

David M. Rosenthal

Consciousness and the Mind

I. Are Mind and Consciousness the Same?

Writing just over 60 years ago, Sigmund Freud described the way people ordinarily conceive of the connection between mind and consciousness:

Everyone—or almost everyone—was agreed that what is [mental] ... *has* a common quality in which its essence is expressed: namely the quality of *being conscious*—unique, indescribable, but needing no description. All that is conscious ... is [mental], and conversely all that is [mental] is conscious; that is self-evident and to contradict it is nonsense.¹

Freud was well aware that this “equation ... of what is conscious with what is mental” is seldom made explicit, but rather is tacitly taken to be axiomatic.² Still, he held that both “the philosopher and the man in the street” in effect subscribe to this view about the connection between consciousness and mind.³

Putting “the philosophers” to one side for the moment, Freud does seem right that many people do tend to equate mind and consciousness. This is no less true today than at the time he wrote. Despite the huge influence Freud’s own views about unconscious mental functioning have had on us and despite impressive and dramatic experimental findings in psychology about nonconscious cognition, many would still insist that mental states and processes, properly so called, must all be conscious.

¹ “Some Elementary Lessons in Psycho-Analysis,” in *The Complete Psychological Works of Sigmund Freud*, tr. and ed. James Strachey (London: The Hogarth Press, 1966–74) (henceforth *Standard Edition*), 23: 283. The Strachey edition uses ‘psychical’, which I’ve replaced by ‘mental’; cf. the parallel passage quoted in the next paragraph, from “The Unconscious,” which uses ‘mental’.

² “The Unconscious,” *Standard Edition* 14: 167.

³ “Some Elementary Lessons,” p. 282.

This intuitive idea seems especially compelling in connection with certain kinds of mental phenomena. Can one be in pain without being conscious of being in pain, or suddenly remember or realize something without being conscious of remembering or realizing it? When one sees or hears, tastes or smells, is it possible not to be conscious of one's auditory, visual, gustatory, or olfactory sensations? And if one feels joyful, angry, relieved, or frightened, can these feelings possibly fail to be conscious feelings? It will seem obvious to many that the answer to all these questions is an unqualified "No."

There does, however, seem to be a kind of mental functioning that constitutes an exception to this close tie between consciousness and mentality. We all believe and desire many things even when we are in no way consciously thinking about those things. For example, we believe that $5 + 7 = 12$ and that the sun is far away—indeed, we presumably believe these things all the time. But we seldom have conscious thoughts about these things. So a person will believe that $5 + 7 = 12$ or that the sun is far away even when none of that person's conscious thinking is concerned with these matters. It is the same with desiring; we all desire to lead long and happy lives as well as many other more specific things, even when those things are not at all consciously before our minds. So we must accept that we can believe and desire things nonconsciously—that not all our beliefs and desires are conscious beliefs and desires.

When we speak of believing and desiring, we often have in mind dispositional states of some sort. A disposition is a long-term state something has that results in its behaving in certain ways under certain circumstances. For example, being soluble is a disposition in virtue of which soluble things dissolve when immersed in water or some other suitable liquid. So perhaps the beliefs and desires that aren't conscious are all of this dispositional variety.

What sort of disposition might believing and desiring be? It is sometimes urged that they are dispositions to act in certain ways given that certain circumstances obtain. Thus my believing that it's raining is perhaps a disposition to take my umbrella if I go outside. This kind of account has difficulty specifying the behavior at the right level of abstraction; I may, for example, take my raincoat, instead. More important, such an account faces the problem that the relevant connections between beliefs and behavior hold only if I also have certain desires and only if I have certain other beliefs. I

won't take my umbrella unless I want not to get wet and I believe that otherwise I will. And it's very likely impossible, in any particular case, to specify all the beliefs and desires relevant to the behavior in question. It is therefore more likely that dispositional believing and desiring aren't dispositions to overt behavior at all, but rather dispositions to be in other mental states that are themselves not dispositional. If I dispositionally believe that $5 + 7 = 12$, then, when the question comes up, I will have the occurrent, nondispositional thought that $5 + 7 = 12$. Similar remarks apply to dispositional desiring.

We must therefore reformulate the intuitive connection often claimed to hold between mind and consciousness to accommodate dispositional mental states. This is not difficult. We can say simply that dispositional mental states need not be conscious, but occurrent mental states must be. And this does seem to capture the traditional idea that Freud was describing. In what follows, I'll take this qualification for granted, and assume that we are talking about occurrent, nondispositional mental states.

I argue below that our commonsense conception of mind does not imply any such connection between mind and consciousness. Common sense both allows for and recognizes mental states that occur without being conscious. But whatever the case about common sense, the idea that mental states are invariably conscious has certainly been a cornerstone of some of the most important theoretical thinking for more than 350 years. Descartes, for example, having identified mental functioning with thinking, went on to insist that "no thought can exist in us of which we are not conscious at the very moment it exists in us." Indeed, Descartes actually defined thought in terms of consciousness: "[T]he word 'thought,'" he wrote, "applies to all that exists in us in such a way that we are immediately conscious of it."⁴

The identification of consciousness and mind was by no means novel with Descartes. Earlier thinkers frequently held that mental functioning is always conscious, typically drawing on Aristotle's remark that thinking implies thinking that one thinks and that perceiving always involves perceiving that one perceives. Indeed, even a century after Descartes, it was still Aristotle's dictum that echoed in the remark of John Locke that "thinking consists in

⁴ *Fourth Replies*, in *Oeuvres de Descartes*, ed. Charles Adam and Paul Tannery (Paris: J. Vrin, 1964–75), 7: 246; and *Geometrical Exposition of the Second Replies*, Adam and Tannery, 7: 160. Translations here are mine.

being conscious that one thinks.” Insofar as the tie between mind and consciousness is concerned, Descartes’s novelty lay not in affirming that connection, but rather in developing a theory of mind on which our very understanding of mentality must be cast in terms of consciousness.⁵

Freud’s psychoanalytic theory took issue with both the more common idea that mental functioning is always conscious and the stronger Cartesian claim that we must understand mind in terms of consciousness. “Psychoanalysis,” he wrote, “energetically den[ies] the equation between what is [mental] and what is conscious.”⁶ There are, he insisted, unconscious mental states, and we can describe them in exactly the terms we use to describe and understand conscious mental states. As he put it, “all the categories which we employ to describe conscious mental acts, such as ideas, purposes, resolutions, and so on, can be applied” equally well to unconscious mental states. Indeed, he continued, “the only way in which” many of “these latent [i.e., unconscious] states differ from conscious ones ... precisely in the lack of consciousness.”⁷ Moreover, though consciousness is a quality of mental occurrences, it’s “an inconstant quality ... —one that is far oftener absent than present.”⁸

It was these considerations which led Freud to insist that “mental processes are in themselves unconscious,” and that “the perception of them by means of consciousness ... [is like] the perception of the external world by means of the sense-organs.”⁹ It’s “unjustifiable and inexpedient,” he argued, “to make a breach in the unity of mental life for the sake of propping up a definition, since it is clear in any case that consciousness can only offer us an incomplete and broken chain of phenomena.”¹⁰

⁵ See Aristotle, *de Anima* 3.2.524b12–17; *Nicomachean Ethics* 9.9.1170a28–33; *Metaphysics* 12.9.1074b35–36; *de Sensu* 2.437a27–29; *de Memoria* 2.452b26–28; *de Somno* 2.455a15–20. See Locke, *Essay Concerning Human Understanding* 2.1.19.

⁶ “Some Elementary Lessons,” p. 283, again replacing ‘psychical’ by ‘mental’.

⁷ “The Unconscious,” p. 168. The Strachey edition has ‘some’ such states; Cecil M. Baines’s translation has ‘many’, in *Freud: General Psychological Theory*, ed. Philip Rieff (New York: Collier Books, 1962), p. 116.

⁸ “Some Elementary Lessons,” p. 283.

⁹ “The Unconscious,” p. 171.

¹⁰ “Some Elementary Lessons,” p. 286.

II. The Mind–Body Problem

Freud of course adduced extensive psychoanalytic evidence for mental functioning that is not conscious, arguing that many conscious mental states and much overt behavior, in normal life as well as in psychopathology, require explanation in terms of unconscious mental states. Thus he held, for example, that we can best explain dreams and parapraxes—certain slips of the tongue and related phenomena—by appeal to the idea that these occurrences result from unconscious mental acts.¹¹

Reasoning in this way from certain data to the best explanation of those data is standard scientific practice. But Freud also put forth a more general reason to resist the equivalence of mind and consciousness, a reason that's independent of both psychoanalytic theory and empirical findings. Seeing mind and consciousness as equivalent, he maintained, inevitably results in our conceiving of mind as independent of the rest of reality. In his words, it “divorc[es mental] processes from the general context of events in the universe and ... set[s] them in complete contrast to all others.”¹² Denying that equivalence was, he concluded, the only way to avoid the unbridgeable gulf that seems to divide reality into mental and physical, mind and matter.

The problem of describing the connection between mind and the rest of reality and of explaining how that connection can obtain is the mind–body problem. What Freud was claiming, then, was that if we see consciousness as essential to mind, the mind–body problem will be insoluble. If mind and consciousness are the same, the gulf that seems to separate mind and consciousness from the rest of reality will indeed be impossible to bridge.

This idea is hardly unique to or new with Freud. Indeed, since Descartes's *cogito* suggested that consciousness is essential to something's being an unextended substance, it has often been held that consciousness somehow lies at the root of the mind–body problem. Among contemporary writers, Thomas Nagel has done much to evoke that special sense we sometimes have of ourselves which makes it difficult to understand how, as conscious selves, we could ever find ourselves located among the physical furniture of the universe. According to Nagel, “the fact that an organism has conscious experience *at all* means, basically, that there is something it is like to be that

¹¹ “The Unconscious,” pp. 166, 168–69.

¹² “Some Elementary Lessons,” p. 283. Again, the Strachey edition uses ‘psychical’ where I have ‘mental’.

organism ...—something it is like *for* the organism.”¹³ What resists explanation in physical terms, he maintains, is precisely what it is like for one to be in a conscious mental state. Because he believes we cannot now explain that in physical terms, Nagel concludes that “[c]onsciousness is what makes the mind–body problem really intractable” (435).

Nagel thus differs from Freud in being relatively pessimistic about whether we can solve the mind–body problem. More important, however, they also differ about just what bearing consciousness has on solving that problem. Freud did not think that the very existence of consciousness causes difficulty in understanding the mind–body relation, since he recognized that many mental states are conscious states, and yet held that the mind–body problem is soluble. Consciousness causes difficulty in understanding the mind–body relation, he argued, only if mental phenomena are all essentially conscious. It is only the equivalence of mind and consciousness that makes the mind–body relation unintelligible.

Nagel sometimes writes as though he is tempted to endorse that equivalence.¹⁴ But in stark contrast to Freud, Nagel insists that consciousness makes the mind–body problem insoluble whether or not all mental phenomena are conscious. The mind–body relation is, he argues, unintelligible simply because we cannot capture in physical terms what it is like for one to be in conscious mental states.

I argue below that the equivalence of mind and consciousness does indeed cause difficulties, difficulties severe enough to justify rejecting that equivalence. But the difficulties this equivalence occasions do not, I argue, bear directly on the mind–body problem.

¹³ “What Is It Like to Be a Bat?” *Philosophical Review* 83/4 (October 1974): 436; reprinted in Nagel, *Mortal Questions* (Cambridge University Press, 1979), pp. 165–180.

Cf. Wittgenstein’s remark that a “feeling of an unbridgeable gulf between consciousness and brain-process” arises “when I, for example, turn my attention in a particular way on my own consciousness, and, astonished, say to myself: THIS is supposed to be produced by a process in the brain!—as it were clutching my forehead” (Ludwig Wittgenstein, *Philosophical Investigations*, ed. G. E. M. Anscombe and R. Rhees, tr. G. E. M. Anscombe [Oxford: Basil Blackwell, 1953], §412).

¹⁴ Though conceding that “[n]ot all mental states are conscious,” he holds that any mental state is, as he puts it, “capable of producing” conscious states (“Panpsychism,” in *Mortal Questions*, p. 188). It’s unclear what that requirement amounts to; is it that each instance of a mental state must be so capable or that each mental state must be of a type some of whose instances are?

Conscious mental states are states we are conscious of being in. Moreover, we are conscious of being in them in some way that seems intuitively to be immediate. What is it about our being thus conscious of our own mental states that could resist all possible explanation?

Nagel and others would appeal here to the subjective features of conscious experiences — what it is like for one to have those experiences. But what it is like for one to have a conscious experience consists of two different aspects, which it's crucial to distinguish. There is, in the first place, the consciousness of the experience. This is what all conscious experiences have in common. But there are also the various properties of experiences in virtue of which those experiences resemble and differ from one another. Conscious experiences of red and blue, for example, have in common that they are conscious, but they differ in that one has the property of mental redness whereas the other has the property of mental blueness.

When an experience of red or blue is conscious, we are conscious of having it. And it is because we are conscious of the property the experience has of being red or being blue that we are conscious of the experience itself; we are conscious of experiences in respect of their distinctive mental properties. What is it about our being conscious of these properties, then, which seems to occasion difficulty?

Natural science proceeds by discovering laws and mechanisms that link different phenomena. One way we explain a phenomenon is to find laws or mechanisms that connect it to other, more basic phenomena that we already understand or expect we will come to. And it may seem that no intelligible mechanisms could possibly link consciousness to anything more basic. Is consciousness problematic because we cannot thus connect it to other, more basic phenomena?

Once we discover the mechanisms that connect something to other, more basic processes, those mechanisms often come to seem inevitable. So it is then easy to see these mechanisms as antecedently rational or intelligible in a way that would have allowed us to tell in advance, simply by reasoning, whether that mechanism actually obtains. Since nothing we now know about consciousness allows us to project in this way what mechanisms might connect it to anything else, it may seem tempting to conclude that no such mechanisms are possible.¹⁵

¹⁵ "Panpsychism," in *Mortal Questions*, pp. 181–195. See also Joseph Levine, "On Leaving Our What It's Like," in *Consciousness: Psychological and Philosophical*

But we cannot, in general, read off the laws and mechanisms connecting things from what we know about those things taken on their own. These laws and mechanisms cannot be deduced; they must be discovered. Today we still know relatively little about brain function. So our current inability to project what mechanisms might link consciousness to more basic phenomena has no tendency to show that there are none, nor that we cannot discover what they are.¹⁶

Intuitively, a mental state's being conscious has something to do with one's being conscious of oneself. Thus Nagel and others sometimes write of the special perspective that such consciousness involves. One's having mental states that are conscious has to do with the way things are from one's own point of view, and this, it seems, cannot be captured from anybody else's point of view; we must describe it in the first-person, so to speak, and not the third-person. So perhaps this explains why consciousness seems to resist physicalist treatment.

But it is far from clear exactly what such a first-person perspective amounts to. Suppose I am in pain, and that I am conscious of my pain. Presumably I am then conscious of a first-person fact, about myself. But suppose now that you think to yourself that Rosenthal is in pain. Relative to you, that fact is a third-person fact, a fact about another person. But it is of course one and the same fact that you think to yourself and that I am conscious of. Facts and states of affairs are not, in and of themselves, either first-person or third-person. They do not reflect or embody a perspective.

III. Understanding Consciousness

I have argued so far that a mental state's being conscious does not, by itself, create difficulties in understanding the mind-body relation. This is independent of whether mental states are all conscious or only some are. Why did Freud think otherwise? Presumably because he took for granted the Cartesian idea that only consciousness could mark the difference

Essays, ed. Martin Davies and Glyn W. Humphreys (Oxford: Basil Blackwell, 1993), pp. 121–136; and *Purple Haze: The Puzzle of Consciousness* (New York: Oxford University Press, 2001).

¹⁶ For more on this, see my "Reductionism and Knowledge," in *How Many Questions?* ed. Leigh S. Cauman, Isaac Levi, Charles Parsons, and Robert Schwartz (Indianapolis: Hackett, 1983), pp. 276–300.

between mental and physical. Freud's importance lay in his seeing that the idea that we must understand mind in terms of consciousness is not a necessary truth but derives from a received theoretical framework, and in taking steps to undermine and replace that framework.

But if a particular mental state's being conscious does not cause difficulty in understanding the mind-body relation, what difference could it make if all mental states were? How could the equivalence of mind and consciousness make the mind-body connection problematic if the consciousness of only some mental states does not?

We may conclude, then, that the idea that mental states are all conscious by itself creates no special difficulties in understanding the mind-body relation. Nonetheless, that idea does cause serious problems of a different sort. It turns out that, if mental states are all conscious, we will very likely be unable to give any informative, nontrivial account of what such consciousness consists in. We will simply be unable to understand the very nature of consciousness itself.

Suppose mental states are all conscious. How could we then explain what it is for a mental state to be conscious? There are two ways we might proceed. One of these appeals to something mental; we explain what it is for one mental state to be conscious in terms of other mental states or processes. This will not do. If all mental states are conscious states, then the other mental phenomena to which our explanation appeals will themselves be conscious. So this kind of explanation results in a vicious regress. We cannot explain what it is for any mental state to be conscious except in terms of another mental state, whose being conscious itself requires explanation. We will end up appealing to indefinitely many mental states, each of them conscious, in order to explain what it is for a single mental state itself to be conscious.

The only alternative is to explain what it is for a mental state to be conscious without appealing to any other mental phenomena. One might think that this is the right way to proceed; after all, scientific explanations typically explain higher-level phenomena by appeal to those at a lower level.

We must bear in mind here that we are not after a scientific explanation. What we want is not the mechanisms or laws governing consciousness, but an answer to the preliminary question of what it is for a mental state to be conscious in the first place. We want to understand just what the phenomenon is that a scientific theory might then explain.

It is highly unlikely that we can understand what it is for a mental state to be conscious appealing only to things that are themselves not even mental. Consciousness is the most sophisticated mental phenomenon there is and the most difficult to understand; nothing in nonmental reality seems to be at all suited to help us grasp its nature. If we are to have any informative explanation of what it is for a mental state to be conscious, it is all but certain that it will have to make reference to mental phenomena of some sort or other.

It follows from this that, if mental states are all conscious, we can give no informative account whatever of what such consciousness consists in—of what its nature is. Nothing outside the realm of conscious mental phenomena is at all suited to explain consciousness, and a vicious regress results if our explanation itself appeals to conscious mental phenomena. If mental states are all conscious, then, we can at best simply trace the various conceptual connections that hold among such cognate notions as mind, consciousness, subjectivity, and self. But for present purposes these connections are trivial, since we cannot expect to learn from them anything substantive about the nature of consciousness.

I said earlier that we are not after a scientific explanation, and now I am rejecting giving a conceptual analysis, as well. But is anything else possible? Is there some way to explain consciousness that is not either a matter of giving a scientific theory or of understanding conceptual connections?

It is arguable that there is. Our commonsense knowledge about things far outstrips mere conceptual connections. It is no kind of conceptual truth that rocks are hard, that grass is green, that food nourishes, or that people are inherently social creatures. Nor is it a mere conceptual truth that speech acts express our thoughts, or that conscious states are those we are conscious of being in. We come to know these things not *a priori*, but by experience, and we could imagine things being otherwise.

Moreover, much of what we know in commonsense terms has to do with regularities and causal connections among commonsense objects and occurrences. We tend to systematize and explain things even while we remain within our commonsense framework. Moreover, we sometimes posit the occurrence of certain commonsense events even if we don't actually observe them, and we hypothesize that these occurrences explain various regularities. Much of our commonsense psychological knowledge is of this kind. We posit general connections between people's actions and their beliefs, desires, and emotions. These connections do not express conceptual

connections; rather, they in effect embody commonsense psychological theories, which can themselves then be refined and even explained in turn by a genuine scientific theory. It is this kind of commonsense theory that I am concerned with here. I want to develop a commonsense theory that explains what it is for mental states to be conscious states.¹⁷

In any case, we have seen that the difficulty that results from equating mind and consciousness is not, as Freud thought, that this equation makes the mind–body problem insoluble. Rather, it is that this equivalence prevents us from ever finding a nontrivial, informative account of the nature of consciousness. Moreover, given that equivalence, we must understand mind itself in terms of consciousness. So accepting the equivalence would also prevent us from ever developing an informative account of mind.

Perhaps, after all, this conclusion confirms Freud's idea that equating mind and consciousness makes it impossible to solve the mind–body problem. If we cannot even understand what it is for mental states to be conscious states, what hope can we have to understand the mind–body relation? It seems that our having an initial grasp of what consciousness may well be necessary for understanding how consciousness connects with physical reality.

It is far from clear that this conclusion is warranted. It may be that, even without understanding the nature of consciousness, we can still find connections between mind and brain that enable us understand the mind–body relation. But, even if we cannot do this unless we first understand what consciousness consists in, it is crucial to see that the principal difficulty about equating mind and consciousness has to do at bottom with understanding the nature of consciousness. If we are confused about what the basic problem is, the steps we take to solve it may very likely fail.

Ironically, we find an example of this in Freud himself. Freud held that, if all mental states are conscious, the mind–body problem is insoluble. He therefore concluded that the mind–body connection would be intelligible if we reject that doctrine.

But Freud did not see that equating mind and consciousness makes it impossible to give an informative account of what consciousness is, perhaps

¹⁷ On the idea that we develop theories and posit things within our commonsense framework, see Wilfrid Sellars, "Empiricism and the Philosophy of Mind," in his *Science, Perception and Reality* (London: Routledge & Kegan Paul, 1963), pp. 127–196.

because he held, somewhat surprisingly, that no such account is possible. Not only did he apparently accept that consciousness is a “unique, indescribable” quality of mental states;¹⁸ he actually maintained that “the fact of consciousness” “defies all explanation or description.”¹⁹ It is doubtful that he would have held this pessimistic view if he had understood that the main difficulty in giving an informative account of consciousness stems from the equivalence of mind and consciousness which he was at such pains to undermine. And recognizing this might have led him to see that consciousness is no more mysterious with one type of mental state than another, which might in turn have kept him from resisting, as he did, the idea that feelings and sensations can occur without being conscious.²⁰

In any case, we must reject the doctrine that mental states are all conscious if we want to come to understand what such consciousness is. In what follows, I take some initial steps in the direction of such understanding. In the remainder of this section I argue that, although many people still do find inviting the equation of mind and consciousness that Freud strove against, our ordinary, commonsense picture of mind plainly does make room for mental states that are not conscious. I then sketch a positive account of what such consciousness actually consists in. I conclude by comparing that account with two well-known traditional views.

Earlier, I gave examples that seem to favor the idea that mental phenomena are always conscious. How could one be in pain, or suddenly remember something, or feel joy or anger, without those states being conscious? But other mental phenomena suggest that nonconscious mental states can and do occur. We do sometimes see that somebody wants something or thinks that something is so before that person is at all aware of the desire or thought. This happens even when we have no reason that isn't question begging to think that the thought or desire is merely dispositional. Similarly with emotions; we 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. Even bodily sensations can, it seems,

¹⁸ “Some Elementary Lessons,” p. 283.

¹⁹ *An Outline of Psycho-Analysis, Standard Edition 23*: 157.

²⁰ *The Ego and the Id, Standard Edition 19*: 22–23; cf. *An Outline of Psycho-Analysis*, p. 197.

occur without being conscious. A pain or headache may last for some time even if distractions intermittently keep us from being in any way at all conscious of it.

These cases are neither unusual nor arcane. Nor must we infer their presence from special theoretical considerations or experimental situations. They occur often enough in everyday situations and fit comfortably and seamlessly within our commonsense ways of thinking about our mental lives. We must conclude that mental states can occur without being conscious.

But why, if that is so, do most of us feel such a pronounced temptation to think that all mental states are conscious? There may be systematic, theoretical reasons for holding that view, but those reasons are unlikely to be at the root of whatever pretheoretic ideas we have on this question. What we need to explain is why, if mental states are not all conscious and this is manifest in everyday life, so many people still see consciousness as essential to mind.

There are, I think, several reasons. One has to do with our ways of interacting socially. Even if you see from my behavior or demeanor that I think, feel, or desire something, if I do not acknowledge that I do, then typically it is best for you not to remark on it either. If you point out to me that I think or feel something and I was unaware of thinking or feeling that thing, my thought or feeling may thereupon become conscious. But, if it doesn't and I deny having that thought or feeling, you would normally take my word for it. We are not epistemologically authoritative about our own states of mind, but we are socially authoritative; typical social situations demand that others accord practical authority to us about what mental states we are in.

Even apart from such social norms, however, we typically have little interest in mental states that are not conscious. If I have an ache or pain or other bodily sensation of which I am not conscious, perhaps that is of some scientific or medical interest, but in everyday situations it will be of no particular concern to anybody. Similarly for my having perceptions of which I remain unconscious. One's beliefs, desires, and emotions, even when occurrent, typically affect behavior whether or not they are conscious. So those around me might care more about my nonconscious emotions, desires, and beliefs than they would about my nonconscious sensations and perceptions. Even so, people tend to ignore such states, presumably in the interest of interacting with others on the basis of their acknowledged mental states. This all suggests that the temptation to think that mental states are all conscious stems in part from our lack of interest in any that might not be.

It is useful to note that the intuitive idea that sensations and perceptions must be conscious is far stronger than the idea that thoughts, desires, and emotions must be. It may seem hard to imagine a nonconscious pain or ache or case of seeing something, whereas nonconscious thoughts, desires, and emotions seem easier to understand. This parallels the difference just noted between these two groups. We have less interest in whether somebody has a pain or other sensation that isn't conscious than in whether the person has a desire, thought, or emotion that isn't. This reinforces the hypothesis that we are tempted to think that mental states are all conscious simply because we have no interest in those which aren't.

Although consciousness is not essential to mental phenomena, it is certainly characteristic of many aspects of our mental lives which are most important to us. For one thing, we typically identify and classify mental states by reference to the conscious cases. Even more important, it is arguable that being a person means that many of one's mental states are conscious. The concept of a person is broader than that of a human being; we recognize the possibility of beings other than human beings that we would regard as persons. What is crucial to being a person is a kind of rational connectedness in the way one is aware of one's mental states. And this cannot occur unless many of one's mental states are conscious. This suggests a third reason why we find it tempting to see mental states as invariably conscious. Since the most important aspects of our mental lives involve mental states that are conscious, we find it easy to acquiesce in the idea that all mental states are.

IV. A Sketch of a Theory

I've argued that identifying mind and consciousness makes it impossible to give an informative account of what consciousness consists in. Let me turn now to a sketch of a positive account we can develop once we reject that equivalence.

Conscious mental states are mental states we are conscious of being in. And we are conscious of being in those states in a way that seems somehow to be immediate. The correlation between a state's being conscious and our being conscious of it holds in the opposite direction as well. If we are conscious of being in a mental state in a way that seems suitably immediate, that state is perforce a conscious state. So, if we can explain how we are

conscious of our conscious states and what that apparent immediacy consists in, we will have explained what it is for a mental state to be a conscious state.

There are two ways we know of in which we can be conscious of something. We are conscious of things when we perceive them and also when we have thoughts about them as being present. I am conscious of you if I see you or hear you. But I may also be conscious of you, without perceiving you, if I have a suitable thought about you. Being conscious of something always takes one form or the other; we are conscious of things either perceptually or by having suitable thoughts about them. Which model will explain how it is that we are conscious of our conscious mental states?

A perceptual model, which we saw that Freud endorses, may initially seem tempting. After all, we perceive things in a way that seems immediate, and perhaps there are other analogies that also make a perceptual model inviting. But that model will not do. Perceiving involves sensory qualities in virtue of which we perceive one thing rather than another. If we were conscious of things around us or of states of our bodies without being in some state with a distinctive sensory quality, that would not count as perceptual consciousness.

But that is just what happens when we are conscious of our conscious states. We do not literally sense those states, since our consciousness of them does not rely on our being in some other state with a characteristic sensory quality. Often this is obvious because there simply is no quality that occurs in connection with a state's being conscious; our thoughts, desires, and other intentional states are often conscious without there being any associated sensory quality. And, even when some mental imagery accompanies our thoughts, there is no stable connection between the two; there is no image that is characteristic of any particular kind of thought. So such associated mental imagery is not relevant to the way we are conscious of specific intentional states.

Some mental states do, of course, exhibit sensory qualities, but even there the perceptual model won't work. The distinctive qualities that occur in connection with conscious sensations are those of the sensations we are conscious of, not of the way we are conscious of them. The qualities attach to the sensations, whether or not they are conscious. Since sensory quality is no part of our being conscious of our sensations, our being conscious of them is not a form of perceptual consciousness.

Indeed, there is reason to hold that the sensory qualities characteristic of our conscious sensations occur even when sensations of the relevant types fail to be conscious, as in peripheral vision and subliminal perception and in laboratory contexts such as experiments involving masked priming.²¹ When sensations occur without being conscious, they often still affect our behavior and mental processes in ways that parallel the effects of conscious sensations. How could that happen unless the nonconscious sensations resemble and differ from one another in respect of their characteristic mental qualities?

The only other way we can be conscious of our conscious states is by having thoughts about them. In the special case in which the mental state we are conscious of is itself a thought, the thought in virtue of which we're conscious of it will be a thought about a thought. So it's convenient to use the term 'higher-order thought' (HOT) for the thoughts in virtue of which we are aware of our conscious states generally.

Normally, of course, we are not aware of any such HOTS, even though many of our mental states are conscious. But that is no problem, since HOTS need not, in general, themselves be conscious thoughts. On the present theory, a mental state is conscious if it is accompanied by a HOT about the mental state. So such a HOT will itself be conscious only if there exists a yet higher-order thought about that second-order thought. It is reasonable to expect that few such third-order thoughts will occur, and so few second-order thoughts will be conscious.

Most of the time, our mental states are conscious in an unreflective, relatively inattentive way. But sometimes we deliberately focus on a mental state, making it the object of introspective scrutiny. The HOT model readily explains such introspective consciousness as occurring when we have a HOT about a mental state and that HOT is itself a conscious thought. Introspection is the special case in which our HOTS are themselves conscious states.

What else do we know about what these HOTS must be like? One thing we know is what their content must be. A mental state's being conscious is one's being conscious of that state. But we can be more specific: A mental state's being conscious consists of being conscious of oneself as being in that state.

²¹ See, e.g., Anthony J. Marcel, "Conscious and Unconscious Perception: Experiments on Visual Masking and Word Recognition," *Cognitive Psychology* 15 (1983): 197–237.

A state is not a conscious state if I am just conscious of it, on its own, and not conscious of myself as being in the state. This constrains the kind of HOT that can be operative. The content of a HOT must be that I am, myself, in that mental state.

We know also that the HOT cannot be a dispositional state of the sort considered earlier. Suppose I have a dispositional belief that I am in some mental state; I am disposed to have an occurrent HOT that I am in the state, but I do not have any actual HOT. This will not do. Simply being disposed to have a thought about something does not result in my being conscious of that thing. For my state to be conscious, my HOT must actually occur. These issues will figure in the final section in sorting out some claims about consciousness put forth by Kant.²²

If the HOT hypothesis is to work, my HOTS must result in my being conscious of their targets in some way that seems immediate. This intuition fits comfortably with the HOT model, since HOTS are typically not themselves conscious thoughts. So nothing will seem to us to mediate between the HOTS and their targets. If I am not aware of having a HOT to begin with, there is no question of something's seeming to mediate between it and its target.

But we can do better. We want to avoid the appearance that, when we have a HOT about a mental state, something mediates between our being in that state and our having the HOT. What sort of mediation could occur? The

²² Peter Carruthers has argued that a satisfactory theory should posit only dispositions to have HOTS, and not the actual HOTS themselves. His reason is that there aren't enough computational and cortical resources to accommodate all the actual HOTS we would need. See his *Language, Thought, and Consciousness: An Essay in Philosophical Psychology* (Cambridge: Cambridge University Press, 1996), and *Phenomenal Consciousness: A Naturalistic Theory* (Cambridge: Cambridge University Press, 2000).

But there is every reason to suppose we have ample cortical resources. The idea that we cannot accommodate actual HOTS very likely stems rather from our introspective sense that the conscious mind cannot accommodate more than a very few HOTS at a time. But introspection cannot tell us that we have mental states unless they are conscious, and we have seen that most HOTS are, by hypothesis, not conscious.

Daniel C. Dennett's view that "[c]onsciousness is cerebral celebrity" also reflects a dispositional view about what it is for a mental state to be a conscious state ("The Message is: There is no Medium," *Philosophy and Phenomenological Research* 53/4 [December 1993]: 929). See also Dennett, *Consciousness Explained* (Boston: Little, Brown and Company, 1991), ch. 10 and esp. p. 315.

HOT might be mediated by an inference, as when somebody tells us that we seem happy or sad and we infer from that that we are. Or it might be mediated by some observation. Just as I can sometimes simply see that you are happy or sad, perhaps also I can sometimes notice from my own behavior that I myself am happy or sad. No other type of mediation seems possible. So we can guarantee the appearance of immediacy by hypothesizing that a state is conscious only if the accompanying HOT relies neither on conscious inference nor conscious observation.²³

This theory of what it is for a mental state to be a conscious state is surprisingly successful at saving the phenomena—at explaining, that is, our commonsense ideas about conscious mental states. An example of this has to do with the sense we have that our mental states' being conscious somehow involves the self. Because a HOT is a thought that one is, oneself, in the mental state in question, a mental state's being conscious presupposes an idea of the self. Of course, being in a conscious mental state does not mean consciously thinking about oneself. The theory captures this, too, since the HOT is typically not conscious.

If we focus only on creatures like ourselves, who have language and also the vocabulary to describe their mental states, a mental state's being conscious coincides with one's being able to tell somebody about that state. Experimental psychologists routinely rely on this in determining the conscious states of human subjects, and the absence of language creates methodological challenges for studying the conscious states of nonhuman animals.

²³ Neither inference nor observation would impede our sense of immediacy if the inference or observation were not itself conscious. So we must rule out only HOTS based on inference or observation of which we are conscious. And, since any HOT mediated by conscious observation would presumably also be mediated by conscious inference, ruling out conscious inference is enough. For more on the HOT model of what for a mental state to be a conscious state, see my "Two Concepts of Consciousness," *Philosophical Studies* 49/3 (May 1986): 329–359; "A Theory of Consciousness," in *The Nature of Consciousness: Philosophical Debates*, ed. Ned Block, Owen Flanagan, and Güven Güzeldere (Cambridge, Mass.: MIT Press, 1997), pp. 729–753; "Explaining Consciousness," in *Philosophy of Mind: Contemporary and Classical Readings*, ed. David J. Chalmers (New York: Oxford University Press, forthcoming 2002); and "Introspection and Self-Interpretation," *Philosophical Topics* 28/2 (Fall 2000). "Two Concepts of Consciousness" and "Introspection and Self-Interpretation" will be reprinted in my *Consciousness and Mind* (Oxford: Clarendon Press, in preparation).

The present theory explains this connection between language and consciousness. A mental state is conscious if it is accompanied by a HOT that one is in that state. And thoughts can typically be expressed in speech. But to express the HOT is simply to say that one is in that mental state. So when a mental state is conscious, one is typically able to tell other people about it.²⁴

If a mental state's being conscious consists in its being accompanied by a suitable HOT, what connection will obtain between the HOT and the mental state that it's about? Doubtless that mental state will very often be causally implicated in producing the accompanying HOT, but such a causal connection is not required by the theory. A mental state's being conscious consists in one's being conscious of being in that state, and the HOT will achieve that effect regardless of whether there is any causal connection. And, given that mental states often occur without any accompanying HOT, it cannot be that the states are, by themselves, causally sufficient to produce HOTs; other mental occurrences must enter into the aetiology. So it is natural to expect that a HOT might result causally from those other mental factors alone.

The HOT model is especially useful in explaining an important feature of consciousness, namely, that a particular mental state need not be conscious in respect of all of its mental properties. When one consciously sees something red, one has a conscious sensation of red. The sensation, moreover, will be of a particular shade of red, depending on what shade of red one sees. But, unless one focuses on that shade, one typically isn't conscious of the red in respect of its specific shade; the sensation is conscious only as a red of some indeterminate shade.

The HOT model readily explains this. In the ordinary case, the content of one's HOT is simply that one has a sensation of red. When one focuses on the shade, the content of that HOT incorporates more specific information about shade. One thereby comes to be conscious of oneself as having a sensation of that particular shade of red. What it's like for one actually

²⁴ For an expanded version of this argument, see my "Thinking that One Thinks," in *Consciousness: Psychological and Philosophical Essays*, ed. Martin Davies and Glyn W. Humphreys (Oxford: Basil Blackwell, 1993), pp. 197–223, to be reprinted in *Consciousness and Mind*.

Thoughts need not be expressed in speech. So the connection between language and consciousness does not imply that nonlinguistic animals lack conscious mental states. Such creatures doubtless often have thoughts, and some may well have HOTs.

changes from having a sensation of an indiscriminate shade of red to having a sensation of a specific shade.

If HOTS need not capture all the mental properties of the states they represent us as being in, is it possible that HOTS can actually misrepresent us as being in mental states that we are not in? Indeed, if a HOT's mental target need not be implicated in causing the HOT, might it happen that HOTS sometimes occur in the absence of any suitable target at all?

It is likely that both sorts of thing do actually happen. We are sometimes conscious of our sensations as being of a type that we know they cannot be. Dental patients, for example, report being in pain as a result of drilling on a tooth that is fully anaesthetized or whose nerve is destroyed; the likely explanation is that the patient is conscious of fear and vibration from the drill as pain. When this explanation is given to patients and drilling resumes, patients are no longer conscious of themselves as feeling pain. Still, they continue to recall the earlier experience as one of pain, suggesting that the way they were conscious of the earlier occurrence determines what it's like for them. The best explanation is that the patients are conscious of these experiences in respect of the mental properties their HOTS represent the experiences as having.

It is also well known that we sometimes confabulate beliefs and desires that there is compelling evidence we do not actually have, often to conform to expectations we or others have.²⁵ In such cases, however, people are as firmly convinced as ever that they actually have those beliefs and desires; what it's like for them is they are in those states. Again, the HOT model provides the best explanation. When one has a HOT to the effect that one has a particular belief or desire, what it's like for one is actually having that state. One is conscious of oneself as being in the state, even though the state does not actually occur.

V. Consciousness, Mental Attitude, and Content

Having traced certain parallels and differences between the present theory and Freud's thinking about consciousness, it may be useful in closing to

²⁵ The classic study is 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.

discuss the views of two other traditional thinkers, Kant and Franz Brentano.

In an important passage in the *Critique of Pure Reason*, Kant writes: “The *I think* must be able to accompany all my representations.” Later in the passage, Kant refers to this *I think* as itself a representation (*Vorstellung*), and he claims that it results from a kind of self-consciousness (*Selbstbewußtsein*). Its being produced by such self-consciousness explains, he held, what it is for the representations it accompanies to be mine.²⁶

Because Kant describes the *I think* as a representation, it’s tempting to see it as an intentional state whose content is that I think the content of whatever representation accompanies it. The content of that accompanying representation wouldn’t need to be part of the content *I think*; rather, the *I think* could refer to the accompanying representation demonstratively.²⁷ So construed, the *I think* is in effect a thought about another intentional state, that is, a HOT.

But Kant claims not that the *I think* actually accompanies every other representation, but only that it must be able to do so. So perhaps Kant holds some version of the dispositional HOT view, on which a representation’s being conscious consists in a disposition to have a suitable HOT. That view, we saw, would face serious difficulties, since its merely being possible for the *I think* to accompany another representation would do nothing to make that other representation conscious.

But there is good reason to think that this is not Kant’s view. If a representation’s being conscious consisted just in its being possible for the *I think* to accompany it, every representation would be conscious, since the *I think* must be able to accompany every other representation. But Kant seems to have countenanced representations that are not conscious. Thus he explicitly allows for the possibility of representations that are mine even

²⁶ Immanuel Kant, *Critique of Pure Reason*, trans. and ed. Paul Guyer and Allen W. Wood (Cambridge: Cambridge University Press, 1998), p. 246, B132; cf. B131–38, B157–59, A122–23, B406.

²⁷ On demonstrative reference to an accompanying vehicle of content, see Donald Davidson, “On Saying That,” in *Words and Objections*, ed. Donald Davidson and Jaakko Hintikka (Dordrecht: D. Reidel, 1969), pp. 158–174; reprinted in Davidson, *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984), pp. 93–108. Davidson construes the ‘that’ of indirect discourse as a demonstrative pronoun that refers to an accompanying sentence, construed as exhibiting its use.

though “I am not conscious of them as such” (B132). And presumably no such representation would be conscious; how, if a representation of mine is conscious, could I fail to be conscious of it as mine?

There are other reasons as well to see Kant as holding that representations are conscious only when the *I think* actually accompanies them, and not also when such accompaniment is merely possible. For one thing, Kant’s main concern in this passage is not actual consciousness, but its transcendental conditions. Thus the unity of the *I think*, he writes, is “the *transcendental* unity of self-consciousness” (B132). So we can understand the possibility of the *I think*’s accompanying a particular representation as the transcendental condition for that representation’s being conscious, rather than what results in its actually being conscious. Seeing a representation as being conscious only when the *I think* actually accompanies it also fits with the idea that the *I think* is produced by self-consciousness, since a representation so generated would in turn confer consciousness on the representations it accompanies.

The consciousness of our various mental states is by itself not enough, according to Kant, to explain our sense of ourselves as unified mental subjects. As he puts it, “the empirical consciousness that accompanies different representations is by itself dispersed and without relation to the identity of the subject” (B133). Our sense that all our mental states occur in a single subject derives, rather, from our “*adding* one representation to the other and being conscious of their synthesis” (B133). The HOT model helps capture what this idea amounts to. Our sense that many conscious states are unified in a single subject results from our HOTS’ making us conscious of them in large groups whose members are related in respect of their content or other mental properties.²⁸

According to Kant, any representation that the *I think* could not accompany would be a representation that could not be thought at all (B132). This suggests an additional role Kant may have had in mind for the *I think*.

²⁸ Although this is doubtless part of what our sense of subjective unity amounts to, it is arguably not all. See my “Persons, Minds, and Consciousness,” in *The Philosophy of Marjorie Grene*, in *The Library of Living Philosophers*, ed. Lewis E. Hahn (La Salle, Ill.: Open Court, forthcoming 2002), §§4–5; “Introspection and Self-Interpretation,” *Philosophical Topics* 29/2 (Winter 2001), special issue on introspection, forthcoming, §7; and “Unity of Consciousness and the Self,” MS, delivered at the 5th Annual Meeting of the Association for the Scientific Study of Consciousness, Duke University, May 2001.

Every intentional state has two aspects, its intentional content and a mental attitude towards that content. An intentional state with the content that it's raining might be a case of believing that it's raining, doubting that it's raining, or desiring that it rain, depending on the mental attitude held toward that intentional content.

Intentional content is a matter of how things are represented. So mental representations are often taken as constituting the mental content of intentional states, in abstraction from any mental attitude. But according to Kant, a representation that the *I think* could not accompany would be a representation that could not be thought at all. So perhaps the role of the *I think* is at bottom that of contributing some mental attitude to the representation it accompanies. The insistence that the *I think* must be able to accompany every other representation simply allows for the possibility that representations can occur on their own, independently of any mental attitude.

Kant does describe the *I think* as a representation, but perhaps he thought that mental attitude itself derives from a suitable representation. And he does describe the *I think* as special in that no further representation can accompany it (B132), which would be so if its role were to contribute a mental attitude. Whatever the case about that, assigning this additional role to the *I think* results in difficulties. Kant plainly held that the *I think* makes one conscious of the representation it accompanies. So, if it also contributes mental attitude to that representation, the same factor that, by providing a mental attitude, converts a mere content into an actual intentional state would then also make one conscious of the resulting state. Every intentional state will accordingly be conscious. Kant would in effect be following Descartes's claim that "no thought can exist in us of which we are not conscious at the very moment it exists in us."

Kant on this reading follows Descartes in various other ways, as well. Descartes encourages seeing mental attitude as resulting from a suitable representation, since he holds that the idea of a mental act is not distinct from the act itself.²⁹ More important, he forges an intimate link between mental attitude and consciousness. The mind is transparent to itself, he holds, only in respect of mental attitude, not content.³⁰ Indeed, it is because I cannot doubt

²⁹ To Mersenne, January 28, 1641, Adam and Tannery 3: 295.

³⁰ Thus, he tells Arnauld, it is the operations of the mind (*operationibus intellexi*) he means when he says "that nothing can be in our mind of which we are not at that time conscious." *Fourth Replies*, Adam and Tannery 7: 232; cf. 7: 246–247. That

or deny that I am holding a mental attitude that Descartes's *cogito* seems to him to show that we automatically are conscious of all our intentional states.³¹

But intentional states do often occur without being conscious; so it cannot be that a single factor contributes both mental attitude and consciousness. A mental state's being conscious is a matter of one's being conscious of oneself as being in that state, and mental attitude alone cannot have that result. A HOT makes one conscious of oneself as being in a particular state because the HOT's content is about that state. On this Cartesian reading, Kant's failure to distinguish mental act from content leads to the mistaken assimilation of mental attitude and consciousness.

It is worth noting a somewhat different family resemblance that holds between the present theory and that of Brentano, whose lectures Freud attended. Brentano maintained that every mental state is about two things; it is about its ostensible object—what it appears to be about—and it is also about itself. And, by being about itself, every mental state makes one aware of it; every mental state is accordingly a conscious state. In effect, Brentano posited an actual HOT as an intrinsic part of every mental state.³²

Like Freud, Brentano unequivocally insisted that there is no contradiction in the notion of a mental state that is not conscious. It is no part of the idea of a state's being mental that it be conscious.³³ In denying any conceptual connection between a state's being mental and its being conscious, Brentano

such transparency does not apply to intentional content is clear from the possibility of ideas that are unclear and indistinct.

³¹ Adam and Tannery 7: 25. Only *seems* to show, since the question of whether one is holding a mental attitude need not arise, and if it doesn't my mere inability to doubt or deny that I am cannot make me conscious that I am, any more than does the mere possibility of an accompanying *I think* or HOT.

For more about Descartes on content and mental attitude, see my "Will and the Theory of Judgment," in *Essays on Descartes' Meditations*, ed. Amélie O. Rorty (Berkeley: University of California Press, 1986), pp. 405–434, repr. in *Descartes's Meditations*, ed. Vere Chappell (Lanham, Maryland: Rowman and Littlefield, 1997), pp. 129–158; and "Judgment, Mind, and Will in Descartes," Report 29/1990, Center for Interdisciplinary Research (ZiF), University of Bielefeld.

³² Franz Brentano, *Psychology from an Empirical Standpoint*, ed. Oskar Kraus, English edition, ed. Linda L. McAlister, tr. Antos C. Rancurello, D. B. Terrell, and Linda L. McAlister (London: Routledge & Kegan Paul, 1973), pp. 129–130.

³³ "An unconscious consciousness is no more a contradiction in terms than an unseen case of seeing" (p. 102).

in effect returned to the Aristotelian tradition, which defined psychological states by their distinctive function, rather than, as Descartes urged, in terms of consciousness.

Brentano thereby acknowledges the pivotal distinction between a mental state's being conscious and its being a state in which a person is conscious *of* something. Many mental states are states in virtue of which one is conscious of something, but not all such states are on that account conscious states. It is tempting to see here an important intellectual debt that Freud may have owed Brentano. It may even be that Brentano's distinction between the direct and indirect objects of intentional states inspired Freud's own distinction between the manifest and latent content of thoughts and desires.

Still, this liberal attitude about consciousness and mind did not carry over into Brentano's actual theory. The absence of a conceptual link between a state's being mental and its being conscious does not show that any mental states actually fail to be conscious. And unlike Freud, Brentano held that none ever do. It is doubtless that conviction which encouraged Brentano to insist that every mental state has awareness of itself built in.

But that view cannot be sustained. One's awareness of some thing must have an assertoric mental attitude; wondering whether one is in a mental state or desiring to be would not make one conscious of oneself as being in that state. And, as Brentano himself recognized (p. 127), an intentional state can have only one mental attitude. So a conscious case of wondering, doubting, or desiring could not contain an assertoric awareness of itself. As with the Cartesian reading of Kant, Brentano's view runs afoul of considerations of mental attitude. We can avoid this difficulty only by letting our awareness of our conscious states be the result of distinct HOTS, with their own independent content and mental attitude.³⁴

*Washington University in St. Louis,
Philosophy, Neuroscience, and Psychology, and
Graduate Center, The City University of New York*

³⁴ This paper was presented at the City University of New York Graduate Center on November 3, 2002. An earlier version was written during a year as a research fellow at the Center for Interdisciplinary Research (ZiF), University of Bielefeld, Germany, and was delivered in June 1990 as part of the first ZiF Colloquium Lecture Series. I am indebted to the Center for their generous support, and for the exceptionally congenial and stimulating environment provided there.