

46 A Theory of Consciousness

David M. Rosenthal

A judicious division of labor often advances our understanding of things. This is no less true in the study of mind than in other areas of investigation. For example, because the intentional and sensory features of mental states differ radically, it is widely accepted that we will make progress by studying each kind of property independently of the other. Of course many mental phenomena, such as perceptual states and emotions, have both kinds of property; but other mental states exhibit only one of the two. And the difference between intentional and sensory character is so pronounced that the study of each raises problems unlike those encountered with the other. Pursuing independent investigations allow us to ignore the irrelevant problems.

Is a similar division of labor desirable in the case of consciousness? Can we profitably investigate consciousness independently of intentionality and sensory character? In particular, can we usefully separate the question of what it is for mental states to be conscious from the question of what it is for those states to have intentional or sensory character?

Some have held that when it comes to consciousness such a division of labor is not only undesirable; it is not even possible. On this view consciousness is intrinsic to being an intentional or sensory mental state; so one cannot understand what it is for states to have sensory or intentional character without knowing what it is for those states to be conscious. Nor, then, can we fully grasp what it is for a state to be a conscious state unless we understand why a state's being conscious is intrinsic to its having intentional or sensory character.

I shall argue that this view is mistaken, and that a separation of these questions is necessary to any satisfactory understanding of what it is for a mental state to be a conscious state. More specifically, I argue that if such consciousness were intrinsic to sensory or intentional character, no

theoretical understanding of what it is to be a conscious state would be possible at all.

My argument proceeds in several steps. After drawing some preliminary distinctions in section I, I argue in section II that mental states can occur without being conscious. In section III, then, I show that if mental states were all conscious, or if being conscious were intrinsic to those states which are conscious, an informative explanation of such consciousness would be impossible. This provides a powerful reason to deny that being conscious is such an intrinsic property and, thus, to embrace the division of labor just described. I then show that accepting this division of labor paves the way for a natural and independently defensible theory of what such consciousness consists in. Section IV draws a distinction that points to this positive theory, which I develop and defend in sections V through VII.

I Preliminary Distinctions

Two distinct issues are often run together in discussions of consciousness. One is the question of what it is for a mental state to be conscious. Assuming that not all mental states are conscious, we want to know how the conscious ones differ from those which are not. And, even if all mental states are conscious, we can still ask what their being conscious consists in. We can call this the question of state consciousness. This is my main concern in what follows.

But we not only describe mental states as being conscious or not; we also ascribe consciousness to creatures. So there is a second question, that of what it is for a person or other creature to be conscious, that is, how conscious creatures differ from those which are not conscious. We can call this the question of creature consciousness. No special puzzle exists about what it is for a creature not to be conscious; so in this case the contrast between being conscious and not being

conscious is reasonably transparent. Accordingly, the answer to this question is tolerably clear, at least in rough outline: To be conscious, a person or other creature must be awake and sentient. Though this answer is schematic, it does tell us what a full account would have to look like.¹

The questions about creature consciousness and state consciousness are plainly distinct. The property of a state's being conscious is different from that of a creature's being conscious. Indeed, since mental states are states of creatures, the properties of such states cannot be the same as the properties of creatures.

Issues about the two kinds of consciousness are of course not entirely independent. Perhaps being in conscious states implies that a creature is awake, and therefore conscious. And some might insist that the converse connection holds, perhaps because being sentient implies that at least some of one's sensory states are conscious.

Nonetheless, answers to one question will not help much with the other. Knowing that a creature's being conscious means that it is awake and sentient does not help us understand what it is for a mental state to be conscious, or place useful constraints on what such state consciousness consists in. Nor is there reason to think that a complete account of creature consciousness would help more. And even if creatures must themselves be conscious for their mental states to be conscious states, knowing what it is for mental states to be conscious would tell us little about creature consciousness.

Despite this, state consciousness and creature consciousness not only are seldom explicitly distinguished, but are often run together. Perhaps this is due in part to a natural tendency to speak of consciousness, rather than of something's being conscious. This seemingly innocuous shorthand encourages one to lose sight of the fact that there are two very different sorts of things we describe as being conscious, and hence which of the two is at issue. Intuitions that pertain to one of the two properties thus get invoked in support of claims about the other.

It will be helpful to draw a second preliminary distinction. When we pay deliberate attention to what mental state we are in, we are introspectively conscious of that state. This is different from the way in which mental states are conscious when we are not intentionally focusing on them. Introspection is attentive, deliberately focused consciousness of our mental states. It is relatively rare, and it is more elaborate than the way in which mental states are ordinarily conscious states. When we introspect, we are not only aware of what mental states we are in; we are aware of being thus aware. The difference between introspective and nonintrospective consciousness emerges vividly if we note that, when a state is nonintrospectively conscious, it normally has the potential to be an object of introspection; ordinary conscious states are normally introspectible. These differences form the basis of the argument I advance in section VI in support of my positive theory of state consciousness.

Consciousness is a notoriously puzzling topic. One way in which this is so has to do with how physical systems can be conscious, or how they can be in conscious states. We have a sense of ourselves that seems to make it difficult to understand how, as conscious selves, we could be located among the physical furniture of the universe.

But there is also an independent, and conceptually prior, sense of mystery simply about what consciousness is. This sense of puzzlement has to do mainly with the nonintrospective consciousness of mental states, rather than either creature consciousness or introspective consciousness. We know in general terms what it is for a creature to be conscious; it is conscious if it is awake and sentient. So there is no special mystery about what creature consciousness is.

By contrast, we lack even the sketch of a generally accepted answer to what kind of property the consciousness of mental states is. This lack of a schematic answer makes it difficult to say even what it could be for a mental state to be conscious. And the absence of even a general idea of

what such consciousness consists in has led some to see it as a primitive, unanalyzable property of mental states. Moreover, the consciousness of mental states is often held to have a number of special features, such as transparency, immediacy, and epistemic privilege of one sort or another. These considerations strongly reinforce whatever initial sense of mystery we may have about the nature of state consciousness.

As with creature consciousness, there is no particular mystery about introspective consciousness. Introspection is the attentive, deliberately focused consciousness of our mental states. So whatever puzzles we have about its nature are not special to it. There may well be problems about attention, but those are not puzzles about consciousness, *per se*. If we understood what it is for a mental state to be conscious, we would have at least the outline of an account of introspective consciousness.

The first puzzle, about how a physical system could be in conscious states, is often discussed independently of the problem about what it is for a state to be conscious. But this puts the cart before the horse. Only if we first understand what a conscious state is can we determine whether there is a real problem about how physical objects could be in such states, and if so exactly what that problem is.

The consciousness of states is in a certain way intermediate between the other two phenomena involving consciousness. Since only conscious creatures can be in conscious states, but the mental states of a conscious creature may not all be conscious states, state consciousness presupposes creature consciousness. Similarly, all states of which we are introspectively aware are conscious states, though not conversely.

II Are All Mental States Conscious?

It is sometimes held that all mental states are conscious states, indeed, that part of what it is for a state to be a mental state is that it be conscious.

This claim is central to the Cartesian concept of mind, and was forcefully articulated by both Descartes and Locke.² And even when it is not explicitly put forth, this idea figures centrally in the writing of many contemporary theorists.³ Whether or not it is true may well be the most important and pressing question about the nature of mind and consciousness.

This issue is connected to the distinction between creature consciousness and state consciousness. It might be held that only conscious creatures can be in conscious mental states. Moreover, to be conscious a creature must be in at least some mental states; so if all such states were conscious, conscious creatures would *perforce* be in conscious states. A creature's being conscious would then be coextensive with its being in conscious states, and the distinction between the two would arguably be idle.

It is far from obvious, however, that all mental states are conscious. For one thing, our ordinary picture of mentality plainly allows room for mental states that are not conscious states. We sometimes see that somebody wants something or thinks that something is so before that person is at all aware of that desire or thought. Similarly with emotions; we will occasionally recognize that we ourselves are sad or angry only after somebody else points it out to us. Subliminal perception and peripheral vision remind us that perceptual sensations can occur without our being aware of them. It is arguable that even bodily sensations such as pains can at times go wholly unnoticed, and so can exist without being conscious. When one is intermittently distracted from a relatively minor pain or headache, it is natural to speak of having had a single, persistent pain or ache during the entire period. It would be odd to say that one had had a sequence of brief, distinct, but qualitatively identical pains or aches.

Despite these considerations, there does exist an intuitive temptation to hold that mental states are invariably conscious. This should not, however, lead us to conclude⁴ that our commonsense concept of mind is irredeemably Cartesian. In

part, this temptation is due to the tendency noted earlier to run together the consciousness of states with that of creatures. If a creature's being conscious coincided with its mental states' being conscious, all its waking mental states would be conscious states. And it is possible to deny that apart from dreams, whose nature is notoriously far from clear, we have any mental states while asleep.⁵

Whatever commonsense pull there is toward thinking that all mental states are conscious states may also result in part from our relative lack of interest in nonconscious mental states. Some kinds of nonconscious mental states do interest us, because they help us understand other people's behavior even when they are unaware of being in those states. Still, our concern with the mental states of others is set in a social context that largely precludes our remarking on mental states of which they are unaware. So we tend not to pay explicit or conscious attention to such states. And of course we normally disregard whatever nonconscious mental states we ourselves may be in.

Even more telling, the tie between mental states and their consciousness is stronger with some types of state than with others. And the more plausible it is to hold that a particular kind of mental state must be conscious, the less interest we would have in cases that are not conscious. This temptation is perhaps strongest with bodily sensations such as pains and tickles, less compelling with emotions, and very likely weakest with intentional states such as thoughts and desires. Correspondingly, we have the least interest in nonconscious bodily sensations, whether our own or anybody else's, and far the most in nonconscious beliefs and desires, because of their role in explaining behavior. This variable lack of interest in the nonconscious cases is very likely a major source of the intuition that at least some kinds of mental state are invariably conscious.

One might try to sustain the idea that all mental states are conscious by insisting that nonconscious states would be mere dispositions. Only occurrent states could qualify as genuine mental

states and these, one might claim, would all be conscious.⁶

This suggestion conflates two ways of being dispositional. By 'dispositional mental state' one might mean a state that is simply a disposition to be in some occurrent mental state. But one might also mean a mental state whose nature we can usefully describe at least partly in dispositional terms. Being dispositional in the first way is incompatible with being occurrent, but being dispositional in the second way is not. Most states we describe dispositionally are dispositional in just this second way. Being flammable, for example, is both a disposition to burn and an occurrent, albeit possibly disjunctive, chemical property. Most mental states are like this. Since we detect nonconscious mental states by way of their dispositional connections to other things, we can assume that nonconscious mental states are dispositional in this second way. But that in no way implies that they are not also occurrent states.

Still, our intuitive reluctance to countenance nonconscious bodily sensations may persist. For one thing, we speak roughly interchangeably of feeling a pain or tickle or itch and of having the relevant sensation. And when we feel a pain or tickle or itch, is not that sensation automatically conscious?

Such terms as 'feeling' do carry this implication of consciousness; a felt pain is perforce a conscious pain. This is also true of something's hurting, and perhaps even of one's simply being in pain. But these things are not the same as having a pain, or a pain's existing. If we are intermittently unaware of a pain by being distracted from it, we feel the pain only intermittently; similarly with its hurting and our being in pain. Still, it is natural to speak of having had a pain that lasted throughout the day, and even to say that one was not always aware of that pain. This provides evidence that commonsense countenances the existence of nonconscious pains.⁷ Feeling pains and having them seem equivalent only because of our lack of interest in the nonconscious cases.

We cannot of course know what it is like to have a nonconscious pain or tickle or itch.⁸ But that is not relevant here. The reason we cannot know what it is like to have a nonconscious pain is simply that unless the pain is conscious there is no such thing as what it is like to have it. What it is like to be in pain, in the relevant sense of that idiom, is simply what it is like to be conscious of being in pain. None of this shows that nonconscious pains do not occur.

Reflection on what it is like to feel sensations does, however, suggest an important source for doubt about whether nonconscious sensations can occur. We classify sensory states and discriminate among their various tokens on the basis of what it is like for us to be in those states. This is true equally for bodily and perceptual sensations; we appeal to what it is like to be in pain and to what it is like to see red or hear a trumpet. And there is no such thing as what it is like to have these sensations unless the sensation is conscious.

One might conclude from this that there is no such thing as a sensation's having some distinctive sensory quality unless that sensation is conscious. If that were so, the very properties by reference to which we classify and discriminate among sensory states could occur only when those states are conscious. And if these qualities cannot occur except consciously, then plainly sensory states cannot either.

The premise of this argument is correct; we do classify and discriminate among sensory states by appeal to the conscious cases. But this hardly shows that the properties by reference to which we classify sensory states cannot occur nonconsciously. The distinctive qualities by means of which we type sensations form families of properties that pertain to color, visual shape, sound, and so forth. The members of these families resemble and differ from one another in ways that parallel the similarities and differences among the corresponding perceptible properties of physical objects. For example, the red sensory quality of visual sensations resembles the orange sensory quality of such sensations more than either re-

sembles the sensory green or blue of such sensations. The same holds for the distinctive qualities of pains. The qualities of being dull, stabbing, burning, or sharp resemble and differ in ways that reflect the similarities and differences among the corresponding physical objects and processes. Moreover, piercing and stabbing pains are both species of sharp pains, and typically result from piercing and stabbing objects or processes. Similarly, throbbing and pounding pains are species of dull pains.⁹

Sensory qualities are sometimes held to be problematic in some way that precludes this kind of account. Common sense does not support this. Sensory qualities are simply those properties in virtue of which we distinguish among our sensations. They are the properties in virtue of which sensory states differ from one another in respect of sensory content. We have access to these properties in the first instance by way of the conscious cases, but that by itself does not show that there is anything problematic about the properties themselves.

Nor is there reason to hold that these similarities and differences cannot obtain except when the sensation in question is conscious. So we have no basis to deny that sensory qualities can occur nonconsciously. The distinctive sensory properties of nonconscious sensations resemble and differ in just the ways that those of conscious sensations resemble and differ. They diverge only in that one group is conscious and the other not. We do rely on conscious sensations in typing sensory states. Indeed, we fix the extensions of our terms for the various kinds of sensation by way of the conscious cases, both our own and those of others. But this in no way shows that all sensory states are conscious states.¹⁰

Our goal is to understand and thereby disarm whatever intuitive doubts we have about whether mental states can occur nonconsciously. In addition to the concern we may have about sensory states specifically, doubts may also linger about mental states in general, whether sensory or intentional. And the foregoing considerations will

not help with intentional states. Can we explain whatever temptation exists to hold that mental states generally cannot occur nonconsciously?

Perhaps the most important source of such temptation stems from our intuitive sense that mental phenomena are in some special way different from everything else. And being conscious may well seem to be the only way to demarcate the distinctively mental. The problem is pressing because no other single mark will do for all mental phenomena. Intentional and sensory properties constitute the most likely candidate; all mental states have one or the other. But mark neither covers all mental states, and the two kinds of property have so little in common that using them disjunctively may well seem artificial.¹¹

Here again it will help to distinguish what fixes the extension of a term from those properties which hold of all the items in that extension. We can satisfy our need for a uniform mark of the mental by having a single property in terms of which we fix the extension of 'mental'. We need not also have a single property that all mental states exhibit. As a way of fixing this extension, consciousness very likely does provide a single mark of the mental. We fix the extension of 'mental' by way of conscious mental states, even though not all mental states are conscious.

Nonconscious mental states apart, there are commonsense reasons to prefer a mark of the mental based on intentional and sensory properties rather than consciousness. For one thing, the way we are aware of conscious mental states resembles the way we are aware of such bodily states and events as the positions of our limbs and perturbations of our viscera. In both cases our awareness is intuitively unmediated and strikingly unlike the way we may know about such states and events in others. So using consciousness as a mark of mental states may leave unexplained why those bodily states and events are not mental. Using intentional and sensory properties to demarcate the mental excludes such cases.¹²

Is a first-person perspective somehow essential to mental states? Is it impossible to grasp the na-

ture of those states from a third-person point of view? This seductive idea is sometimes invoked in support of the suggestion that being conscious is an inseparable aspect of what it is to be a mental state, or perhaps a mental state of a particular kind. Since the very existence of a first-person viewpoint depends on a state's being a conscious state, consciousness would then be intrinsic to its nature.¹³

But any serious theory of mental states must capture both their first-person and third-person dimensions, and do so without shortchanging either. If you feel pain or think that it's raining, your first-person viewpoint provides access to the very same states which I may know about from my third-person perspective. Since this third-person knowledge can be independent of those states' being conscious, consciousness cannot be essential to the state you may be able to grasp from your first-person point of view.

This conclusion does not diminish the importance of consciousness. We fix the extension of our terms not only for sensations but for mental states generally by reference to the conscious cases; so our first-person access to mental states figures in fixing those extensions. It does not follow that the first-person perspective is essential to the states thus determined.

Still, one might urge that without some essential tie to consciousness, the genuine intentionality of mental states would be indistinguishable from the merely derivative intentionality of speech.¹⁴ It is far from clear that any such difference of status does distinguish the intentionality of thought from that of speech.¹⁵ But we can in any case describe the distinction without reference to consciousness. The intentionality of speech is derivative, if at all, because speech acts are caused by mental states with the same content and a mental attitude parallel to the illocutionary force of the speech acts. Since mental states are not caused by anything that has the same content and mental attitude, their intentionality is in this way not derived.¹⁶

III Can Consciousness Be Explained?

On balance, then, common sense rejects the claim that all mental states are conscious states. But there are also serious theoretical reasons to reject that claim. It turns out that if all mental states were conscious states, it would be impossible to give any nontrivial, informative explanation of the nature of state consciousness.

An explanation of what it is for mental states to be conscious either will itself appeal to mental states or it will not. Suppose now that all mental states are conscious, and that our explanation of what it is for mental states to be conscious does invoke mental states. Such an explanation will be circular, since the appeal to mental states is then automatically an appeal to conscious states. Invoking the very phenomenon we want to explain trivializes the explanation and prevents it from being informative.¹⁷

Alternatively, our explanation might appeal only to what is nonmental. Some mental phenomena can, perhaps, be fully explained in that way. To give such an explanation is, in effect, the project of naturalizing the mental.¹⁸ Many would accept the possibility of an essentially nonmental explanation of what it is for states to have intentional properties, at least in the case of nonconscious intentional states. And given some such account as that sketched in the previous section, perhaps the same is true of sensory states as well.

But it is far less likely that we can successfully do this in any direct way with consciousness. Nothing in nonmental reality seems suited to explain what it is for a mental state to be conscious; as Thomas Nagel insists, the gulf between mental and physical seems unbridgeable primarily in respect of consciousness.¹⁹ Moreover, the consciousness of mental states is the most sophisticated and difficult to understand of any mental phenomenon; so such consciousness is unlikely to yield directly to explanation in nonmental terms. Both intuitively and theoretically, the opposition between consciousness and matter seems to defy

the possibility of directly explaining one in terms of the other.²⁰

To those with Cartesian leanings, this picture will seem exactly right. We cannot explain consciousness in terms of what is not mental, and explaining consciousness in terms of conscious states will be trivial and uninformative. We must therefore acquiesce in our inability to give any genuine explanation of consciousness at all; at best we can render the phenomenon more intelligible by tracing the merely conceptual connections that hold among such cognate notions as mind, consciousness, subjectivity, viewpoint, first-person perspective, and self.²¹

But the dichotomy is false. Since mental states are not all conscious, there is a third strategy. We can explain what it is for mental states to be conscious by appeal to mental states that are not, themselves, conscious states. We bridge the intuitive gulf between consciousness and matter by explaining consciousness in terms of states that are mental, but not conscious.²²

This strategy not only preserves the possibility of an informative explanation of consciousness; it also holds out hope for a naturalist account of mind. If we can explain consciousness by appeal to states that are mental but not conscious, perhaps we can in turn explain those nonconscious mental states in nonmental terms. This tiered picture would sustain naturalism. We still would not explain consciousness directly in nonmental terms, but the explanatory chain would reach nonmental reality by way of some nonconscious but mental intermediate. The intermediate step would dispel the apparent unintelligibility of explaining consciousness straightaway in nonmental terms.²³ Since it explains conscious states in terms of nonconscious mental states, moreover, this approach would vindicate the division of labor described at the outset. But naturalism is not the main issue here. Whatever the case about that, no nontrivial, informative explanation of state consciousness will be possible unless we reject the idea that all mental states are conscious.

As we have seen, the idea that all mental states are conscious is considerably more compelling with sensory states than with intentional states. So one might concede that nonconscious intentional states exist, but insist that nonetheless sensory states are all conscious. Nothing so far shows that this modified position would preclude an informative explanation of consciousness. So we might try to salvage this much of the Cartesian position.

Once we drop the idea that mental states are all conscious, what reason could we have to insist that all sensory states are? The considerations of section II suggest an answer: The distinctive sensory quality of sensory states can exist only if those states are conscious. If so, being conscious would be part of what it is to be a sensory state. It would be intrinsic to having sensory character, and thus intrinsic to being a sensation in the first place. Some such line of thought very likely underlies the idea that sensory states cannot occur nonconsciously.

A property is intrinsic if something's having it does not consist, even in part, in that thing's bearing some relation to something else. If being conscious is at least partly relational, a mental state could be conscious only if the relevant relation held between the state and some other thing. Because the other thing would be a distinct existence, it would be natural to conclude that being conscious is a contingent property of mental states; any particular conscious state could be the very same state and yet not be conscious. Some special reason would be needed to avoid that conclusion. So, if all sensory states were conscious, the best explanation of that generalization would be that being conscious is an intrinsic property of sensory states. Similarly with the more general claim that all mental states, of whatever sort, are conscious.

If mental states are all conscious states, we cannot explain what it is for mental states to be conscious. This difficulty also arises if being conscious is an intrinsic property of mental states, even if not all mental states are conscious.

No informative explanation of state consciousness is possible unless we can represent it as having some articulated structure. But it will be hard to justify the idea that being conscious is an intrinsic property of mental states if that property does have some informative structure. Once an explanation assigns such structure, it will be equally plausible to regard being conscious as an extrinsic property of mental states.²⁴ So the only reason to see consciousness as an intrinsic property of mental states will be that it lacks such structure, and is therefore simple and unanalyzable. And its being simple would effectively preclude our explaining it by appeal to anything else. Simple properties are those we take to be primitive in our hierarchies of explanation.²⁵

The reason for insisting that being conscious is intrinsic to sensations is its supposed inseparability from sensory quality. On this view, the property of being conscious is not only intrinsic to sensations, but it is actually an intrinsic aspect of the qualitative properties that distinguish types of sensations.²⁶ This would mean the failure of the division of labor envisaged at the outset between sensory character and consciousness.

The inseparability of sensory quality and consciousness would also exacerbate the difficulty about simplicity. Not only would the property of being conscious be simple and unanalyzable; it would also include the relevant sensory quality. There seems no hope on this model for any useful, informative explanation of what it is for a sensory state to be conscious.

These considerations reinforce the earlier argument that no satisfactory explanation of state consciousness is possible if mental states are all conscious. Even if all mental states are indeed conscious, perhaps we cannot preclude the possibility of a noncircular explanation that appeals to conscious mental states, but not to the property of their being conscious (see note 17). Only if being conscious were an intrinsic property of mental states would we have reason to hold that mental states are invariably conscious. And it will be arbitrary to hold that being conscious is an in-

trinsic property unless that property lacks the articulated structure necessary for an informative explanation.

The idea that being conscious is an intrinsic property has strong intuitive appeal. A parallel may be useful. On the pre-Galilean idea, bodies move toward a natural resting place, and having a particular natural resting place is an intrinsic property of every kind of body. This conception is intuitively natural; it is inviting to see bodies as intrinsically tending toward upward or downward movement.²⁷ Still, we get vastly more accurate and powerful explanations of bodily motions if we see a body's tendency to move in terms of its relations to other bodies. Similarly, it is pretheoretically tempting to see the property of being conscious as intrinsic to sensations, or perhaps to all mental states. Nonetheless, it is only if we regard being conscious as a relational property that we can explain what such consciousness consists in.

IV Transitive and Intransitive Consciousness

To be able to explain what it is for a mental state to be conscious, we must reject the claim that all mental states are conscious as well as the related idea that being conscious is an intrinsic property. We must see consciousness as a property of only some mental states, and as a relational property of whatever states have it. Does any reasonable account of consciousness meet these conditions?

Putting creature consciousness to one side, we can distinguish two ways in which we use the word 'conscious'. One is when we speak of our being conscious of something. Because of the direct object, I shall call this the transitive use.

But we also apply the term 'conscious' to mental states, to say of them that they are conscious states. This is what I have labeled state consciousness. The lack of a direct object suggests calling this the intransitive use. This intransitive use figures only when talking about mental states, whereas we speak of being conscious, transitively,

of mental and physical things alike. We may be transitively conscious of a stone or a symphony, or of a mental state.²⁸

We want an account of what it is for mental states to be intransitively conscious on which that property is relational and not all mental states are conscious. The natural suggestion is to explain intransitive consciousness in terms of transitive consciousness. As a first try, we can say that a mental state is intransitively conscious just in case we are transitively conscious of it.²⁹

Explaining intransitive by way of transitive consciousness would be circular if transitive consciousness were a type of state consciousness. But transitive consciousness can occur without intransitive state consciousness. One is transitively conscious of something if one is in a mental state whose content pertains to that thing—a thought about the thing, or a sensation of it. That mental state need not be a conscious state. And if, as is likely, mental states are possible during sleep, transitive consciousness will not even presuppose creature consciousness.

Explaining intransitive in terms of transitive consciousness squares reasonably well with our pretheoretic intuitions about what it is for a mental state to be conscious. Plainly, if a state is a conscious state, we are transitively conscious of it. But the converse claim requires a bit of qualification. One can be conscious of being in a mental state even when we would not count that state as being a conscious state. We regard mental states as conscious states only if we are transitively conscious of them in some suitably unmediated way.

Suppose I am angry, but unaware of it, and suppose that you see this from my behavior, and tell me. I trust your judgment, but still consciously experience no anger. In this case I am conscious that I am angry even though my anger is not a conscious state. Or suppose that I myself notice my angry behavior and infer that I am angry, but again experience no conscious feeling of anger.

These kinds of case require only a minor adjustment. We must specify that our transitive consciousness of our mental state relies on neither inference nor observation.³⁰ Mental states are intransitively conscious just in case we are non-inferentially and nonobservationally conscious of them.

Conscious states are states we are conscious of, without benefit of inference or observation. But suppose that inference can be nonconscious,³¹ and that such inference sometimes figures in our becoming conscious of our mental states. Common sense would say these are cases of conscious states, but our excluding all reliance on inference rules them out.

The natural rejoinder is to require that conscious states be those we are **intransitively** conscious of without benefit of inference or observation in the ordinary sense. The trouble is that the natural way to understand that ordinary sense is circular; ordinary inference is conscious inference. The proposal about transitive and intransitive consciousness disarms this difficulty. Ordinary inference is inference of which we are transitively conscious. So conscious mental states are mental states we are transitively conscious of, but without relying on any inference or observation of which we are transitively conscious.

This proposal captures our commonsense intuitions. It also has the advantage, at least on the most natural construal, of providing a relational account of the property of being a conscious state. Bracketing the qualifications about inference and observation, a state is intransitively conscious if one is transitively conscious of it. And being transitively conscious of something is a relation that a person or other creature bears to that thing. Accordingly, a conscious mental state is a compound of two things: the mental state, which by itself is not conscious, and one's being transitively conscious of it. So being in the state and being transitively conscious of it are distinct, and the property of being conscious is not intrinsic to the state itself. The proposal thus avoids the problems raised earlier about explaining consciousness.

Care is necessary on this point. If by 'conscious state' one means the state, which as it happens is conscious, then the property of being conscious is not intrinsic to conscious states. But one might instead mean by 'conscious state' to refer to a state together with the property of its being conscious. Then being conscious is of course intrinsic to the state. Being intrinsic in this way is trivial, since one can in this way make any property intrinsic. It is also question begging, since only if one assumed that being conscious was intrinsic to conscious states would one use 'conscious state' in that way.

The distinction between transitive and intransitive consciousness also dispels one widespread reason for insisting that intransitive consciousness is an intrinsic property. Being conscious is often thought of as a reflexive property of conscious states, as though such states were somehow conscious of themselves. Ryle vividly captures this idea with his disparaging metaphors of conscious states as "self-intimating" or "self-luminous."³²

It is doubtful that any states have the epistemic privilege that self-intimating states would presumably exhibit. But whatever the case about that, explaining what it is for a mental state to be conscious in terms of our being conscious of those states undercuts the metaphor of a state's being conscious of itself. Being transitively conscious of something is a relation that a person or other creature bears to that thing. So only creatures can be transitively conscious of things. A mental state may well be that state in virtue of which somebody is conscious of a thing, but the state cannot itself literally be conscious of anything.

One might object that being transitively conscious of things does not in general make them intransitively conscious. My being conscious of a stone does not make it conscious. What is relevantly different in the case of mental states? Why should being transitively conscious of such states make them conscious states?³³

We must distinguish two ways of understanding the present proposal. Being transitively conscious of a mental state does in a sense make it intran-

sitively conscious. But that is not because being conscious of a mental state causes that state to have the property of being intransitively conscious; rather, it is because a mental state's being intransitively conscious simply consists in one's being transitively conscious of it. The mistake here is to suppose that a state's being intransitively conscious is an intrinsic property of that state. If it were, then being intransitively conscious could not consist in one's being transitively conscious of being in that state unless being thus conscious induced a change in that state's intrinsic properties. This objection is at bottom just a disguised version of the doctrine that being intransitively conscious is an intrinsic property.

The foregoing considerations help explain the intuitive appeal of the idea that being conscious is an intrinsic property. For one thing, one-place predicates typically suggest nonrelational properties; so speaking of mental states as intransitively conscious doubtless encourages this picture. An even more telling point derives from the need for our transitive consciousness of conscious states to be independent of inference and observation. Because of this independence, we normally have no idea how we come to be conscious of those states and thus, on this proposal, no idea of why those states are intransitively conscious. Since we cannot trace this consciousness to anything, it can seem tempting to suppose that a state's being intransitively conscious is something intrinsic to the state itself.

It is tempting to suppose that consciousness reveals its own nature. How better, one might think, to learn about what being conscious consists in than from the very dictates of consciousness? And, since consciousness does not reveal itself as an extrinsic property of conscious states, can we not conclude that it is an intrinsic property? But the deliverances of consciousness are compatible both with its being an intrinsic property and with its being an extrinsic property. And there is no reason to think that if it were in fact intrinsic, that would help explain the intuitive appeal of the idea that it is intrinsic; the best ex-

planation of that intuition is that consciousness itself is mute about its own nature. Our intuition that it is intrinsic, therefore, is arguably idle. Indeed, if consciousness gave us privileged knowledge about what mental states we are in, it might still provide no special insight into its own nature.

V A Theory of Consciousness

On the foregoing proposal, a mental state's being intransitively conscious is simply one's being transitively conscious of that state in a certain way. What kind of positive account might be given of how we are transitively conscious of our conscious mental states?

Cases of being conscious of things fall into two broad categories. One is perceptual; we are conscious of things when we see or hear them, or perceive them in some other way. The other way we are conscious of things has to do with thinking; when we think about something, we are conscious of that thing. Which of the two ways is needed here?

There are analogies between sense perception and the way mental states are conscious that may encourage the adoption of a perceptual model. For one thing, the various sense modalities each enable us to respond differentially to a distinctive range of stimuli. Because being conscious of our mental states is an ability to respond differentially to the mental states we are in, it may seem somewhat like a sense modality dedicated to mental states. For another thing, some sensory modalities exhibit a characteristic directness. The sensory mechanisms subserving taste, touch, proprioception, and enteroceptive perception enable each of these senses to operate without any intervening medium. Given the apparent immediacy of our consciousness of our mental states, these sense modalities may seem to be apt models for such consciousness.

These two analogies are important in the way we discriminate the several sensory modalities.

Each modality has a characteristic range of properties or stimuli to which it responds, and a characteristic sense organ or sensory mechanism by means of which it operates. But other commonsense intuitions suggest a perceptual model as well. Sensory quality and the consciousness of mental states both seem intuitively inexplicable, as though the two must somehow be emergent phenomena. That the two seem mysterious in similar ways may make it inviting to adopt a single model for both. Another apparent similarity is that the consciousness of our mental states seems to occur organized into a kind of field, somewhat like the fields that characterize the various sensory modalities.³⁴

However, the perceptual model does not withstand scrutiny. Whereas a range of stimuli is characteristic of each sensory modality, mental states do not exemplify a single range of properties. Rather, as noted in section II, mental states exhibit intentional or sensory character, and these two have little in common; nor do the sensory qualities special to the various modalities resemble each other.³⁵ The other analogy fares no better. Doubtless some central neural process subserves our being conscious of our mental states, and presumably that process is responsible for the apparent immediacy in the way we are conscious of those states. But there is no reason to suppose that the operation of that process explains that apparent immediacy in anything like the way that the mechanisms of sensory modalities like touch and proprioception do explain the apparent immediacy of such perception.

The perceptual model fails even apart from these points. Perceiving something involves the occurrence of some sensory quality, which in standard circumstances signals the presence of that thing. If our being conscious of a mental state is like perceiving something, our being conscious of it will involve the occurrence of some mental quality; otherwise the analogy with perception will be idle.

What might that mental quality be? In the case of our being conscious of a sensory state, it is

tempting to hold that the mental quality in virtue of which we perceive that state is the sensory quality of the sensation in question. But this answer is theoretically unmotivated. When we see a tomato, the redness of our sensation is not literally the same property as the redness of the tomato. If being conscious of a sensory state is like perceiving that state, why should the quality involved in our being conscious of that state be the same quality as that of the state itself? Moreover, this answer, by reducing the quality involved in our being conscious of a state to the quality of that state, in effect reduces our being conscious of the state to our simply having that state. But that results in consciousness being intrinsic to such states,³⁶ and the point of the perceptual model was to avoid that conclusion.

But, if being conscious of a mental state is like perceiving, there is no other remotely plausible candidate for the mental quality involved in being conscious of a sensory state.³⁷ The difficulty is even more acute with intentional states; being conscious of them seems plainly to involve no mental quality whatever. Moreover, even when the conscious state is sensory, it is the quality of that state we are conscious of; so how could the sensory quality of a higher-order state explain our being conscious of the lower-order quality? We must therefore reject the perceptual model of how we are transitively conscious of our conscious mental states.

Nor do the two intuitive factors help. As already argued, the mystery that seems to attend our mental states being conscious is due largely to our conceiving of such consciousness as an intrinsic property. And only when the states that are conscious occur in a sensory field does the consciousness seem to come organized in a field. In such cases as emotions and thoughts, the property of their being conscious resists being so located.

The other model for transitive consciousness, which relies on thinking, encounters no such difficulties. Thinking is not specific to a particular range of objects, and it involves no dedicated or-

gan in the way sense modalities do. Moreover, the relation between thinking and its object is intuitively at least as unmediated as that between perceiving and the object of perception. Nor does having thoughts about things involve any characteristic qualities. It seems, therefore, that such a model provides the best hope for explaining what it is for mental states to be conscious in terms of our being conscious of those states.

We are conscious of something, on this model, when we have a thought about it. So a mental state will be conscious if it is accompanied by a thought about that state. The occurrence of such a higher-order thought (HOT) makes us conscious of the mental state; so the state we are conscious of is a conscious state. Similarly, when no such HOT occurs, we are unaware of being in the mental state in question, and the state is then not a conscious state. The core of the theory, then, is that a mental state is a conscious state when, and only when, it is accompanied by a suitable HOT.

Because being conscious of something is factive, using HOTs to explain the relevant transitive consciousness may seem less plausible than a perceptual model. After all, perceiving something is also arguably factive, whereas having thought is not. This should not lead one to adopt the perceptual model, however; if the relevant consciousness really is factive, we can stipulate that for our HOTs. In any case, there is good reason to doubt that the way we are conscious of our conscious mental states guarantees truth; special views about privileged access notwithstanding, we can and do make mistakes about what conscious states we are in.³⁸

When a mental state is conscious, it is not simply that we are conscious of the state; we are conscious of being in that state. This places constraints on what the content of these HOTs must be; their content must be that one is, oneself, in that very mental state. Independent considerations point to the same conclusion. One cannot think about a particular mental-state token, as opposed to thinking simply about a type of mental state, unless what one thinks is that some in-

dividual creature is in that mental state. So HOTs will not be about mental-state tokens unless their content is that one is, oneself, in the mental state.³⁹

How demanding are these requirements? One way to answer this is to see whether creatures without language can have such thoughts. Language is not required in order for a creature to be able to have any thoughts at all.⁴⁰ To have thoughts a command of some concepts is of course necessary. But those conceptual abilities need not be anything like as powerful as ours, nor need the concepts be nearly so fine grained. More important, a creature's thoughts need not be articulated into syntactic components. Just as people can deploy one-word sentences, so thoughts can occur without syntactic structure.

We would not, however, be warranted in interpreting a creature's behavior as the using of sentences or the expressing of thoughts unless we had reason to see that behavior as involving affirmation or denial. To do that we must be able to distinguish affirmations and denial in the creature's behavior, and perhaps other illocutionary acts as well. This suggests that the ability to have thoughts may depend more on a creature's being able to have some minimal variety of mental attitudes, such as believing and desiring, than on its having resources for fine-grained content with some compositionally articulated structure.

But is language necessary for a creature to have HOTs with the content required by the present theory? HOTs refer both to oneself and to one's mental states; can a nonlinguistic creature have thoughts with sufficiently sophisticated content? A minimal concept of self will suffice for reference to oneself; no more is needed than a concept that allows distinguishing between oneself and other things. Such a minimal concept need not specify what sort of thing the self is. Thus it need not imply that the self has some special sort of unity, or is a center of consciousness, or is transparent to itself, or even that it has mental properties.⁴¹

Nor is there reason to suppose that rich conceptual resources are necessary for a thought to refer to one's own mental states. We refer in

thought to physical objects by way of their position in our visual field. It is natural to suppose that a thought can similarly refer to sensory states by way of their position in the relevant sensory field. In any case, conscious differentiation of sensory detail quickly outstrips one's conceptual resources; so some such means of referring to sensory states is necessary.

Doubtless one cannot thus refer in thought to all the sensory states to which richer conceptual resources would enable us to refer. Nor can one's HOTs refer in that way to intentional states that lack suitable perceptual content. Richer conceptual resources would thus expand the range of one's conscious states. Though this may initially seem surprising, on reflection it is plain that this happens in our own experience. Having more fine-grained conceptual distinctions often makes us aware of more fine-grained differences among sensory qualities. Vivid examples come from wine tasting and musical experience, where conceptual sophistication seems to generate experiences with more finely differentiated sensory qualities.

The term 'thought' sometimes has a generic sense, in which it applies to any propositional mental state, regardless of its mental attitude. On this construal, such diverse states as believing, hoping, expecting, desiring, wondering, and suspecting all count as thoughts. Will any propositional state make a mental state conscious, as long as it has the right content?

The requisite content is that one is, oneself, in a particular mental state. If one doubts or wonders whether one is in a particular mental state, or desires, hopes, or suspects that one is, that plainly will not make the state a conscious state.

Still, we often are conscious of things by virtue of such states as doubting and wondering. For example, if I doubt or wonder whether some physical object is red, I am conscious of the object; similarly if I expect, hope, or desire that it is. So perhaps if one doubts or wonders whether a mental state has some particular property, or hopes, desires, or expects that it does, one will thereby be conscious of the mental state.⁴²

But it is not the doubt, wonder, hope, or desire that makes us conscious of the object. If I doubt whether that object is red, or desire or suspect that it is, I must at least think assertorically that the object is there. Similarly with doubting, hoping, or expecting that my mental state has some property; I must at least have the assertoric thought that I am in that state. Similarly, we must not suppose that these nonassertoric attitudes will make one conscious of one's being in that state unless having the attitude also leads one to have an affirmative thought that one is in that state.⁴³ The same holds of such states as anger and pleasure, when these are construed propositionally.

Nor will one's mental states be conscious if accompanied merely by a dispositional higher-order mental state. Being disposed to have a thought about something does not by itself make one conscious of it.⁴⁴ To be transitively conscious of something, therefore, we must have a thought about it in a relatively narrow sense: It must be an assertoric, occurrent propositional state.⁴⁵

The HOT theory sustains in one direction the division of labor proposed at the outset. Being conscious and having sensory character are totally independent on this theory. Moreover, since being conscious is not intrinsic to intentional states, we can study intentionality independently of consciousness. And, although the HOT theory does presuppose that we understand intentionality, it remains neutral about the main issues that divide theorists about intentional states.

On the HOT theory, every conscious mental state is accompanied by a HOT about that state. This may seem hard to accept; after all, we are seldom aware of such HOTs. But the theory actually predicts that we would not be. A mental state is conscious only if it is accompanied by a HOT. So that HOT will not itself be a conscious thought unless one also has a third-order thought about the second-order thought. And if our second-order thought is not a conscious thought, we will be unaware of it. Only if one assumed that all mental states are conscious would one urge that being unaware of a state shows it is not there.

Given the multitude of conscious states we are in during most of our waking life, is it plausible to suppose we have such a great number of HOTs? We do, in any case, have very many thoughts that are not conscious. There are vast numbers of nonconscious first-order thoughts by means of which we negotiate our way through the world and adjust both our plans and our conscious trains of thought. Why should there not also be a multitude of nonconscious HOTs? Moreover, we may need fewer than might at first appear. The content of HOTs may typically be reasonably specific for mental states that are near our focus of attention. But it is unlikely that this is so for our more peripheral states. For example, the degree of detail we are conscious of in our visual sensations decreases surprisingly rapidly as sensations get farther from the center of our visual field. It is natural to suppose that the content of one's HOTs becomes correspondingly less specific, and that a progressively smaller number of HOTs will refer to successively larger portions of the visual field.⁴⁶

The nonconscious character of most HOTs leads, however, to another objection. HOTs are supposed to make the mental states they are about conscious. How can nonconscious HOTs do this? How can HOTs be a source of consciousness if they are not themselves conscious?⁴⁷

This objection disregards the distinction between transitive and intransitive consciousness. HOTs confer intransitive consciousness on the mental states they are about because it is in virtue of those thoughts that we are transitively conscious of those mental states.

The situation emerges vividly when we consider a certain ambiguity of the phrase, 'state of consciousness'. A state of consciousness can be a conscious state, that is, a state one is conscious of being in. Or it can be a mental state in virtue of which one is conscious of something. Adapting our terminology, we can call these intransitive and transitive states of consciousness, respectively. For a mental state to be conscious is for it to be a state one is conscious of being in. So a

HOT can be a source of consciousness for the mental state it is about because the HOT is a transitive state of consciousness; it does not also need to be an intransitive state of consciousness.

One might insist that this misunderstands the objection about a source of consciousness; such a source must be conscious in just the way conscious states are conscious. But without some independent substantiation, this claim begs the question at hand, since on the HOT theory a state's being intransitively conscious just is one's being transitively conscious of it in a suitable way. More important, the objection so understood is simply a version of the claim that we can have no noncircular explanation of state consciousness, since a satisfactory explanation would itself have to appeal to state consciousness.

But are transitive and intransitive states of consciousness really distinct in this way? Or must every transitive state of consciousness be an intransitive state of consciousness as well? Suppose I see or think about some physical object, and am thereby conscious of it. Must not my visual experience or thought be a conscious experience or thought?

The cases that come immediately to mind do indeed work that way, but often this connection does not hold. To make one's way when driving or walking somewhere, one must be visually conscious of many obstacles. But one may well be wholly unaware that one is conscious of those obstacles if, for example, one is immersed in conversation. Other cases abound. The so-called cocktail-party effect occurs when one screens out the sounds of conversations other than one's own. Still, if one's name is mentioned in a conversation one had screened out, one's attention immediately shifts to that conversation, showing that one must have had auditory consciousness of what was being said. Such cases do not immediately occur to us precisely because in these cases we are not aware of being conscious of these things.⁴⁸

Perhaps one reason why the two notions often fail to be distinguished may be that all intransitive states of consciousness are also transitive states

of consciousness. To be a conscious state implies being a mental state, and mental states are all states of being conscious of something or other. But the two notions still apply differently; although all intransitive states of consciousness are transitive states of consciousness, the converse does not hold.

Must the relationship between mental state and HOT be closer than mere accompaniment? In particular, must we postulate a causal connection between the two? It is natural to suppose that the mental state must in some way be implicated in causing the HOT. But requiring that the mental state be the principal factor in causing the HOT is very likely too strong.⁴⁹ For one thing, mere accompaniment is all one needs to spell out the proposal to explicate intransitive in terms of transitive consciousness. Moreover, it is not uncommon that a particular mental state is sometimes conscious and sometimes not. In these cases some causal factor other than the mental state must figure in explaining why the HOT does or does not occur. It is possible that the mental state does by itself cause the HOT while other factors, such as mental distractions, intermittently block that causal link. But we have no reason to prefer that model to one in which other factors, such as the mental state's immediate mental environment, contribute to causing the HOT.⁵⁰ In any case, since mere accompaniment suffices to explain our being conscious of our conscious states, we can remain noncommittal about the causal history of HOTs.

One motive for supposing that mental states will be implicated in causing HOTs is a desire for a mechanism to inhibit the occurrence of HOTs in the absence of the mental states they are about. But the occurrence of such HOTs is not a difficulty so long as they do not occur too often. Indeed, there is good reason to think that we sometimes in effect invent the mental states that we take ourselves to be in.⁵¹ It is natural to interpret such confabulation as the occurrence of suitable HOTs in the absence of the mental states they purport to be about.

Strictly speaking, having a HOT cannot of course result in a mental state's being conscious if that mental state does not even exist. Our ordinary notion of transitive consciousness reflects this fact; since being conscious of something is factive, we cannot be conscious of something that is not there. Still, a case in which one has a HOT along with the mental state it is about might well be subjectively indistinguishable from a case in which the HOT occurs but not the mental state. If so, folk psychology would count both as cases of conscious states.

It is sometimes argued that one can do justice to our pretheoretic intuitions about consciousness only on the assumption that being conscious is intrinsic to mental states. This is not so. The HOT theory can save the phenomenological appearances at least as well as one can by assuming that consciousness is an intrinsic property, and very likely more successfully. Several examples will illustrate this.⁵²

When we focus on a particular conscious state, what we focus on is the state we are conscious of; it is notoriously difficult to isolate that aspect of the situation in which its being conscious consists. The consciousness of mental states is somehow transparent, or diaphanous.⁵³ The idea that consciousness is intrinsic seems superficially to help; perhaps it would be hard to isolate the property of being conscious if that property were intrinsic to the state itself. We get a more satisfactory explanation from the HOT theory. A conscious state's being conscious consists in there being an accompanying HOT. Since that HOT is not usually a conscious thought, we normally are unaware of anything that could count as the property of the mental state's being conscious.

A strong intuitive connection obtains between being in a conscious state and being conscious of oneself. Again, the idea that consciousness is intrinsic may seem to help, by implying some sort of self-reference. This is a confusion. If being conscious were intrinsic to mental states, perhaps those states would refer to themselves, but what is needed is reference to the self, which is different.

Again, the HOT theory explains the phenomenon more successfully, since one's HOTs must refer to oneself.

Sometimes the idea that being conscious is intrinsic to mental states makes it actually harder to save the appearances. Some mental states seem to be conscious only some of the time, largely through shifts in attention; examples considered earlier are pains or auditory sensations from which we are temporarily distracted. It is doubtful that we can explain such shifts between a state's being conscious and not being conscious if the property of being conscious is intrinsic to those mental states which are.⁵⁴

There are, finally, phenomena that seem to point directly to the HOT theory, with which no other theory could deal at all. One has to do with sensory qualities, which supposedly lend intuitive support to the idea that consciousness is intrinsic. As noted earlier in this section, we are often aware of more fine-grained differences among sensory qualities when we have more fine-grained conceptual distinctions. The HOT theory predicts this. The degree to which we are conscious of differences among sensory qualities is a function of how fine-grained the concepts are that figure in our HOTs. It is difficult to see how a theory that does not rely on HOTs could explain this role that concepts have.⁵⁵

As noted at the end of section IV, we can explain the apparent force of the idea that being conscious is intrinsic to conscious states as due to our normally having no idea how we come to be conscious of them. The HOT theory enables us to expand on that explanation. Because the HOTs are typically not conscious thoughts, we are unaware of what the consciousness of our mental states consists in. This not only makes the property of being conscious seem diaphanous; it also makes it tempting to hold that such consciousness is intrinsic to those states. The sense that a state is inseparable from our consciousness of it may be further strengthened by the reference the HOT makes to that state. But this inseparability is of course merely notional; it is the inseparability

of a thought from its object. So it cannot help show that being conscious is intrinsic to mental states.

VI The Argument from Introspection

In closing I turn briefly to two arguments that provide independent support for the HOT theory. As noted in section I, introspection involves more than a mental state's being a conscious state. Introspecting a mental state is deliberately and attentively focusing on that state. Nonintrospective consciousness, by contrast, requires no special act of attention. Every introspected state is therefore a conscious state, but not conversely.

Mental states are conscious just in case they are introspectible.⁵⁶ All and only conscious states can become subjects of introspective consciousness by a more or less deliberate shift of attention. This equivalence of being conscious with being introspectible may lead to eliding the difference between introspective and nonintrospective consciousness.

If a mental state's being conscious consists in our being conscious of being in that state, what more occurs when we introspect that state? Focusing on a mental-state token means focusing on somebody's being in that state; so introspecting one's own mental state implies deliberately focusing on one's being in that state. And that involves not merely being conscious of being in that state, but being actually aware that one is thus aware. Introspection is a kind of higher-order consciousness; it is the transitive consciousness of being conscious of one's mental states.⁵⁷

The HOT theory readily accounts for the difference between introspective and nonintrospective consciousness. A mental state is nonintrospectively conscious when accompanied by a relevant HOT; introspection occurs when there is a third-order thought that makes the second-order thought conscious. Introspective consciousness is the special case of conscious states in

which the accompanying HOT is itself a conscious thought.⁵⁸

HOTs are sometimes invoked to explain introspective consciousness; introspecting a mental state is, on such an account, one's having an accompanying thought about that state.⁵⁹ But when we introspect a mental state, the HOT that accompanies the state is always conscious; we are conscious of thinking about that state. The idea that introspection consists simply in having HOT about the introspected state must thus rely on the tacit assumption that HOTs are all conscious. So that view would seem plausible only if one already held that mental states are automatically conscious.

In any case, the idea that HOTs can explain introspection seems to many to be more intuitively compelling than the proposal that HOTs can explain nonintrospective consciousness. Conscious HOTs may seem to provide a more intuitively satisfactory explanation of introspection than HOTs, of whatever sort, can provide of nonintrospective consciousness.

Nonetheless, the idea that we should explain introspection by appeal to conscious HOTs actually provides the basis of an argument for the HOT explanation of nonintrospective consciousness. There is more to being a mental state's being introspectively conscious than to its just being nonintrospectively conscious. Similarly, there is more to being nonintrospectively conscious than to not be conscious at all. So nonintrospectively conscious states are intermediate between introspected mental states and those which are simply not conscious.

Introspected states are accompanied by conscious HOTs. And nonconscious mental states are presumably unaccompanied by HOTs, whether conscious or not. What can we say of non-introspectively conscious states that preserves their intermediate status between introspected states and nonconscious states? The natural answer is that they are accompanied by HOTs, but unlike the introspective case those HOTs are not conscious thoughts.

The only alternative is that nonintrospective consciousness consists instead in a disposition for a conscious HOT. A dispositional account would represent such consciousness as intermediate between introspected and nonconscious mental states, and might at the same time capture the equivalence of being nonintrospectively conscious and being introspectible.

This is hardly satisfactory. For one thing, there is no reason to adopt this view save the unfounded insistence that all mental states are conscious states. And as noted earlier, it is unclear how being disposed to have a HOT could make one conscious of one's mental states. In any event, because being disposed to have a conscious thought is the same as having that thought without its being conscious, the putative alternative very likely reduces to the HOT theory itself.

VII The Argument from Reporting and Expressing

Brentano held that all mental states are, in part, about themselves: All mental acts, he wrote, "apprehend [themselves,] albeit indirectly" (p. 128). His reason was that there is no difference between the mental act of, say, perceiving or thinking something, and the mental act of thinking that one so perceives or thinks. Since mental states are individuated by way of the mental act involved, HOTs are indistinguishable from the mental states they are about (pp. 127-128). The property of being conscious is thus intrinsic to mental states.

It is not easy to know what to make of Brentano's notion of a mental act. But in any case the two mental acts are not indistinguishable. The thought that it's raining does have the same mental analogue of performance conditions as the thought that I think it's raining. There is no difference between the circumstances in which it would be appropriate to think the two things. Still, the two have distinct truth conditions.⁶⁰ And, if the lower-level state is anything other than

an assertoric propositional state, the mental attitudes of the two will differ as well. Identity of performance conditions does not imply identity of intentional acts.

These considerations suggest another argument for the HOT theory. Reporting one's mental states is distinct from verbally expressing those states. Suppose I have the thought that it's raining. I can report that thought by saying that I think that it's raining, thereby explicitly telling you about my thought. Or I can instead verbally express the thought, without referring to it, by saying simply that it's raining.

As with Brentano's argument, it is crucial to distinguish performance conditions from truth conditions. The speech acts of saying that it's raining and saying that I think it's raining are indistinguishable with respect to performance conditions. But they are different in truth conditions. And though the speech act of asking whether it's raining would have the same performance conditions as the speech act of saying that I ask, the two are distinct not only in truth conditions, but in illocutionary force as well.

Reporting, moreover, has a special connection to consciousness. If we restrict attention to creatures with the relevant linguistic ability, a mental state is conscious if, and only if, one can report being in that state. I cannot tell you about my mental states if I am not conscious of them, and given the requisite linguistic capacity, I can tell you about those I am conscious of.⁶¹ Every sincere, meaningful speech act, finally, expresses a propositional state with the same content and a mental attitude analogous to the illocutionary force of the speech act. I cannot meaningfully and sincerely say it's raining unless I have the assertoric thought that it's raining.⁶²

When I report being in a mental state, therefore, my speech act expresses my HOT that I am, myself, in that very mental state. The ability to report a mental state is thus the same as the ability to express the very thought posited by the HOT theory. So a mental state is conscious if, and only

if, one can express a suitable HOT about that state.

The best explanation of our ability to express these HOTs about all our conscious states is that our conscious states are actually accompanied by such HOTs. Similarly, we can best explain our inability to report mental states when they are not conscious by assuming that we lack the relevant HOTs. The HOT theory therefore provides the best explanation of our ability to report mental states if, and only if, they are conscious states.⁶³

Any theory of conscious mental states must do justice to our pretheoretic intuitions. The Cartesian doctrines that all mental states are conscious and that being conscious is an intrinsic property of mental states make explaining such consciousness impossible. So it is a reasonable goal to reject those claims if we can still do justice to our intuitions. That the HOT theory provides our best hope of doing this, and also receives support from independent arguments, gives us good reason to adopt it as a working hypothesis.⁶⁴

Notes

1. Thus creatures are conscious only for various periods of time. Describing creatures as conscious may also mean just that unlike stones, for example, they have the capacity to be awake and sentient. I reserve 'creature consciousness' for occurrent consciousness, rather than this extended notion.
2. Perhaps Descartes's clearest remarks are that "no thought can exist in us of which we are not conscious at the very moment it exists in us" (*Fourth Replies, Oeuvres de Descartes*, ed. Charles Adam and Paul Tannery, Paris: J. Vrin, 1964–75 [henceforth *AT*] VII, 246); and "the 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, AT* VII, 160). See also *First Replies, AT* VII, 107, letter to Mersenne, December 31, 1640, *AT* III, 273, *Fourth Replies, AT* VII, 232, and *Principles I*, ix, *AT* VIII–1, 7. (Translations of Descartes are mine; *AT* volume and page numbers can be found in *The Philosophical Writings of Descartes*, translated by John Cottingham,

Robert Stoothoff, and Dugald Murdoch, 3 volumes, Cambridge: Cambridge University Press, 1984–91.) Cf. Locke's rather compressed remark that "thinking consists in being conscious that one thinks" (*Essay II*, i, 19); see also *Essay II*, i, 10–12 and IV, vii, 4 for somewhat clearer versions.

3. Even writers with explicitly anti-Cartesian views. Thus J. C. C. Smart, for example, holds that "[t]o say that a process is an ache is simply to classify it with other processes that are felt to be similar" ("Materialism," *The Journal of Philosophy* 60, 22 [October 24, 1963]: 651–62, p. 655).

On the Cartesian assumptions of contemporary mind-body materialists, see my "Mentality and Neutrality," *The Journal of Philosophy* 73, 13 (July 15, 1976):386–415.

4. With Stuart Hampshire, "Critical Notice of *The Concept of Mind*, by Gilbert Ryle," *Mind* 59, 234 (April 1950):237–225; and Richard Rorty, "Incorrigibility as the Mark of the Mental," *The Journal of Philosophy* 67, 12 (June 25, 1970):399–424, p. 406, fn. 11.

5. For a useful study, see Daniel C. Dennett, "Are Dreams Experiences?," *The Philosophical Review* 73, 2 (April 1976):151–171.

6. I am grateful to Ernest Sosa for having put this idea especially forcefully.

7. One could insist here that only a single, temporally discontinuous state of pain occurs, on the model of the temporally discontinuous bursts of sound as a single siren sound. (I owe this idea to Jaegwon Kim.) But all the argument here needs is that common sense be open to nonconscious pains, and it is doubtful that common sense countenances single, temporally discontinuous pains.

8. See Thomas Nagel's important article, "What Is It Like To Be a Bat?," *The Philosophical Review* 83, 4 (October 1974):435–450; and chapter 32 of this book. See also Nagel's "Panpsychism," in *Mortal Questions*, Cambridge: Cambridge University Press, 1979, pp. 181–195, and *The View From Nowhere*, chapters 1–4, New York: Oxford University Press, 1986.

9. I develop and defend this claim in "Armstrong's Causal Theory of Mind," in *Profiles: David Armstrong*, ed. Radu J. Bogdan, Dordrecht: D. Reidel Publishing Co., 1984, pp. 79–120, section V: pp. 100–108; in my review of *Perception: A Representative Theory*, by Frank Jackson, *The Journal of Philosophy* 82, 1 (January 1985):28–41; and in "The Colors and Shapes of

Visual Experiences," Report No. 28/1990, Center for Interdisciplinary Research (ZiF), Research Group on Mind and Brain, University of Bielefeld.

10. For the notion of fixing extensions, see Saul A. Kripke, *Naming and Necessity*, Cambridge, Mass.: Harvard University Press, 1980, pp. 54–159; "Identity and Necessity," in *Identity and Individuation*, ed. Milton K. Munitz, New York: New York University Press, 1971: pp. 135–164, pp. 156–161; and Hilary Putnam "The Meaning of 'Meaning'," in Putnam's *Philosophical Papers*, vol. 2 (Cambridge: Cambridge University Press, 1975): pp. 215–271, pp. 223–35. Both Kripke and Putnam maintain that, in general, what fixes the extension of a term need not coincide with what is essential or necessary to the things in that extension. See, however, note 26.

11. For this kind of argument, see Rorty, "Incorrigibility as the Mark of the Mental," pp. 409, 412.

12. We could also exclude them by insisting that, unlike the mental case, our intuitively immediate access to them relies on observation. But such a stipulation would in this context be ad hoc.

13. For such claims about the first-person perspective and viewpoint, see Nagel, *op. cit.*; John R. Searle ("Indeterminacy, Empiricism, and The First Person," *The Journal of Philosophy* 84, 3 [March 1987]:123–146; "Consciousness, Explanatory Inversion, and Cognitive Science," *The Behavioral and Brain Sciences* 13, 4 (1990):585–696; and *The Rediscovery of the Mind*, Cambridge, Mass.: MIT Press, 1992, esp. chapter 7; and Colin McGinn, *The Problem of Consciousness*, Oxford: Basil Blackwell, 1991.

14. Searle defends a version of this claim in "Consciousness, Explanatory Inversion and Cognitive Science" and in "Consciousness, Unconsciousness and Intentionality." On the idea that the intentionality of speech is merely derived, see Searle, *Intentionality: An Essay in the Philosophy of Mind* (Cambridge: Cambridge University Press, 1983). See also Zeno Vendler, *Res Cogitans* (Ithaca: Cornell University Press, 1972).

15. See Dennett, "Evolution, Error, and Intentionality," chapter 8 of his *The Intentional Stance*, Cambridge, Mass.: MIT Press/Bradford Books, 1987.

16. See my "Intentionality," *Midwest Studies in Philosophy* 10 (1986):151–184.

17. Things are actually a bit more complicated. Such an explanation can refer to conscious states without being circular if it does not rely on the property of their being

conscious; perhaps only other properties of conscious states are relevant to explaining what it is for states to be conscious. (I am grateful to Frances Egan for insisting on this.) But it will be difficult, if possible at all, to disarm a skeptic's charge that relying on conscious states harbors a tacit appeal to the property of being conscious.

18. See, for example, Dennett, *The Intentional Stance*, especially "Evolution, Error, and Intentionality"; and Jerry A. Fodor, *Psychosemantics*, Cambridge, Mass.: MIT Press/Bradford Books, 1987, esp. chapter 4.

19. "What Is It Like To Be a Bat?", *op. cit.* p. 435; also chapter 32, this book.

20. The difficulty here, and in what follows, affects only the nonintrospective consciousness of mental states; there is no difficulty about explaining introspection in terms of nonintrospective consciousness. See section VI.

21. Such acquiescence is not uncommon; see, for example, Nagel and McGinn. Nagel explicitly insists that any satisfactory account must be based on a full conceptual analysis ("What Is It Like To Be a Bat?," *op. cit.*, pp. 436–437; also chapter 32, this book.

Such acquiescence also occurs in those who take Cartesian ideas about consciousness to support an eliminativist position. See Georges Rey, "A Question about Consciousness," in *Perspectives on Mind*, ed. H. R. Otto and J. A. Tuedio, Dordrecht: D. Reidel Publishing Company, 1986, pp. 5–24; and to some extent Kathleen V. Wilkes, *Real People: Personal Identity Without Thought Experiments*, Oxford: Oxford University Press, 1988, chapter 6.

22. One version of this gulf has to do with how creatures with the capacity to be in conscious states could have evolved from nonmental organisms. One might speculate that this capacity relied on the intermediate capacity to be in nonconscious mental states.

23. Compare trying to explain cell mitosis by appealing directly to the laws of particle physics.

24. Franz Brentano's idea that a mental state's being conscious is due to its being in part about itself is a rare attempt to assign informative structure to state consciousness, conceived of as an intrinsic property. (*Psychology from an Empirical Standpoint*, tr. Antos C. Rancurello, D. B. Terrell, and Linda L. McAlister [London: Routledge & Kegan Paul, 1973], pp. 129–130.) But he gives no reason to sustain his insistence that this awareness of conscious mental states is intrinsic to those states. Other difficulties for Brentano's view are noted in section VII.

25. Thus, mental properties aside, we generally regard properties as truly simple only in sciences such as particle physics, whose properties we think cannot be explained, but must be taken for granted.

It is striking that even Freud saw consciousness as in this way simple. Moreover, he explicitly accepted the consequence that we could give no explanation of it. Freud did maintain that "[l]ike the physical, the psychical is not necessarily in reality what it appears to us to be." ("The Unconscious," in *The Complete Psychological Works of Sigmund Freud*, tr. and ed. James Strachey, London: The Hogarth Press, 1966–74, vol. 14: pp. 166–215, p. 171.) But while he countenanced unconscious intentional states, he denied that feelings can strictly speaking be unconscious (*The Ego and the Id, Works*, vol. 19: pp. 3–68, pp. 22–23). Perhaps for this reason, he insisted that consciousness is a "unique, indescribable" quality of mental states ("Some Elementary Lessons," *Works*, vol. 23: pp. 279–286, p. 282), and that "the fact of consciousness" "defies all explanation or description" (*An Outline of Psycho-Analysis, Works*, vol. 23: pp. 141–208, p. 157).

26. Thus Kripke urges that the way pains appear to us cannot diverge from how they really are: "For a sensation to be *felt* as pain is for it to *be* pain" ("Identity and Necessity," p. 163, n. 18; author's emphasis throughout) and, conversely, that "for [something] to exist without being *felt as pain* is for it to exist without there *being any pain*" (*Naming and Necessity*, p. 151). Thus "[i]f any phenomenon is picked out in exactly the same way that we pick out pain, then that phenomenon *is* pain" (*Naming and Necessity*, p. 153).

These claims lead Kripke to his famous insistence that what fixes the extension of the term 'pain' must coincide with what is essential to the items thereby picked out ("Identity and Necessity," pp. 157–161; *Naming and Necessity*, pp. 149–154). But this is so only if it is necessary that pains affect us in the way they do, that is, only if being conscious is intrinsic to pains. Kripke gives no independent reason to accept these claims.

27. Similarly, Michael McCloskey has elegantly and convincingly argued that commonsense predictions about bodily motions systematically err in ways that reveal the tacit assumption that those motions are due to an internal force imparted by the source of motion. ("Intuitive Physics," *Scientific American* 248, 4 [April 1983]:114–122.)

28. I use interchangeably the notions of being conscious of something and being aware of it.

We of course speak also of creature consciousness intransitively, but in that case there is no distinction to mark between transitive and intransitive.

29. This proposal is a substantive hypothesis about what intransitive consciousness is, not a recommendation about how to conceive of it. The goal is to explain our ordinary folk-psychological notion of a mental state's being a conscious state in terms of our being conscious of that state.

30. Including proprioception and enteroceptive perception as well as ordinary exteroceptive perception. Perhaps whenever our transitive consciousness relies on observation it also relies on inference; I rule out both so as to remain neutral on that question.

31. See, for example, Gilbert Harman, *Thought*, Princeton: Princeton University Press, 1973, especially chapters 1 and 2; and Irvin Rock, *The Logic of Perception*, Cambridge, Mass.: MIT Press/Bradford Books, 1983, esp. chapter 9.

32. Gilbert Ryle, *The Concept of Mind* (London: Hutchinson and Company, 1949), pp. 158, 159.

33. I am grateful to Peter Bieri and Fred Dretske for forcefully advancing this objection.

34. For more on the analogy between sense perception and consciousness, see D. M. Armstrong, "What is Consciousness?," *Proceedings of the Russellian Society* (1978), reprinted in his *The Nature of Mind*, St. Lucia, Queensland: University of Queensland Press, 1980, pp. 55-67, especially pp. 60-62; and *A Materialist Theory of the Mind*, New York: Humanities Press, 1968, pp. 94-107, 323-338. See also D. H. Mellor, "Conscious Belief," *Proceedings of the Aristotelian Society*, New Series, 88 (1977-78):87-101, pp. 98-100; and "Consciousness and Degrees of Belief," *Prospects for Pragmatism*, ed. D. H. Mellor (Cambridge: Cambridge University Press, 1980): pp. 139-173.

35. According to Armstrong, all mental properties are intentional, so on his account the perceptual analogy may not face this particular difficulty. But it is doubtful whether perceptual states can be satisfactorily explained by way of intentional properties. See my "Armstrong's Causal Theory of Mind," section V.

36. There is another way in which the perceptual model seems to suggest that being conscious is intrinsic. Since perceiving depends on a dedicated organ or mechanism, the perceptual model raises the question of what special organ or mechanism subserves being conscious of one's

mental states. The absence of any such organ or mechanism may tempt one to conclude that being conscious must be something internal to conscious states.

37. The difficulty is reminiscent of Aristotle's concern about whether or not the sense we use to see that we see is the same as the sense of sight (*de Anima* III 2, 425b13-4). Aristotle concludes that it is, since otherwise the sense of sight and the other sense would both have color as their proper object, and no two distinct senses can share the same proper object.

38. See below, especially note 51.

39. Only if one's thought is about oneself as such, and not just about someone that happens to be oneself, will the mental state be a conscious state. Otherwise it might turn out in any particular case that the state was a state of somebody else instead. For this special sort of reference to oneself, see, for example, G. E. M. Anscombe, "The First Person," in *Mind and Language*, ed. Samuel Guttenplan (Oxford: Oxford University Press, 1975), pp. 45-65; Hector-Neri Castañeda, "On the Logic of Attributions of Self-Knowledge to Others," *The Journal of Philosophy* 65, 15 (August 8, 1968), 439-456; Roderick M. Chisholm, *The First Person* (Minneapolis: University of Minnesota Press, 1981), chapters 3 and 4; David Lewis, "Attitudes *De Dicto* and *De Se*," *The Philosophical Review* 88, 4 (October 1979):513-543; and John Perry, "The Problem of the Essential Indexical," *Nous* 13, 1 (March 1979), 3-21.

40. *Pace* Donald Davidson ("Thought and Talk," in *Mind and Language*, ed. Samuel Guttenplan, Oxford: Oxford University Press, 1975, pp. 7-23). Davidson argues that believing something presupposes understanding the difference between believing truly and believing erroneously, and that only language provides a context in which that distinction can take hold. His arguments seem better adapted, however, to showing that what the distinction between truth and error is necessary for is our having the concept of belief, and it is far from obvious that having beliefs presupposes having the concept of a belief. Moreover, though language and its interpretation may well presuppose the distinction between believing truly and erroneously, the converse is far less clear.

41. Moreover, it is probable that being nonobservational will ensure that HOTs refer to oneself in the way suggested in note 20.

To illustrate such a minimal idea of oneself, I once proposed that each HOT could refer to the self as

whatever individual has that very HOT. The content of HOTs would then be that whatever individual has this very thought is in that very mental state ("Two Concepts of Consciousness," *Philosophical Studies* 49, 3 [May 1986]: 329–359, p. 344). Thomas Natsoulas assumes that a self-referential thought is automatically self-intimating, and infers that since HOTs thus construed are about themselves, they would make themselves conscious. He also supposes that I would accept this inference ("An Examination of Four Objections to Self-Intimating States of Consciousness," *The Journal of Mind and Behavior* 10, 1 [Winter 1989]:63–116, pp. 70–72).

It is worth stressing that I disavow the idea that HOTs are ever self-intimating, and also reject the inference from mere self-reference to a thought's being self-intimating. Plainly an intentional state could refer to itself and yet not be conscious. To be conscious, an intentional state one is in would, in any case, have to affirm that one is, oneself, in that very state (see section V). On my earlier proposal, the HOT does not say of an independently identified self that it has that HOT, but rather identifies the self as being whatever has the HOT. Natsoulas might argue that this difference is too esoteric to matter to whether a state is self-intimating, especially given my acceptance of a minimal concept of the self. But in any case the earlier proposal about how we can understand the content of HOTs is at best peripheral to the overall theory.

42. Such nonassertoric propositional attitudes would of course be relatively rare. Our knowledge of our own mental states is neither exhaustive nor infallible, though we typically act as though it is.

43. Terms for propositional mental states, such as 'belief', 'doubt', and 'expectation', can refer to tokens or types. And as just noted, the term 'thought' can refer generically to such states. So 'thought' can refer to a type, independently of what mental attitude is involved. In this case, the term in effect refers to some propositional content, rather than those states themselves.

All propositional states about one's mental states will involve the thought, so construed, that I am in those states. But to make that state conscious, that content must occur with an assertoric mental attitude.

44. As noted in section II, a particular property may be described in both dispositional and nondispositional terms. Relative to the categories of folk psychology, we need nondispositional states to make mental states con-

scious; but at a subpersonal level of description we might end up describing those states in dispositional terms. I am grateful to Daniel Dennett for arguing the virtues of a dispositional treatment, and for much useful conversation on these topics in general.

45. In discussing my "Two Concepts of Consciousness," Rey urges that it is arbitrary for consciousness to result from any particular level of nested intentionality, rather than from some other (section II). This misunderstands the motivation of the theory. It is because a HOT of the requisite sort constitutes one's being conscious of one's mental state that HOTs make one's mental states conscious.

Without the foregoing restrictions on the type of HOT, the connection between higher-order cognition and a state's being intransitively conscious will seem implausible. Both Rey and Robert Van Gulick, ("A Functionalist Plea for Self-Consciousness," *The Philosophical Review* 97, 2 [April 1988]:149–181) disregard those restrictions.

46. Moreover, HOTs presumably often refer to perceptual states demonstratively, aided by location in the relevant sensory field.

47. See, for example, David Woodruff Smith, "The Structure of (Self-) Consciousness," *Topoi* 5, 2 (September 1986): 149–156, p. 50.

48. As Brentano put it, with an unnecessary air of paradox, "[a]n unconscious consciousness is no more a contradiction in terms than an unseen case of seeing" (p. 102).

49. Natsoulas assumes I would appeal to some causal connection to explain how HOTs refer to the mental states they are about (p. 65). But whatever puzzles arise about how HOTs refer to mental states are due to general problems about reference, and require nothing special in this case. If something like causal ties is needed to explain reference generally, it will figure here as well.

50. Nor can we conceive the connection as one of inferring from a mental state to the relevant HOT (though some other inference may well be involved). Suppose I think it's raining; the relevant HOT is then that I have the thought that it's raining. But the inference from it's raining to my having the thought that it is would at best be irrational. A similar point is attributed to Michael Rohr by D. M. Armstrong in *A Materialist Theory of the Mind*, New York: Humanities Press, 1968, p. 200.

51. This supports our choice of HOTs over a perceptual model to capture the way we are transitively conscious

of our conscious states, since perceiving is factive and thinking is not.

On the confabulation of intentional states, see Richard E. Nisbett and Timothy DeCamp Wilson, "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review* 84, 3 (May 1977):231–259.

Our visual field seems to be replete with visual detail throughout. This is because eye movements provide foveal vision over a wide area, and we retain the visual information thus gained. Nonetheless, as noted in section V, we are at any given moment aware of little visual detail outside the center of our visual field. It is natural to speculate that our seeming to see much of this detail may in effect be due to our confabulating detailed visual sensations.

52. For a more extended discussion of these intuitions, see my "Two Concepts of Consciousness."

53. The terms are G. E. Moore's: "The Refutation of Idealism," in his *Philosophical Studies*, (London: Routledge & Kegan Paul, 1922):1–30, pp. 20, 25.

54. It is a familiar Cartesian tactic to define these phenomena away. Descartes, for example, having insisted that, "we do not have any thoughts in sleep without being conscious of them at the very moment they occur," goes on to concede that "commonly we forget them immediately" (letter to Arnauld, July 29, 1648, *ATV*, 221).

55. Another striking phenomenon that it is likely only a HOT theory can explain is that thoughts are generally conscious whenever we express them verbally, though not necessarily when we express them nonverbally. For this explanation, see my "Why Are Verbally Expressed Thoughts Conscious?", Report No. 32/1990, Center for Interdisciplinary Research (ZiF), Research Group on Mind and Brain, University of Bielefeld; and "Consciousness and Speech."

56. Van Gulick proposes detaching our concept of introspection from that of self-consciousness and, presumably, of a state's being conscious (p. 162). But he understands self-consciousness and state consciousness in terms of the subpersonal possession of "reflexive meta-psychological information" (pp. 160ff.), and it is unclear that this has much to do with the relevant folk-psychological concepts.

57. Any alternative account of introspection—say, in terms of attention or phenomenological *epoché*—would at the very least face the challenge of showing that these

concepts themselves need not be explicated in turn by reference to higher-order transitive consciousness.

58. We intuitively regard introspective consciousness as somehow active, in contrast to the nonintrospective consciousness of mental states. (I am grateful to Jay Rosenberg for insisting on this.) The HOT theory readily explains this. Introspection is having a conscious HOT, whereas in nonintrospective consciousness the HOT is not conscious; and, although the having of conscious thoughts is intuitively an active matter, the having of nonconscious thoughts plainly is not.

59. See, for example, Armstrong, *A Materialist Theory of the Mind*, pp. 94–107 and 323–338, and "What is Consciousness?", especially pp. 59–63; Dennett, "Toward a Cognitive Theory of Consciousness," in *Minnesota Studies in the Philosophy of Science*, ed. C. Wade Savage (Minneapolis: University of Minnesota Press, 1978): pp. 201–228, esp. pp. 216–222; Lewis, "An Argument for the Identity Theory," *The Journal of Philosophy* 63, 1 (January 6, 1966):17–25, p. 21, and "Psychophysical and Theoretical Identifications," *Australasian Journal of Philosophy* 50, 3 (December 1972):249–258, p. 258; and Wilfrid Sellars, "Empiricism and the Philosophy of Mind," in *Science, Perception and Reality* (London: Routledge & Kegan Paul, 1963): pp. 127–196, 188–189, and 194–195. Except for Armstrong's reliance on the perceptual model (see note 34), these views about introspective consciousness are similar in spirit to the theory of nonintrospective consciousness developed here.

Mellor, like Armstrong, relies on a perceptual metaphor; but unlike Armstrong he seeks to explain nonintrospective consciousness. And unlike both Armstrong and the present theory, Mellor insists that his account cannot explain the consciousness of mental states generally, but only beliefs.

60. And insofar as truth conditions are relevant to content, distinct content.

Brian Loar argues that the concepts in a lower-order thought will occur in a HOT, and concludes that the conceptual role of the lower-order thought is included in that of the HOT. ("Subjective Intentionality," *Philosophical Topics* 15, 1 [Spring 1987]:89–124, p. 103.) This claim about conceptual role seems to capture the spirit if not the letter of Brentano's view about HOTs (though as Loar notes [p. 89], it is of course independent of Brentano's antinaturalism).

61. This does not mean that the mental states of nonlinguistic creatures are never conscious states. The

ability to report our mental states in effect fixes the extension of the term 'conscious state'. And as with other mental categories, what fixes the extension of a term need not be a property common to everything in that extension. So, *pace* Van Gulick, the fact that many mental states of nonlinguistic creatures are conscious provides no reason to deny the connection between a state's being conscious and our being able to report about it (p. 162).

62. On the connection between thought and sincere speech, see my "Intentionality."

63. This argument is developed in detail in my "Thinking That One Thinks," in *Consciousness*, ed. Glyn W. Humphreys and Martin Davies (Oxford: Basil Blackwell, 1993), pp. 197–223; originally issued as Report No. 11/1989, Center for Interdisciplinary Research (ZiF), Research Group on Mind and Brain, University of Bielefeld.

64. This chapter was written during a research year spent at the Center for Interdisciplinary Research (ZiF), University of Bielefeld, Germany, and was originally issued there as Report 40, 1990, Research Group on Mind and Brain, Center for Interdisciplinary Research (ZiF). I am indebted to the Center for their generous support, and for the exceptionally congenial and stimulating environment provided there. I am also grateful to many friends and colleagues there and elsewhere for helpful discussion, especially Peter Bieri, Daniel Dennett, Jerry Fodor, Jaegwon Kim, Jay Rosenberg, and Robert Schwartz.

THE NATURE OF CONSCIOUSNESS

Philosophical Debates

edited by Ned Block, Owen Flanagan, and Güven Güzeldere

**A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England**

Second printing, 1998

© 1997 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

This book was set in Times Roman by Asco Trade Typesetting Ltd, Hong Kong.
Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

The nature of consciousness: philosophical debates /
edited by Ned Block, Owen Flanagan, and Güven
Güzeldere.

p. cm.

"A Bradford book."

Includes bibliographical references (p.) and index.

ISBN 0-262-02399-7 (hb : alk. paper). —

ISBN 0-262-52210-1 (pb : alk. paper)

1. Consciousness. 2. Philosophy of mind. I. Block,
Ned Joel, 1942- . II. Flanagan, Owen J.

III. Güzeldere, Güven.

B808.9N37 1997

126—dc20

96-17500

CIP