be in that state. But how can being transitively conscious of a sensory state have this result? What way of being transitively conscious of our sensory states could, by itself, give rise to there being something it's like for us to be in those states? Perhaps, after all, Block is right that a sensory state's being conscious is not a matter of one's having suitable access to it.

The difficulty seems particularly pressing for the HOT hypothesis. An attraction of the perceptual model was that it might help explain the qualitative dimension of our conscious sensory states. Since perceiving involves sensory qualities, if a state's being conscious consisted in our

perceiving it, perhaps we could explain the way we are conscious of the qualities of our conscious sensations. As we saw, that explanation fails, since the higher-order qualities it appeals to would themselves need to be explained. But the HOT hypothesis may seem even less well-suited to deal with this problem. How can one's being in an intentional state, of whatever sort, result in there being something it's like for one to be in a conscious sensory state?

There are two ways the HOT theorist might try to show that being in a suitable intentional state can have this result. One would be to show that it's evident, from a first-person point of view, that one has a suitable HOT when, and only when, there is something it's like for one to be in some sensory state. We could then argue that one's having that HOT is responsible for there being something it's like for one to be in that state.

But if the HOT hypothesis is correct, we cannot expect to find any such first-person correlations. That's because, on that hypothesis, the HOTs in virtue of which our sensory states are conscious are seldom conscious thoughts. And when a thought is not conscious, it will seem, from a first-person point of view, that one does not have it.

So if the HOT hypothesis is correct, it will rarely seem, from a first-person point of view, that HOTs accompany one's conscious sensory states. Our first-person access reveals correlations only with conscious HOTs, not HOTs generally. And HOTs are conscious only in those rare cases in which one has a third-order thought about the HOT. But on the HOT hypothesis, HOTs need not be conscious for there to be something it's like to be in the target sensory states. So we cannot hope to test the hypothesis by correlating in a first-person way the occurrence of HOTs with there being something it's like to be in conscious sensory states.

But we need not rely solely on first-person considerations; there are other factors that help establish the correlation between having HOTs and there being something it's like for one to be in conscious sensory states. In particular, there is a striking connection between what HOTs we are able to have and what sensory qualities we are able to be aware of. And the best explanation of this connection is that accompanying HOTs do result in there being something it's like for one to be in states with those sensory qualities.

Consider wine tasting. Learning new concepts for our experiences of the gustatory and olfactory properties of wines typically leads to

our being conscious of more fine-grained differences among the qualities of our sensory states. Similarly with other sensory modalities; acquiring new concepts for specific musical and artistic experiences, for example, enables us to have conscious experiences with more finely differentiated sensory qualities. Somehow, the new concepts appear to generate new conscious sensory qualities.

There are two ways this might happen. One is that coming to have new concepts results in our sensory states' coming to have distinguishing properties that they did not previously have. This is highly implausible. How could merely having new concepts give rise to our sensory states' having new properties? On a widespread view, concepts are abilities to think certain things; how could having a new ability change the properties of the sensory states that result from the same type of stimulus?

But there is another possibility. The new concepts might result in new conscious qualities not by generating those properties, but by making us conscious of properties that were already there. The new concepts would enable us to be conscious of sensory qualities we already had, but had not been conscious of.³⁷

Possessing a concept allows us to form intentional states that have a certain range of contents. So which contents our intentional states can have must somehow make a difference to which sensory qualities can occur consciously. Moreover, the new concepts, which make possible conscious experiences with qualities that seem new to us, are the concepts of those very qualities.³⁸ So being able to form intentional states about certain sensory qualities must somehow result in our being able to experience those qualities consciously. It must result, that is, in there being something specific that it's like for us to be in the relevant sensory states.

How could this happen? The only plausible explanation is that a sensory quality's being conscious does actually consist in our having a HOT about that quality. This is true not only of the relatively finely differentiated qualities we have just now been considering. We can extrapolate to any sensory quality, however crudely individuated, and extrapolate even to whether or not we are conscious of any quality at all.

Take the conscious experience of hearing the sound of an oboe. If one's HOTs couldn't classify one's sensations in terms of the sound of an oboe but only that of some undifferentiated woodwind, having that sensation could not be

for one like hearing an oboe. And if one also lacked any concept of the sound of a woodwind, what it would be like for one to have that sensation would then be correspondingly more generic. If one lacked even the concept of a sensation's being of a sound as against being of some other type of stimulus, having the sensation would for one be like merely having some indiscriminate sensory experience or other. This sequence makes it plausible that peeling away that weakest HOT would result, finally, in its no longer being like anything at all to have that sensation. Even though HOTs are just intentional states, and so have no qualitative properties, having HOTs does make the difference between whether there is or is not something it's like for one to have particular sensations.

Because HOTs seldom occur consciously, we cannot, from a first-person point of view, note the occurrence of HOTs when, and only when, we are in conscious sensory states. Still, the argument from wine tasting does draw on first-person considerations. We know in a first-person way that learning new concepts for sensory qualities is enough for us to come to be conscious of our sensory states as having those qualities. And on that basis, we can infer that nonconscious HOTs are responsible for there being something it's like for one to be conscious of our sensory states in that way. It's just that the direct correlation between nonconscious HOTs and conscious sensory states is unavailable from a first-person point of view.

Is it enough to have correlations inferred from first-person considerations? Or must we work completely within a first-person point of view if we are to show that HOTs are responsible for there being something it's like for one to be in conscious sensory states?

A theory of consciousness must explain the first-person aspects of our conscious states. But the explanation need not itself rely only on first-person aspects. Indeed, to demand otherwise is to make any such explanation viciously circular. So the factor responsible for there being something it's like to be in a sensory state need not itself be a first-person aspect of that state, nor even something available from a first-person point of view. The HOTs in virtue of which our mental states are conscious need not, themselves, be conscious thoughts.

Compare the causal relations conscious sensory states have to stimuli, behavior, and other mental states. These relations are typically unavailable from a first-person point of view; we must infer them from other considerations, both

first- and third-person. Similarly, we may expect that whatever is responsible for there being something it's like for one to be in conscious sensory states is not directly accessible from a first-person point of view, but must instead be learned about by way of theoretical inference.

Some theorists have insisted that no correlations or theoretical developments could ever enable us to understand fully how physiological occurrences give rise to there being something it's like for one to be in conscious qualitative states. If so, perhaps we also cannot fully understand how HOTs could give rise to conscious qualities.

Joseph Levine calls this difficulty the "explanatory gap" and argues that it results from our being able to conceive of physiological occurrences without conscious qualities. By contrast, he claims, it's inconceivable that water could boil at a different temperature, at least holding constant the rest of chemistry.³⁹ But our ability to understand things and the apparent limits on what we can conceive are always relative to prevailing theory, whether scientific or folk theory, as Levine's holding chemistry constant illustrates.

Since the appearance of an explanatory gap simply attests our current lack of a well-developed, suitable theory, theoretical advances pertaining to conscious qualitative states should substantially narrow whatever gap seems now to obtain. And, though we may never fully eliminate that gap, we seldom if ever have a complete understanding of how any commonsense, macroscopic phenomenon arises.⁴⁰

The HOT model proceeds independently of physiology, but a similar explanatory gap seems to arise, since we need to understand how non-conscious HOTs can result in conscious qualities. Causal connections are irrelevant here, since there need be no causal tie between a HOT and its target. Rather, HOTs result in conscious qualities because they make us *conscious of ourselves as being in certain qualitative states*, which results in the subjective impression of conscious mental qualities. And the considerations raised earlier in this section provide reason to hold that HOTs can actually do this.

V. Consciousness, Confabulation, and Function

In closing I turn briefly to two unexpected implications of the HOT hypothesis, indeed, of any theory on which a mental state's being con-

scious consists, as I've argued it must, in one's being transitively conscious of that state.

As we have seen, the HOTs in virtue of which mental states are conscious represent those states in more or less fine-grained ways. And the way our HOTs represent the states they are about influences what those states are like from a first-person point of view. What it's like for me to have a particular gustatory sensation of wine depends on how much detail and differentiation goes into the HOT in virtue of which that sensation is conscious. Given any particular sensory state, different HOTs would yield different ways it's like for one thing one to be in that state.

Since the HOT that accompanies any particular sensory state can be more or less fine-grained, it is not the sensory state alone that determines what HOT one will have. That will depend also on such additional factors as the size of one's repertoire of concepts, one's current interests, how attentive one is, and how experienced one is in making the relevant sensory discriminations.

This raises an interesting question. Since the sensation itself does not determine what HOT one has, why can't the HOT misrepresent the sensory state one is in? Why can't one be in a sensory state of one type, but have a HOT that represents one as being in a sensory state of some different sort? The HOT one has, moreover, determines what it's like for one to be in the relevant sensory state. So why wouldn't an erroneous HOT make it seem, from a first-person point of view, as though one were in a sensory state that one is not in fact in?

There is reason to believe that this actually happens. Dental patients sometimes seem, from a first-person point of view, to experience pain even when nerve damage or local anesthetic makes it indisputable that no such pain could be occurring. The usual hypothesis is that the patient experiences fear or anxiety along with vibration from the drill, and consciously reacts as though in pain. Explaining this to the patient typically results in a corresponding change in what it's like for the patient when drilling resumes, but the patient's sense of what the earlier experience was like generally remains unaltered. The prior, nonveridical appearance of pain is indistinguishable, subjectively, from the real thing.

Other striking examples occur in connection with our perceptual sensations. As Daniel Dennett notes in *Consciousness Explained*, parafoveal vision can produce only low-

resolution sensations of most of the Marilyns in Warhol's famous painting,⁴¹ but we are aware of them all as clear and focused. What it's like for us is a function not of the character of our sensations, but of how we're conscious of those sensations.

There is also a well-known tendency people have to confabulate being in various intentional states, often in ways that seem to make *ex post facto* sense of their behavior;⁴² here it's plain that HOTs misrepresent the states that subjects are in. Similarly, it is very likely that repressed beliefs and desires are often actually conscious beliefs and desires whose content one radically misrepresents. Thus one might experience one's desire for some unacceptable thing as a desire for something else instead. In such a case, the desire is not literally unconscious; it is a conscious desire whose character is distorted by inaccurate HOTs. What it's like for one to have that desire fails accurately to reflect its actual content.⁴³

The HOT hypothesis is not the only theory to make room for these things; any theory on which a mental state's being conscious consists in one's being transitively conscious of that state will do so. As long as a conscious state is distinct from one's transitive consciousness of it, the content of that transitive consciousness may misrepresent the state. Conscious states are states *we are conscious of ourselves as being in*, whether or not we are actually in them.

The idea that what it's like for one to be in a state is determined not by that state's intrinsic properties but by the way one's HOT represents it enables us to understand certain cases that seem otherwise intractable to explanation. Suppose you're walking through the woods, stepping over branches as needed, but so deeply engrossed in conversation that you pay no conscious attention whatever to the branches. From a first-person point of view, you appear to have no thoughts about the branches; any thoughts about them you do have are not conscious thoughts.

To negotiate through the branches, however, you presumably need more than just thoughts about them; you must also have sensations of the branches. But from a first-person point of view, it may well also seem as though you have no such sensations. Unlike your thoughts, however, there is reason to doubt that your sensations of the branches literally fail to be conscious. It's not that there are no conscious sensations where one would expect sensations of branches to occur in one's visual field; the vi-

sual field does not seem to have gaps where the relevant sensations would be. Rather, the sensations that seem to you to be there are, roughly, just sensations of the undifferentiated rustic environment.

Why, then, are you unaware of your sensations of the branches? Plainly you have such sensations; that's how you manage to negotiate through the branches. And the sensations you have of the relevant part of the environment are all conscious; that's why your visual field doesn't seem to contain gaps. So it must be that the sensations are conscious not as sensations of branches, but only as sensations of the undiscriminated environment. We can explain this kind of occurrence only if the way one is transitively conscious of our sensations determines what it's like for one to have them. Compare Dennett's vivid example of looking straight at a thimble but failing to see it as a thimble. It's clear that one's sensation of the thimble is conscious, but one is conscious of it not as a sensation of a thimble but only, say, as a sensation of part of the clutter on a shelf.⁴⁴

In the thimble and branches cases, what it's like for one to be in particular sensory states is informationally less rich than the states themselves. But the opposite also happens, as when we experience our low-resolution sensations of the parafoveal Marilyns as though they were clear and focused. The best explanation is that our HOTs about our blurry parafoveal sensations represent them as having high resolution; the way we are conscious of our sensations actually corrects them by, as it were, bringing them into focus and touching them up.⁴⁵ Indeed, this drives home the need to posit occurrent higher-order states, since the high-resolution information must be embodied in some occurrent state.

This disparity between the properties of our sensations and the way we're conscious of them has important implications. For an example, consider Wilfrid Sellars' well-known argument that the sensory qualities of sensations exhibit an "ultimate homogeneity" that sets them apart from the particulate character of ordinary physical properties.⁴⁶ Sellars holds that this ultimate homogeneity derives from the way we conceive, in commonsense terms, of the perceptible properties of physical objects. Whatever the case about that, it is likely that those sensory qualities of sensations are themselves particulate. Being neurally based, the relevant sensory information will occur in the form of particular

pixels that represent color, shape, motion, and the like. We experience such information, however, as ultimately homogeneous simply because that is how we are conscious of the relevant informational states. The way we are conscious of our sensations smooths them out, so to speak, and elides the details of their particulate, bit-map nature.

Dretske has noted that theories on which a state's being conscious consists in one's being transitively conscious of the state seem unable to explain how a mental state's being conscious could have any function.⁴⁷ Being transitively conscious of a state, on these theories, makes no difference to the state's nonrelational properties. So the state's being conscious will make no difference to its causal role nor, therefore, to its function.

It's easy to overestimate the degree to which a state's being conscious does actually play any role. It's inviting to think, for example, that a state's being conscious somehow enhances any planning or reasoning in which that state figures. But the role a state plays in planning and reasoning is due to the content the state has, and that content will be invariant whether or not the state is conscious. So whether or not a state is conscious will not affect the state's role in planning and reasoning. We find it tempting to insist that a state's being conscious affects planning and reasoning when we consider actual cases in which the planning and reasoning are conscious. But those cases tell us nothing unless we compare them to nonconscious cases, to which we have no first-person access. Intuitions cannot help here.

In any event, Dretske has misdescribed the situation. On the HOT hypothesis, a conscious state is a compound state, consisting of the state one is conscious of together with a HOT. So the causal role a conscious state plays is actually the interaction of two causal roles: that played by the state itself and that played by the HOT.⁴⁸ This explains how a state's being conscious may to some extent matter to its causal role. Moreover, the way one is conscious of a conscious state may not fully match the target state one is actually in. In those cases, the causal role played by the HOT will matter even more. State consciousness does, after all, make some small difference to the function mental states have.⁴⁹

But what, then, of the compelling intuition that a mental state's being conscious does make a large and significant difference to its mental

functioning? That intuition is very likely due to the sense we have that our conscious thoughts, desires, and intentions occur freely and that this apparent freedom enhances our ability to reason and make rational choices. But our sense that these states occur freely itself arguably results from the way we are conscious of those states. Because we are seldom if ever conscious of anything as causing our conscious thoughts and desires, we have the subjective impression that they are uncaused, and hence free. So it seems that just being conscious of these states makes a significant difference to the role they can play in our lives. It is because the way we are conscious

of our intentional states presents them as free and uncaused that their being conscious seems to matter to our ability to reason and make rational choices.

I have argued that the HOT hypothesis explains how conscious states differ from nonconscious mental states, and why, to the extent that it does, state consciousness has a function. Moreover, the hypothesis squares well with there being something it's like to be in conscious sensory states. We can provisionally conclude that the hypothesis deals satisfactorily with the phenomenon of state consciousness, even for the special case of sensory states.

NOTES

1. *Pace* John R. Searle, *The Rediscovery of the Mind*, Cambridge, Massachusetts: MIT Press, 1992; see note 8, below.

I use 'intentional state' here to refer to states, like beliefs and desires, that exhibit propositional content along with some mental attitude.

2. For some related observations about different uses of 'consciousness' see Edmund Husserl, *Logical Investigations*, London: Routledge & Kegan Paul, 1970, II, pp. 535–36.
3. 'Ordinary' is to exclude so-called hypnagogic dreams, which occur in a semi-waking state.

Intuitions here are in any case hardly decisive. Are very vivid dream states conscious states? Must we be conscious when we're in them? Since it's far from clear what to say about these matters, it may well be that conscious states can occur without the creature itself being conscious.

4. There is, of course, nothing it's *like* for such a creature to be conscious—nothing it's like *for* the creature. But that doesn't mean there's nothing it *is* to be conscious.
5. Even if all sensations were conscious, what it is for a sensation to be *of* something would be a function not of its being conscious, but rather of the ways it qualitatively resembles and differs from other comparable sensations.

Strictly speaking, mental states aren't conscious of things; rather, it's creatures that are conscious of things in virtue of their being in mental states.

6. The best known version of the example is due to D. M. Armstrong, "What Is Consciousness?" in his *The Nature of Mind*, St. Lucia, Queensland: University of Queensland Press, 1980: 55–67, p. 59. See Dretske's *Naturalizing the Mind*, Cambridge, Massachusetts: MIT Press/Bradford Books, 1995, pp. 104–5.
7. "Conscious Experience," *Mind* 102, 406 (April 1993): 263–83; reprinted in Dretske, *Perception, Knowledge, and Belief: Selected Essays*, Cambridge: Cambridge University Press, 2000, 113–37, p. 123; *Naturalizing the Mind*, chapter 4.
8. "Conscious Experience," pp. 125–28; cf. *Naturalizing the Mind*, pp. 112–3.
9. One might object that we are, in any case, conscious

of our conscious states when we are introspectively aware of them. To forestall this objection, Dretske has recently argued that introspection resembles what he calls displaced perception. Just as we come to know how full the gas tank is by looking at the gauge, so we come to know what mental state we're in by noticing what we're seeing. We thereby come to be conscious that we're in some particular mental state, but not conscious *of* that state. (Dretske, "Introspection," *Proceedings of the Aristotelian Society*, CXV [1994/95]: 263–78, and *Naturalizing the Mind*, ch. 2.)

On this ingenious proposal, introspection is a matter of coming to know how one represents things (274–75). But introspection is better construed as knowing what mental state one is in, independently of how that state represents nonmental reality. But even if Dretske's right about what introspection is, just seeing that I represent things as being a certain way won't yield introspection unless I see this consciously. So either the argument rests on Dretske's assumption that all mental states are conscious, or he must give a different account of what it is for states to be conscious.

10. "Conscious Experience," 128; cf. 117–18.
11. In his useful "Dretske on HOT Theories of Consciousness," William Seager independently gives a similar account of how Dretske's argument fails to undermine the HOT hypothesis (*Analysis* 54, 1 [January 1994]: 270–76, esp. pp. 275–76).
12. "But readers who were only thing-aware of the difference between Alpha and Beta [the two arrays in Dretske's example] were not fact-conscious of the difference between Alpha and Beta." ("Conscious Experience," p. 128.)
13. John R. Searle also denies that we are conscious of our conscious mental states, though for reasons different from Dretske's. "[W]here conscious subjectivity is concerned, there is no distinction between the observation and the thing observed" (*The Rediscovery of the Mind*, p. 97). The context makes clear that Searle is denying not just that we can observe our conscious states, but that we are conscious of them at all, in the way we're conscious of other

things: "We cannot get at the reality of consciousness in the way that, using consciousness, we can get at the reality of other phenomena" (96–97). This is because "where conscious subjectivity is concerned, there is no distinction between the observation and the thing observed" (97).

Searle argues for this by appeal to the idea that we can describe consciousness only in terms of what it's consciousness of (96). But even if that's so, there will be states in virtue of which we are conscious of things. So it doesn't follow that there aren't states in virtue of which we are conscious of our conscious states.

14. "[T]he word 'thought' applies to all that exists in us in such a way that we are immediately conscious of it" (Geometrical Exposition of the *Second Replies*, *Oeuvres de Descartes*, ed. Charles Adam and Paul Tannery, Paris: J. Vrin, 1964–75, VII, 160).

15. Nonconscious mediation, moreover, might well occur; factors of which we're in no way conscious often causally mediate among distinct mental states, even when we're aware of them from a first-person point of view.

Our intuitive sense that we're not conscious of our conscious states in any way that's mediated may be what leads Searle to claim that there's no way in which we're conscious of our conscious states (see n. 13). It also distinguishes this case from the way we're perceptually conscious of things, in which we are sometimes conscious of the intervening medium.

16. A slight adjustment to this is needed. One might hold a theory on which an inference mediates between our being conscious of our conscious states and the states themselves, though we're conscious of that inference only by another inference based on the theory. (I thank Eric Lormand for raising this possibility.) We would still count as conscious the same states, even though the theory makes us conscious of the inferential mediations. We can provide for this by stipulating that if a state is conscious, we're conscious of it in a way that does not require that we be conscious of any inference that may occur. Our being conscious of the state may rely on some inference, but not on our being conscious of it.

This handles a related possibility as well. Suppose that inferences of which we're not conscious normally mediate between our being conscious of our conscious states and those states. Even if we somehow became conscious of some of those inferences without benefit of theory, we'd count the same states as conscious. The adjusted stipulation provides for this. Since nothing in what follows hinges on this sort of thing, I'll omit this qualification.

If the way we're conscious of our conscious states were sometimes based on conscious inference, we'd then know how we come to be conscious of those states. Though we're conscious of our conscious states, we generally don't, from a first-person point of view, have any idea how we come to be conscious of them. That ignorance helps explain the air of mystery that surrounds state consciousness.

17. The inference that consciously mediates between mental states and one's being conscious of them need not begin with the mental state to one's being conscious of it; typically, the conscious inference would start, instead, from noticing one's behavior or from the remarks of others. And because those

things are causally due to one's mental state, such an inference counts as mediating between a mental state and one's being conscious of it.

We need not independently preclude reliance on observation. Intuitively, one's being conscious of a mental state can be immediate even if it relies on observation, so long as one is not aware of its doing so. And that will be so if there's no reliance on any conscious inference. E.g., if one observes one's happy gait and so, without any inference of which one is aware, takes oneself to be happy, the way one is conscious of being happy is intuitively immediate.

18. Although we recognize on reflection that mediation does in fact occur, no conscious inference normally mediates, and as we've seen, that's what matters for the intuition of immediacy.

19. On this, see David M. Rosenthal, "The Colors and Shapes of Visual Experiences," in *Consciousness and Intentionality: Models and Modalities of Attribution*, ed. Denis Fisette, Dordrecht: Kluwer Academic Publishers, 1998, pp. 137–69; and "The Independence of Consciousness and Sensory Quality," in *Consciousness: Philosophical Issues, I, 1991*, ed. Enrique Villanueva, Atascadero, California: Ridgeview Publishing Company, 1991, pp. 15–36, reprinted in *Consciousness and Mind*, Oxford: Clarendon Press, forthcoming.

20. These considerations are reminiscent of an argument of Aristotle's at *de Anima* III, 2, 425b12–14, though Aristotle also held that the redness of our perceptions is the very same quality as the redness of physical objects (e.g., *de Anima* II, 5, 418a4; II, 11, 423b31; II, 12, 424a18; III, 2, 425b23).

Perhaps the qualities of the higher-order states are those our sensory states seem to have, and the lower-order qualities do not figure in what it's like for us to be in sensory states. But locating the qualities that figure in what it's like to be in sensory states at the higher level doesn't help explain the qualitative dimension of those states.

21. The concern that nonlinguistic creatures can't be in intentional states with such sophisticated content may also motivate preference for the perceptual model, since perceiving is a less sophisticated mental phenomenon. But little conceptual richness is needed to be in such intentional states. The concept of self, e.g., need involve no more than the distinction between oneself and everything else. And the state itself can be conceptualized in a relatively minimal way, say, just as some way the creature is.

22. So HOTs are not simply about intentional contents, but about full-fledged intentional states: contents plus mental attitudes.

If I doubt or wonder whether a particular physical object is red, I'm conscious of that object; similarly if I expect, hope, or desire that it is. But it's not the doubt, wonder, hope, or desire that makes me conscious of the object. Rather, if I doubt whether the object is red or desire or suspect that it is, I must also think assertorically that the object is there, or exists, and I'm conscious of the object in virtue of my having that assertoric thought. This is evident because, in such a case, I wouldn't be conscious of the object as red, but just as something that exists. The content of my consciousness is determined not by the content of my nonassertoric intentional state, but by the assertoric state. Similarly with intentional states

about our own mental states; being in nonassertoric intentional states about one's mental states make one conscious of being in those states only if they require one also to have the assertoric thought that one is in that state.

It's worth noting an argument of Robert M. Gordon that many emotions must be accompanied by corresponding beliefs; being angry that *p*, e.g., requires believing that *p*. (*The Structure of Emotions: Investigations in Cognitive Philosophy*, Cambridge: Cambridge University Press, 1987, pp. 47ff.) If so, the required belief would explain why, when one's angry that *p*, one is conscious of whatever '*p*' is about. In any case, this result depends on describing the emotion in terms of its intentional content. Thus, if one describes a person not as being angry that *p*, e.g., but as being angry because *p*, no corresponding belief is implied.

- 23. This helps deal with an interesting objection. Freudian theory may seem to posit states that are nonconscious despite their being accompanied by suitable HOTs. (This idea has been pressed by Georges Rey and Stephen Schiffer.) But it's not easy to come up with convincing examples. Pleasure or guilt about repressed states won't do because pleasure and guilt aren't assertoric; so we often aren't conscious of the objects of our pleasure or guilt—even when those states are conscious.

Even if we could come up with plausible examples, moreover, it is far from obvious that Freudian theory requires that we describe the situation as involving nonconscious states accompanied by HOTs, since there typically are several equally good explanations for any such phenomenon. It's also important to note that repressed states are seldom nonconscious states. Rather, they're typically states we disguise by radically misrepresenting their content, or distract ourselves from by creating elaborate mental noise. See p. 29, below.

- 24. According to Searle, the intentional content of perceptual states always refers to those very states; if I see a yellow station wagon, the content of my visual perception is "that there is a yellow station wagon there and that there is a yellow station wagon there is causing this visual experience" (*Intentionality: An Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press, 1983, p. 48). If the content of every perceptual state were partly that one is in that state, then on the HOT hypothesis, just being in the state would make one conscious of it, and non-conscious perceptions would be impossible. (I am grateful to Gilbert Harman for raising this concern.) Moreover, perceiving something does presumably make one conscious of that thing, arguably because the mental attitude of perceiving is assertoric.

Searle's argument for this claim appeals to the truth conditions of perceptions; a state's intentional content "determines under what conditions it is satisfied" (p. 48), and one perceives a thing only if it causes one's perception. But the conditions under which the perception is satisfied are simply that there's a yellow station wagon there, not also that the perception is caused by there being a yellow station wagon there. The causal condition is relevant not to the truth of what I perceive, but of whether I perceive it.

These considerations do, however, point toward an explanation of how many perceptual states do

come to be conscious. We assume as a general belief about about the world that the states of affairs we perceive normally cause the relevant perceptual states. When one has the (typically nonconscious) thought that a perceived state of affairs has caused the perceptual state, that thought results in a HOT that one is in the perceptual state, and thus results in that state's being conscious.

- 25. See, e.g., Dretske, "Conscious Experience," esp. Section 4; also Ned Block, review of Daniel C. Dennett, *Consciousness Explained*, *The Journal of Philosophy* XC, 4 (April 1993): 181–93, who alludes on p. 182 to the HOT hypothesis.
- 26. This argument is developed in detail in my "Thinking That One Thinks," in *Consciousness: Psychological and Philosophical Essays*, ed. Martin Davies and Glyn W. Humphreys, Oxford: Basil Blackwell, 1993, pp. 197–223. On the connection between thought and genuine speech, see my "Intentionality," *Midwest Studies in Philosophy*, X (1986): 151–84. Both will be reprinted in *Consciousness and Mind*.

The argument relies on creatures that can describe their mental states. But noninferential reportability simply helps fix the extension of 'conscious state'; many nonlinguistic creatures are also in conscious states.

Special issues arise about qualitative states, since there is no such thing as verbally expressing a perceptual sensation. We can express perceptions, but only because perceptions, unlike sensations, have an intentional aspect and it's that intentional component that we can verbally express. The same may also hold for bodily sensations; though we can express a pain by uttering 'ouch,' it's unclear that 'ouch' counts as a verbal, as opposed to nonverbal, form of expressing. And, though saying 'It hurts' is linguistic, that reports the pain, rather than expressing it. Still, creatures with suitable linguistic ability can noninferentially report their conscious states, whether the states are intentional or sensory.

These considerations have a bearing on the perceptual model. When a state is conscious, creatures with the relevant linguistic ability can express their transitive consciousness of the state. If there were a higher-order perception of the state, one's report would verbally express only the intentional component of that higher-order perception. But that's in effect just to express a HOT. So the argument from reporting and expressing shows that if the transitive consciousness of a conscious state did have a sensory aspect, that sensory aspect would be irrelevant to the state's being intransitively conscious.

- 27. Dennett and Harman have independently pressed this reply in conversation, and it receives tacit expression in Dennett's view that "[c] onsciousness is cerebral celebrity" ("The Message Is: There is no Medium," *Philosophy and Phenomenological Research* LIII, 4 [December 1993]: 919–31, p. 929). See also Dennett, *Consciousness Explained*, ch. 10 and esp. p. 315.
- 28. Peter Carruthers, *Language, Thought, and Consciousness: An Essay in Philosophical Psychology*, Cambridge: Cambridge University Press, 1996, and *Phenomenal Consciousness: A Naturalistic Theory*, Cambridge: Cambridge University Press, 2000.
- 29. See Lawrence Weiskrantz, *Blindsight*, Oxford: Oxford University Press, 1986, and *Consciousness Lost*

and Found: A Neuropsychological Exploration, Oxford: Oxford University Press, 1997.

There is reason to think that discrimination of stimuli with different form may be due to discrimination of orientation, rather than of form itself (*Blindsight*, 84). Van Gulick has argued that this shows that blindsight does not involve states with phenomenal properties like those of conscious visual sensations. ("Deficit Studies and the Function of Phenomenal Consciousness," in *Philosophical Psychopathology*, ed. George Graham and G. Lynn Stephens, Cambridge, Massachusetts: MIT, 1994.) But that conclusion follows only if one assumes that sensory qualities must be integrated in just the way they are in normal conscious cases.

30. A classical example is the so-called cocktail-party effect. We typically screen out the sounds of conversations other than our own, though mention of one's name in a screened-out conversation normally causes one's attention suddenly to shift to that conversation.
31. Compare parallel arguments that certain nonconscious states have mental properties because of the roles they play in mental processes; e.g., J. A. Fodor, "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology," *The Behavioral and Brain Sciences* III, 1 (March 1980): 63–73.
32. See Thomas Nagel's "What Is It Like to Be a Bat?," *The Philosophical Review* LXXXIII, 4 (October 1974): 435–50; "Panpsychism," in *Mortal Questions*, Cambridge: Cambridge University Press, 1979, pp. 181–95; and *The View from Nowhere*, New York: Oxford University Press, 1986, chapters 1–4.
33. "On a Confusion about a Function of Consciousness," *The Behavioral and Brain Sciences*, 18, 2 (June 1995): 227–47, p. 231; emphasis Block's. See also Block, review of Dennett's *Consciousness Explained*, p. 184; "Begging the Question against Phenomenal Consciousness," *The Behavioral and Brain Sciences* 15, 2 (June 1992): 205–6; "Consciousness and Accessibility," *The Behavioral and Brain Sciences* XIII, 4 (December 1990): 596–98.
34. It's not that the states we report are nonconscious, but nonconscious states influence what we report and how we do it.
35. Block's definition of access consciousness in terms of a state's being "poised" for certain things gives a dispositional mark of such consciousness. (In the review of Dennett's book he uses the phrase 'freely available' [p. 182].) That is compatible with access consciousness's consisting in a subject's being transitively conscious of a mental state, rather than simply being disposed to be conscious of it. States we are transitively conscious of have many dispositional properties, among them being reportable and introspectible.
36. Block distinguishes a third concept of consciousness, which he calls reflective consciousness (review of Dennett, p. 182) or monitoring consciousness ("On a Confusion," p. 235). According to Block, a state is conscious in this way if one has a HOT about it. But the states he counts as reflectively or monitoring conscious are states that we're introspectively conscious of: states that we're conscious of being conscious of. This is a distinct notion of consciousness, but Block is mistaken to define it in terms of having HOTs. Rather a state has monitoring consciousness, in his terms, only if one has a *conscious* HOT about it. See n. 25.
- For more on Block, see Rosenthal, "Phenomenal Consciousness and What It's Like," *The Behavioral and Brain Sciences*, 20, 1 (March 1997), pp. 64–65, "The Kinds of Consciousness," MS, and "How Many Kinds of Consciousness," MS.
37. Of course, the relevant sensory states will often have been conscious before one acquired the more fine-grained concepts, but conscious only with respect to qualities individuated in a more coarse-grained way. E.g., one might initially be conscious of a particular type of olfactory sensation solely as being unlike, and subsequently become conscious of it in terms of more fine-grained sensory qualities.
38. One might argue that the new concepts pertain not to the distinguishing properties of our conscious sensory experiences, but rather to the perceptible properties of the perceived physical objects and processes, e.g., the wine or the musical performance. (See Harman, "The Intrinsic Quality of Experience," *Philosophical Perspectives*, 4: *Action Theory and Philosophy of Mind*, 1990, pp. 31–52.) But it's clear that in the cases just imagined we also focus introspectively on the distinguishing properties of our conscious sensory states. So those cases involve new concepts of the distinguishing properties of sensory states.
39. "On Leaving Out What It's Like," in *Consciousness: Psychological and Philosophical Essays*, ed. Martin Davies and Glyn W. Humphreys, Oxford: Basil Blackwell, 1993, 121–36, p. 134, and *Purple Haze: The Puzzle of Consciousness*, New York: Oxford University Press, 2001, pp. 79. See also "Materialism and Qualia: The Explanatory Gap," *Pacific Philosophical Quarterly* LXIV, 4 (October 1983): 354–61. For related arguments see David J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*, New York: Oxford University Press, 1996.
- Similarly, Nagel claims we have a purely rational understanding of why "heat caus[es] water to boil, rocks caus[e] glass to break, magnets induce[] electric current, [and] the wind mak[es] waves" ("Panpsychism," 186), but currently lack any understanding of how physical heat, e.g., or a brain process, could causally necessitate a pain or other sensation ("Panpsychism," 187).
40. See my "Reductionism and Knowledge," in *How Many Questions?*, ed. Leigh S. Cauman, Isaac Levi, Charles Parsons, and Robert Schwartz, Indianapolis: Hackett Publishing Co., 1983, 276–300.
41. Daniel C. Dennett, *Consciousness Explained*, Boston: Little, Brown and Company, 1991, p. 354. See pp. 53–54 for Dennett's striking illustration of these limits in attempting to discern the color of playing cards seen parafoveally at arm's length.
42. For a classic study, see Richard E. Nisbett and Timothy DeCamp Wilson, "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review* LXXXIV, 3 [May 1977]: 231–59.) Nisbett and Wilson's influential study focused not only on cases in which subjects confabulate stories about the causes of their being in particular cognitive states, but also on cases in which they confabulate accounts about what states they're actually in.

43. Perhaps such erroneous HOTs might figure also in apparent self-deception.
It may sometimes be difficult to tell whether a HOT misrepresents an actual target or the HOT has only a notional target, and there is an actual state that simply isn't conscious. Indeed, it may well be arbitrary within a certain range of cases which way we describe a case.
44. *Consciousness Explained*, p. 336.
Similarly, in the cocktail-party effect, one's attention shifts to a previously unattended conversation in which one's name was mentioned. So one must have been hearing the articulated words in that conversation, though to consciousness it seemed just to be background din.
Robust experimental findings, e.g., those involving masked priming, also provide compelling evidence that what it's like to have a sensation sometimes diverges from the properties of the sensation itself. In masked priming, subjects report being unaware of qualitative input whose presence is evident from its effect on subsequent cognitive behavior. For a classic study, see Anthony J. Marcel, "Conscious and Unconscious Perception: Experiments on Visual Masking and Word Recognition," *Cognitive Psychology* 15 (1983): 197–237.
Experimental work on change blindness also provides vivid evidence for divergence of how we're conscious of our sensations from their actual properties. Subjects here fail consciously to register visible changes so salient that it's overwhelmingly likely that corresponding changes do occur in their visual sensations. So subjects' sensations diverge from how they're aware of them. Moreover, the compelling impression we all have of being continuously conscious of salient qualitative detail is evidently erroneous. See John Grimes, "On the Failure to Detect Changes in Scenes across Saccades," in *Perception*, ed. Kathleen Akins, New York: Oxford University Press, 1996, pp. 89–110; Daniel J. Simons, "Current Approaches to Change Blindness," *Visual Cognition* 7 (2000): 1–16; and Ronald A. Rensink, "The Dynamic Representation of Scenes," *Visual Cognition*, 7, 1/2/3 (January 2000): 17–42, and "Seeing, Sensing, and Scrutinizing," *Vision Research*, 40, 10–12 (2000): 1469–87.
45. In aesthetic experience, also, how we are conscious of a sensation presumably outstrips that sensation's qualitative character.
For more on sensations' diverging from the way we are conscious of them and the way HOTs function in that connection, see "Sensory Qualities, Consciousness, and Perception," forthcoming in *Consciousness and Mind*, and "Consciousness and Metacognition," in *Metarepresentation: Proceedings of the Tenth Vancouver Cognitive Science Conference*, ed. Daniel Sperber, New York: Oxford University Press, 2000, 265–95.
46. Often referred to as Sellars' "grain argument." Wilfrid Sellars, "Philosophy and the Scientific Image of Man," in *Frontiers of Science and Philosophy*, ed. Robert G. Colodny, Pittsburgh: University of Pittsburgh Press, 1962, pp. 35–78; reprinted in *Science, Perception and Reality*, 1–40, p. 36; also p. 35, and "Phenomenalism," also in *Science, Perception and Reality*, 60–105, pp. 103–5.
Cf. Peter Carruthers' claim that "perceptual information is *analogue* (that is, 'filled in' and continuous)," and "the subjective aspect of an experience just *is* analogue information about [physical] red, presented to a cognitive apparatus having the power classify states as information carriers, as well as to classify the information carried" (Peter Carruthers, *Language, Thought, and Consciousness: An Essay in Philosophical Psychology*, Cambridge: Cambridge University Press, 1996, pp. 167, 214).
47. *Naturalizing the Mind*, p. 117.
48. The interaction of the two roles may not be additive; the causal properties of the HOT may interact with those of the state in such a way that the original causal properties of the state are modified, or even blocked altogether.
49. So conscious inessentialism, on which every intelligent activity we perform consciously could be performed without its being conscious, is mistaken. The label is due to Owen Flanagan, who rejects the thesis (*Consciousness Reconsidered*, Cambridge, Massachusetts: MIT Press/Bradford Books, 1992, pp. 5, 129ff.).

PHILOSOPHY OF
MIND

Classical and
Contemporary Readings

David J. Chalmers

New York Oxford
OXFORD UNIVERSITY PRESS
2002

Oxford University Press

Oxford New York

Auckland Bangkok Buenos Aires Cape Town Chennai

Dar es Salaam Delhi Hong Kong Istanbul Karachi Kolkata

Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi

São Paulo Shanghai Singapore Taipei Tokyo Toronto

and an associated company in Berlin

Copyright © 2002 by Oxford University Press, Inc.

Published by Oxford University Press, Inc.

198 Madison Avenue, New York, New York, 10016

<http://www.oup-usa.org>

Oxford is a registered trademark of Oxford University Press

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data

Chalmers, David John, 1966–

Philosophy of mind : classical and contemporary readings / David J. Chalmers.

p. cm.

Includes bibliographical references.

ISBN-13 978-0-19-514581-6

ISBN 0-19-514580-1 (hardback : alk. Paper)—ISBN 0-19-514581-X (pbk : alk.paper)

I. Philosophy of mind. I. Title.

BD418.3 .C435 2002

128'.2—dc21

2002072403

Printing number: 9 8 7 6 5

Printed in the United States of America
on acid-free papers