I. INTRODUCTION

Introspection is a process by which people have focused access to their own mental states. We have access, of course, to all our conscious states, since a mental state's being conscious in the first place means in part that one is conscious of that state. That we are conscious of all our conscious states is evident from the fact that, if one is altogether unaware of a state, that state does not count as a conscious state. But our introspective access to a mental state is something more than that state's simply being conscious. Access to a state is distinctively introspective only when it is deliberate, attentive, and reflective.

It is sometimes held that all mental states are introspectible, that is, subject to introspective access. But there are two distinct ways to understand this claim, which are not always distinguished. When we introspect a state, we attend to it in respect of its mental properties, its intentional content and mental attitude in the case of intentional states and its sensory modality and qualitative properties in the case of sensory states. So mental states are the right kind of states to be subjects of introspective awareness, and we can readily imagine introspecting any state that has these mental properties.
But a state's being introspectible can also mean that one is actually able to introspect the state. And it may well be that individual mental states occur that we cannot, for whatever reason, access introspectively. Many mental states occur without being conscious at all, and perhaps there are mechanisms or other factors that actually prevent some of these states from becoming conscious. And, if a state cannot come to be conscious at all, it cannot become introspectively conscious. It might even be that in some cases states that are conscious but not introspectively so cannot, because of some mechanism or other factor, come to be introspectively conscious. Mental states are all the right kinds of things to be subjects of introspective consciousness, but it may be that specific factors prevent some of them from ever becoming introspectively conscious.

Introspective access to our mental states is subjectively unmediated, in that nothing seems, from a subjective point of view, to mediate between one's state and one's introspective consciousness of it. Such access also seems, subjectively, to be within our control, in that nothing more is needed for one to become introspectively aware of a state than for one to decide to be. The introspective access one has to one's own mental states seems to afford the only access anybody has to what mental states one is in other than inferences from one's behavior. Such inferences are, of course, notoriously fallible. So it is tempting to conclude that introspection provides an epistemically privileged way to determine what mental states a person is in. Nothing, it may seem, could overrule a person's introspective pronouncements. Indeed, an even stronger conclusion seems to some to be warranted. It is arguable that we need some independent way to tell whether inferences from behavior to mental states are correct. So, since introspection is evidently the only alternative way we have to tell what mental states a person is in, it may seem inviting to conclude that introspection is not only privileged, but decisive.

But this traditional line of argument is unconvincing. We rely both on introspective access and behavioral inference to determine what mental states a person is in, but neither route to this information need be infallible. It could well be instead that we simply establish what mental state somebody is in by taking into account all available information, both introspective and inferential. That would not guarantee a decisive conclusion about what mental state the person is in, but that is the way generally with empirical information. Introspection is no less fallible and subject to correction than are third-person considerations. Just as there is no way to determine the existence and behavior of subatomic particles independent of balancing observation and theoretical inference, so we cannot determine what mental states a person is in independently of balancing introspection with third-person evidence.
For these and other reasons, I will not discuss the alleged special epistemic status that some theorists have held introspective access has, nor the distinctive methodological role introspection is sometimes held to play. Instead, I shall focus on trying to characterize the nature of introspection and, in particular, its relation to other mental phenomena. I begin in section II by urging that the widespread theoretical model of introspection on which it is a kind of inner perceiving fails, and that introspecting must therefore be a kind of conscious, attentive thinking about our own mental states. In section III I turn to the contrast mentioned at the outset between introspective consciousness of our conscious states and our ordinary, unreflective access to those states. Section IV, then, discusses whether we ever are, properly speaking, conscious in a first-person way of our own mental states, and section V takes up a challenge about what the content of such introspective consciousness of our mental states could be. In section VI, I discuss confabulatory introspective reports and the interpretive character of introspection. I conclude in section VII with some brief remarks about introspection in connection with personhood and the unity of consciousness.

II. THE PERCEPTUAL MODEL

Introspection gives us conscious access to our mental states. But what sort of process or mechanism is responsible for that conscious access? There is a popular idea that introspection is a kind of "inner sense" or inner perception of our mental states. This perceptual model, which reflects the very etymology of 'introspect' (from the Latin spicere [look] and intra [within]), derives considerable plausibility from the idea that introspecting is a kind of internal monitoring of one's mental states. Just as exteroceptive perception monitors the external environment and proprioception monitors the positions and movements of our limbs, so introspection performs a parallel monitoring function for the contents of our minds.

We know of no organ that subserves introspection in the way the eyes subserve the transduction of visual information and the ears auditory information. But this is hardly decisive. For one thing, there might well be some relatively modular portion of the brain that subserves introspective monitoring by being sensitive to the occurrence of conscious mental states. But even if that is not the case, not all perceiving proceeds by way of such organs; proprioception, for example, relies on no dedicated organ.

Indeed, exteroceptive, proprioceptive, and enteroceptive perception seem intuitively to have so little in common that one might well wonder what it is in virtue of which we count all of them as kinds of perceiving. Two things seem necessary. One is that all three involve the monitoring of
some process or condition. The other is that they all seem to proceed by way
of some kind of qualitative property that reflects the state or condition being
monitored. There is in each case a distinctive state whose very qualitative
character serves to carry the relevant information.

One thing that encourages a perceptual model of introspection is the
temptation to regard introspecting as a clear case of monitoring. I shall
argue in section VI that thinking of introspection as a kind of monitoring
misleads us in serious ways about the nature of introspection. But indepen-
dent of that question, it is plain that introspecting does not involve any dis-
tinctive feel or other qualitative property. This is especially obvious when
we introspect intentional states, such as thoughts and desires. Such purely
intentional states have no qualitative aspect, and nothing qualitative occurs
when we are introspectively conscious of them. When we introspect quali-
tative states, we are aware of those states in respect of the qualities they
seem to exhibit. Still, the only qualities that figure in our introspecting such
states are those of the states themselves; there are no additional qualities by
means of which introspection represents the qualities of the states it makes
us aware of. Introspection adds no additional sensory qualities of its own.

Sensory quality is not an incidental aspect of perceiving. Perceiving
represents bodily or environmental states of affairs, but one need not repre-
sent these states of affairs by perceiving them. One can instead represent the
state of affairs by thought, by simply thinking that it obtains. Perceiving a
red object or one’s arm’s being extended differs from simply thinking that a
red object is in front of one or that one’s arm is extended because in per-
ceiving we represent those states of affairs in part by the qualitative charac-
ter of the perception. Thinking, by contrast, involves no such qualitative
aspect. If we could subtract from perceiving its qualitative properties, we
would be left just with our thinking about the state of affairs in question.

William Lycan has argued that the lack of qualitative properties is not
decisive here, and that the monitoring function of introspection by itself suf-
fices for introspection to count as a kind of perceiving, or at least as rele-
vantly like perception. And he urges that, just as perceiving monitors the
environment, so introspecting monitors one’s mental states. But being a kind
of monitoring does not establish any useful resemblance to perception.
Bodily processes often track states of affairs, by occurring whenever the
state of affairs obtains and not otherwise, and indeed by occurring with an
intensity that reflects that of the condition being monitored; states of the
liver, for example, track blood glucose levels. Any bodily state that carries
the requisite information monitors the condition in question, but few such
states do so by perceiving that condition; states of the liver do not perceive
glucose levels, except metaphorically. More than monitoring is necessary
for a state to count as perceiving.
An advocate of the perceptual model might urge that, even if not all monitoring is perceiving, perhaps all mental monitoring is. But even when monitoring does take place by way of mental processes, the states that carry the relevant information could be thoughts, rather than perceptions. Suppose we were so constituted that our thoughts about whether a particular state of affairs obtains tracked that very state of affairs; our thoughts about whether the state of affairs obtains would be reliable indicators of whether it does. Those thoughts, then, would monitor that state of affairs. There would be no temptation to regard the mental states in virtue of which this monitoring takes place as perceptions. We have an independent mark of whether a mental state is a thought or a perception; perceptions, unlike thoughts, involve sensory qualities. Standard cases of mental monitoring are typically perceptual, but that does not mean that all mental monitoring is.

Lycan recognizes that introspection involves no qualitative properties, but denies that this speaks against a perceptual model since, on his view, qualitative properties are not necessary to perceiving. Lycan’s view here relies on his view about what the qualitative properties of perceiving consist in. “[S]ensory properties presented in first-order states are,” he maintains, “the represented features of physical objects; e.g., the color presented in a (first-order) visual perception is [just] the represented color of the physical object.” But if introspection monitors anything, it monitors mental states, not physical objects. So, if Lycan is right that first-order perceptual states have no qualitative properties, but only represent those of perceived objects, there is no reason to expect the perceiving of first-order mental states to have qualitative properties either. That introspecting involves no qualitative character of its own does not, he concludes, undermine the perceptual model of introspecting.

I discuss such representationalist views of mental qualities in section V. But whatever the merits of representationalism about mental qualities, anybody who adopts that view must still draw the ordinary distinctions between, for example, a visual perception of a red object and a nonperceptual thought that there is a red object in front of one. Both mental states represent the color of the object, and both represent the object as being in front of one. We cannot distinguish the two by saying that perceiving a red object, unlike merely thinking that such an object is in front of one, represents one as responding visually to the object. Seeing represents only the object’s visible properties, not also the modality of one’s access to it. However one draws the distinction between perceiving something and thinking of it as present, it seems clear that introspecting will fall not on the side of perceiving, but of thinking about something as present.

Introspection occurs when we focus our attention on some particular mental occurrence. The attended mental occurrence is already conscious;
introspection does not transform a state that was not conscious at all into one we introspect. Introspecting a state is, rather, our focusing on that state from among the range of those which are already within our stream of consciousness. This suggests one more reason why many theorists see introspection in perceptual terms; the most convenient model for attentively focusing on something is selective attention in vision, hearing, and the other perceptual modalities.

But directing one’s attention to something need not be a matter of perceiving at all. Even if I am looking straight at something, I may not be attending to it; I attend to it by consciously thinking about it in a concentrated, detailed way. Attention is often a matter of where one’s conscious thoughts are focused. And this is likely the case when we introspect. When we introspect a state, we typically become conscious of it in respect of its detailed mental properties; we mentally describe it in terms of those salient distinguishing characteristics. And mentally describing something is a matter of having thoughts about it. One introspects being in a mental state by having conscious thoughts to the effect that one is in a state with the relevant mental properties.

In any case, the absence of introspective qualities is by itself enough to show that introspection is not a kind of perceptual monitoring. Still, introspection does give us access to our mental states by making us conscious of those states. The only way it could do this other than by responding perceptually to the states is by way of thoughts to the effect that we are in those states. Introspection is not the perceiving of our mental states, but the having of thoughts about them.

III. INTROSPECTIVE AND NONINTROSPECTIVE CONSCIOUSNESS

A mental state’s being introspectively conscious differs from the way in which mental states ordinarily are conscious in everyday life. A state is introspectively conscious only when one is conscious of it in an attentive, deliberate, focused way, whereas states are nonintrospectively conscious when our awareness of them is relatively casual, fleeting, diffuse, and inattentive.

Consider, for example, the conscious states that make up one’s visual field at any particular moment. Though all conscious, none of these states is typically subject to the focused, deliberate, attention characteristic of introspection. This is true even for states that occur in the center of our visual field, where we have foveal vision; these states have far more fine-grained detail and greater resolution than others, but still typically occur without benefit of introspective scrutiny. Similarly for our sensory input from the other sensory modalities, both exteroceptive and bodily.
Conscious intentional states are also seldom introspected. One might, for example, be considering what to do on a particular occasion, and thus having a series of thoughts and desires, many of them conscious, but it is rare in so doing that one will pause to introspect any of these conscious thoughts or desires. Introspecting is not just being conscious of a mental state; it is being conscious of it in a deliberate, focused, reflective way.

The distinction between conscious states we introspect and those we do not is often overlooked, doubtless in part because in both kinds of case we are conscious of the states in question. Indeed, the term ‘introspection’ is sometimes applied to both kinds of case. But, however the term is used, there is plainly a distinction we must mark between the conscious states we introspect and those we do not, between the states we are attentively and deliberately conscious of and those we are conscious of in only a casual and inattentive way.

Introspecting a state is not perceptual. So it must instead involve one’s having a thought to the effect that one is in that state. How, then, does introspective consciousness differ from ordinary, nonintrospective consciousness? A simple, straightforward answer would be possible if part of what it is for a state to be a mental state at all were that the state be conscious. We could then say that nonintrospective consciousness is just what happens when that state occurs, without anything else, and introspection is what happens when, in addition to the state, one has an accompanying thought that one is in that state.

This simple picture may seem inviting. When we introspect, we are conscious of consciously attending the introspected state in respect of certain mental properties, and hence conscious of having a thought about that state. When a state is nonintrospectively conscious, by contrast, we are never conscious of any such accompanying thought. And, if being a mental state is, in part, being conscious, our not being conscious of a thought is enough to show that no such thought occurs.

But this picture cannot be sustained. Mental states plainly do occur that are not in any way conscious. This is widely recognized for intentional states, such as beliefs, desires, expectations, and the like. But there is also compelling evidence that perceptual sensations also occur without being conscious. Masked priming experiments provide situations in which detailed qualitative information occurs of which subjects are wholly unconscious, results that fit with everyday cases of peripheral vision and subliminal perceiving. And striking work with blindsight patients suggests the same conclusion. Even bodily sensations, such as pains and aches, arguably occur without being conscious; if one is concentrating on something, one may be wholly unaware of a pain evident to others by one’s behavior.

So, we need to explain not only how introspectively conscious mental states differ from those which are nonintrospectively conscious, but also
how both of those differ from states that are not conscious at all. In addition, since not all thoughts are conscious, one’s not being conscious of any thought about a mental state cannot show that no such thought occurs; at most it can show that, if any such thought does occur, it is not a conscious thought. A more complicated picture is needed to explain this three-fold distinction between states that are introspectively conscious, those which are nonintrospectively conscious, and those which are not conscious at all.

When a state is conscious, whether introspectively or not, we are conscious of that state. As noted at the outset, no state of which one was in no way at all conscious would count as being conscious, introspectively, nonintrospectively, or in any other way. So we have a way to approach our three-fold distinction. When a state is not conscious, we are not conscious of it at all. And a state’s being nonintrospectively conscious must differ from our introspecting a state in virtue of the different ways we are conscious of those states.

It is sometimes held that explaining a state’s being conscious, in whatever way, in terms of one’s being conscious of it is unavoidably circular, since that is simply explaining consciousness by appeal to consciousness. But there are two distinct notions of consciousness that figure in this explanation. We are conscious of many things other than mental states; whenever we perceive something or think about it as present, we are conscious of that thing. Moreover, we understand what it is to have thoughts about things and to perceive them independently of understanding what it is for mental states to be conscious. So we understand what it is to be conscious of a state independently of a state’s being a conscious state. There is no circularity in the explanation.

A state will be introspectively or nonintrospectively conscious depending on how one is conscious of that state. What difference in how we are conscious of our mental states is relevant here? Introspecting a state involves one’s having a thought about that state as having certain mental properties. Since the thought is about the state as one’s own current state, it is about that state as present. This explains how it is that introspecting a state makes one conscious of it.

When we introspect a state, we are conscious of it in a way that seems attentive, focused, deliberate, and reflective. When a state is conscious but not introspectively conscious, by contrast, we are conscious of it in a way that is relatively fleeting, diffuse, casual, and inattentive. Introspective and nonintrospective consciousness do not seem to differ in any other ways. There is no other phenomenological or subjective difference, and no theoretical reason to posit any additional difference.

The natural conclusion is that both nonintrospective and introspective consciousness involve an accompanying thought about the target state, but
that the cases differ because of differences in the accompanying thought. When we introspect a state, the accompanying thought is attentive, conscious, and deliberate, whereas the accompanying thought when a state is nonintrospectively conscious is fleeting, casual, and inattentive. In addition, since we are unaware of any accompanying thoughts when our mental states are only nonintrospectively conscious, the accompanying thought in that case is not a conscious thought.

This account of nonintrospective consciousness is just the higher-order-thought (HOT) hypothesis about such consciousness that I have developed elsewhere. But it is worth stressing that the present argument does not rely on that hypothesis to explain introspective awareness. Rather, the argument here goes in the opposite direction. An independent account of introspective awareness, together with the need to explain the difference between introspective and nonintrospective consciousness, by themselves led us to the HOT hypothesis.

Indeed, the HOT hypothesis about nonintrospective consciousness is very likely the only credible way to explain how a state’s being nonintrospectively conscious is weaker than introspection but stronger than a state’s not being conscious at all. On that hypothesis, a state is nonintrospectively conscious just in case it is accompanied by a HOT that is not itself conscious. By contrast, a state is introspectively conscious if it is accompanied by a HOT that is conscious. And it fails to be conscious at all if no HOT, conscious or not, accompanies it.

It might be thought that there is another way to explain how ordinary, nonintrospective consciousness falls in between introspection and a state’s not being conscious at all. Perhaps a state is nonintrospectively conscious not if it is actually accompanied by a HOT, but rather if there is a disposition for such a HOT to occur. A state would be introspectively conscious, then, if accompanied by an occurrent HOT. This view about nonintrospective consciousness has been advanced by Peter Carruthers, who argues that only dispositions for actual HOTs to occur are necessary for mental states to be conscious, not the actual HOTs themselves.

This dispositional version of the HOT model may seem inviting. When a mental state is conscious, it seems phenomenologically that there is no HOT present but that one could readily occur. But that phenomenological appearance is irrelevant. HOTs are posited as the best explanation of what it is for a mental state to be a conscious state, not because we find them in our stream of consciousness. Indeed, since the HOTs posited to explain nonintrospective consciousness are themselves not conscious thoughts, we would expect to be phenomenologically unaware of them. And, in any case, a dispositional variant of the HOT hypothesis cannot work. Since a disposition to have a thought about something cannot make one conscious of that thing,
dispositions to have HOTs cannot explain how we come to be conscious of our conscious mental states.\footnote{11}

I mentioned at the outset that introspection makes us aware of our mental states in a way that seems, from a first-person point of view, to be unmediated. We can readily explain this subjective sense that introspection is direct and unmediated by stipulating that the conscious HOTs in virtue of which we introspect mental states do not rely on any inference of which we are aware. Similarly, as I have argued elsewhere, for the nonconscious HOTs that explain nonintrospective awareness. We need not suppose in either case that nothing actually does mediate between the HOT and its target, only that nothing seems subjectively to mediate.

The same holds for our subjective sense that our introspective awareness of mental states is attentive, focused, and deliberate. It seems to us when we introspect a state that we are conscious of the state in a distinctively attentive way and that this is the result of a deliberate decision to focus on the state in question. We are conscious of ourselves as deliberately attending to our mental states. But being conscious of oneself as deliberately attending to something does not establish that one actually is.\footnote{12} It is an independent question whether the way one is conscious of an introspected state engages any mechanisms of attention. Indeed, the argument of section VI raises a doubt about whether introspective consciousness is, properly speaking, attentive.

There is even a question about how reliable our subjective sense is that introspecting is deliberate. In the case of overt actions, being deliberate is a matter of the action's resulting from a process of deliberation, at least part of which is normally conscious. Such conscious deliberation seldom if ever precedes our introspectively focusing on a mental state. Perhaps being deliberate amounts to something else in this case, but it is not clear just what that might be.

Sometimes attention is drawn to a mental occurrence by some external stimulus; a bright flash of light may, for example, cause one to focus on the almost painful brightness of the resulting visual sensation. By contrast, no external event seems to figure in our focusing on a mental state when we introspect. This suggests that our subjective sense that introspective focusing is deliberate may be due just to its resulting from wholly inner factors, factors of which we typically remain unaware.\footnote{13}

What is clear, however, is that when we introspect a state, we are not only conscious of the state, but also conscious that we are thus conscious. We are aware, when we introspect, that we are focusing on the state in question. In ordinary, nonintrospective consciousness, by contrast, we are conscious of the conscious state, since otherwise it would not be a conscious state at all, but we are not also aware of being so conscious. Our being conscious of our mental states passes, in these cases, without any apparent notice.
Ned Block has sometimes characterized the HOT hypothesis as suited to explain a special type of consciousness, which he calls monitoring consciousness. This characterization may be due in part to Lycan’s and Armstrong’s appeal to higher-order perceiving to explaining what they see as the monitoring aspect of introspective consciousness. Block characterizes such monitoring consciousness as metacognitive, and at one point identifies it with attention (279). These remarks strongly suggest that Block’s notion of monitoring consciousness is just introspective, as against ordinary, non-introspective consciousness.

But the notion of consciousness that the HOT hypothesis seeks to explain is in the first instance not that of introspective consciousness, but that of ordinary, inattentive, fleeting, nonintrospective consciousness. Nor does the HOT hypothesis offer any explanation whatever about attention. A mental state is nonintrospectively conscious, on that hypothesis, if it is accompanied by a HOT that is not, itself, a conscious thought. Introspection is the special case in which that HOT is conscious, which happens when a yet higher-order thought occurs—a third-order thought about the second-order thought. One would see the HOT hypothesis as a dedicated explanation of introspective consciousness only if one tacitly assumed that any HOTs one has would have to be conscious thoughts.

IV. ARE WE CONSCIOUS OF OUR MENTAL STATES?

When a state is conscious, whether introspectively or not, one is conscious of being in that state. Such states differ from those which are not conscious, since in those cases one is not conscious of the state, at least not in the seemingly immediate way characteristic of conscious states. The difference between introspective and nonintrospective consciousness, then, is due to a difference in how we are conscious of states in the two kinds of case.

It is sometimes argued, however, that a mental state’s being conscious does not consist in our being conscious of that state and, indeed, that conscious states actually exist of which we are not conscious. An ingenious argument for this conclusion has been put forth by Fred Dretske. Adapting his case in inessential ways, consider two scenes, one consisting of ten trees and another just like it but with one tree missing. Suppose that you consciously see first one scene and then the other, and that in each case you consciously see all the trees. But suppose that, despite all this, you notice no difference between the two scenes. This kind of thing happens all the time; we often consciously see everything in a scene and then everything in a slightly later version of that scene, altered in some small, unnoticed way.

Dretske assumes that, since you consciously see all the trees in each
scene, you have conscious experiences not only of both scenes but of all the
trees in each. Still, there is some part of the conscious experience of ten
trees that is not part of the conscious experience of nine trees, and that part
is, itself, a conscious experience: a conscious experience of a tree. Since you
notice no difference between the scenes, you are not conscious of the dif-
ference between them. But the conscious experience of the extra tree is the
only difference between the two overall conscious experiences. Accordingly,
Dretske concludes, you are not conscious of that experience of the extra
tree. And, because the experience of the extra tree is a conscious experience,
there is a conscious experience of which you are not conscious.16

But this conclusion is unwarranted. It is notorious that one can be con-
scious of something in one respect but not in another. And this happens
when we are conscious of our mental states just as with anything else. One
might, for example, be conscious of a particular visual experience as an
experience of a blurry patch, but not as an experience of a particular kind of
object, say, a table. Yet it might be that that very experience, though not con-
scious as an experience of a table, is nonetheless an experience of a table. It
might leave memory traces of a table, detectable by its priming effects.
Similarly, one might be conscious of the experience of the extra tree as an
experience of a tree, or even just as part of one's overall experience, but not
conscious of it as the thing that makes the difference between the experi-
ences of the two scenes.17 Indeed, this is very likely the best explanation of
what happens in the case Dretske constructs.18

Since Dretske holds that a mental state's being conscious does not con-
sist in one's being conscious of that state, he must have some other account
of what it is for a state to be conscious. He suggests that a state's being con-
scious consists, instead, in its being a state in virtue of which one is con-
scious of something or conscious that something is the case ("Conscious
Experience," 280). This proposal has the disadvantage that no mental state
could then fail to be conscious, since every mental state is such that being
in it makes one either conscious of something or conscious that something
is the case.

More pressing for present purposes, if a state's being conscious is not a
matter of one's being conscious of it, Dretske must also offer an account of
introspection which does not imply that we are conscious of the states we
introspect. On Dretske's proposal, introspection resembles what he calls dis-
placed perception. Just as we come to know how full a car's gas tank is by
looking at the gauge, so one comes to know what visual experience one has
by noticing what physical object one is seeing. One comes thereby to be con-
scious that one is in a particular mental state, but not conscious of that state.19

Coming to be conscious that one is in a particular state presumably
means having a thought to that effect; so on Dretske's account introspecting

212
a mental state means having a thought that one is that state. In introspecting a visual state, for example, one extrapolates from the physical object one sees to a thought that one has the resulting visual experience. Still, when one introspects, the inference by which this extrapolation takes places is presumably never a conscious inference; it is never an inference of which one is conscious. Moreover, one’s thought that one has the visual experience will, on Dretske’s view, be a conscious thought, since he holds that any state in virtue of which one is conscious that something is the case is a conscious state.

So, on Dretske’s view, introspecting a state consists in having a conscious, noninferential thought to the effect that one is in that state. That conscious HOT is inferred, albeit not consciously, from some thought about the thing that the introspected experience is about; one infers from a thought about the object one sees to a thought that one sees that object. Perhaps that is often the origin of the thoughts by means of which we introspect our conscious states, though I will argue in section VI that our introspective awareness often arises in other ways. That question aside, Dretske’s view differs from that defended above only in its commitment to the idea that mental states are all conscious, which cannot be sustained.

According to Dretske, when one introspects a state one is not conscious of that state, but conscious only that one is in that state. But it is tempting to think that whenever one is conscious that a state of affairs obtains, one is thereby conscious of whatever things are involved in its obtaining. Dretske’s denial that we are conscious of the states we introspect is accordingly very likely due to his desire to reject a perceptual model of what happens when we introspect our mental states. Displaced perception is not perception, properly speaking, but something we infer from a perception.

Like Dretske, Searle also holds that we are never conscious of our mental states, again in large measure because he adopts a perceptual model of what it is to be conscious of something. Being conscious of something, he assumes, is a matter of observing it. But “where conscious subjectivity is concerned,” he argues, “there is no distinction between the observation and the thing observed, between the perception and the object perceived.” He concludes that no distinction is tenable between the introspecting of a state and the state introspected; any “introspection I have of my own conscious state is itself that conscious state.”

Searle argues that the distinction between introspecting and the introspected state collapses because we can describe consciousness only in terms of what it is consciousness of (96). This has a certain plausibility. Since consciousness is a matter of the way things appear to us, we must describe it in terms of those things. We describe consciousness in terms of its content.

But we cannot describe consciousness by only describing the things that
appear to us; we must also describe the way they appear to us and, indeed, the fact that they do appear to us. We must also say how we are conscious of the things we are conscious of. Suppose that I consciously see a red object; I may describe my consciousness simply as being of a red object. But there are different ways one can be visually conscious of red objects. My being conscious of it might be focused and careful or casual and offhand; I might be conscious of it in respect of its specific hue, brightness, and environmental contrasts, but my consciousness might not register those features. There are differences in the ways we are conscious of things that do not reduce to the contents of our consciousness. And we are not only conscious of things in these various ways, but also conscious that we are conscious of things in one way rather than in another. There is more to say about our being conscious of things than just what things we are conscious of.

On Searle's construal of being conscious of things as a matter of observing them, it is plausible that we are never conscious of our conscious states, since we never observe those states, properly speaking. But this undermines only the perceptual model of introspective and nonintrospective consciousness, which, however inviting metaphorically, we have already seen to be indefensible.

Indeed, Searle concedes as much. "[I]f by 'introspection' we mean simply thinking about our own mental states, then there is no objection to introspection. It happens all the time, and is crucial for any form of self-knowledge." Searle would very likely also concede that our thoughts about our own mental states can sometimes be subjectively unmediated. His complaint is with the perceptual model of introspection and with the idea that it has some special epistemic standing, with introspection construed as "a special capacity, just like vision only less colorful" (144). Like Dretske, Searle adopts a view on introspection that differs from that put forth above only in denying that mental states ever occur without being conscious.

Dretske and Searle both construe being conscious of things on a perceptual model, in effect denying that having thoughts about our mental states constitutes a way of being conscious of those states. But having a thought about something does, under the right circumstances, make one conscious of that thing. Perhaps having a thought about Julius Caesar does not make one conscious of him, properly speaking. But that is because Caesar is not present and we do not think of him as being present. Similarly, we do not describe ourselves as conscious of the abstract objects we think about. But when we think about something as being present, we are thereby conscious of it. Suppose that you and I are in the same room, but you neither see me nor hear me or sense me in any other way. Still, if you realize I am there and have a thought about my being there, you are thereby conscious of me. And HOTs represent their targets concretely and as present, since they represent one as currently being in particular token mental states.
V. THE CONTENT OF INTROSPECTION

When a mental state is conscious, whether introspectively or not, one is conscious of being in that state. But we are never conscious of our conscious states perceptually. Dretske and Searle recognize that, but, since they conceive of being conscious of things perceptually, they deny that we are ever conscious of our conscious states.

That denial leads Searle to insist that any "introspection I have of my own conscious state is itself that conscious state"; if being conscious of something is perceiving it, the only things we are conscious of are the things we perceive by way of our conscious perceptual states. Introspective awareness of conscious states is still possible, according to Searle and Dretske, if it is not construed perceptually but in terms of having thoughts about our conscious states.

But a more thoroughgoing challenge is available to the idea that introspection makes us conscious of our mental states, a challenge about what it is we are conscious of when we seem to introspect. That question arises especially vividly in connection with perceptual experiences. As G. E. Moore noted, consciousness seems to be "transparent" or "diaphanous," in that when we try to focus on our conscious perceptual states, it may seem that we simply look through those states to the things we perceive in virtue of those states. As Searle notes, it seems that we can describe our perceptions only in terms of what they are perceptions of. When we try to focus on the visual sensation we have in seeing a red tomato, it may seem that we end up focusing only on the tomato itself and its redness.

These considerations lead Searle to conclude that there is no distinction between introspecting a state and the state introspected. But one could cast one's conclusion in slightly different terms. Instead of maintaining that introspecting is nothing over and above the introspected state, one could simply insist that, when we do seem to introspect an experience, the only properties we are aware of are the properties of the things that experience represents. As Gilbert Harman puts it, when you have a conscious experience of seeing a red tomato, "[y]ou have no conscious access to the qualities of your experience by which it represents the redness of the tomato. You are aware [only] of the redness of the tomato." However, perhaps Harman's remark pertains only to perceptual awareness; the only property one sees or perceives when one has a conscious visual experience of a red tomato is the redness of the tomato. But there are other ways to be aware of properties when they are instantiated, and when we conceive of awareness more broadly, the redness of the tomato is not the only property one can be aware of. One may also sometimes be aware of properties of the visual experience itself.
If one has an experience of a red tomato, for example, one could come to be aware that one has that experience. And then one will be aware of the property the experience has of representing the redness of the tomato. And, if it is a visual experience of a red tomato, the experience will represent that redness in a way unlike the way a nonperceptual thought might represent that redness. Nonperceptual thoughts represent redness simply by being about that color, whereas visual experiences represent it in a distinctively qualitative way. So, when one is aware of having an experience of a red tomato, one is aware of the experience’s representing the redness in that distinctively qualitative way.

As Harman notes, a visual experience of the redness of a tomato is not red in the way the tomato is. But that does not preclude its having some mental quality, distinct from the red of the tomato but characteristic of the visual sensations we have when we see red objects in standard conditions of illumination. Mental qualities are unlike the perceptible qualities of physical objects in several ways. For one thing, they are properties of states, rather than objects. Equally important, they are not perceptible; we do not come to be aware of them by perceiving them. The question of whether we might be perceptually aware of our perceptual experiences and their qualitative properties seems, indeed, to arise only because of this misconception about mental qualitative properties. Only if mental qualities resembled the perceptible properties of physical objects in being literally perceptible could one suppose that we might be perceptually aware of them.

Although mental qualities are distinct types of property from the perceptible properties of physical objects, we fix the reference to mental qualities by appeal to their perceptible physical counterparts. The distinguishing mental quality of red sensations, for example, resembles and differs from other mental color properties in ways that parallel the ways in which the perceptible red of physical objects resembles and differs from other physical color properties. Thus the mental quality of red resembles mental orange more than mental green just as physical red is closer to physical orange than to physical green; similarly with the qualitative properties of other perceptual modalities. The similarities and differences that matter here are those which figure in our commonsense taxonomies, not the similarities and differences that hold among reflectance properties or wavelengths of visible light, described in terms of physical theory.

It is these mental qualities by which perceptual experiences, unlike nonperceptual thoughts, represent things in a distinctively qualitative way. Harman insists, however, we “have no conscious access to the qualities of your experience by which it represents the redness of the tomato.” He concludes that we have no idea whether experiences represent things by means of such mental qualities, since we are never aware of any.
But, if we do not model these qualities on the perceptible properties of physical objects, thereby restricting ourselves to perceptual awareness, there is no reason to expect we would never be aware of these qualities. We sometimes have thoughts about our experiences, thoughts that sometimes characterize the experiences as the sort that visually represent red physical objects. And to have a thought about an experience as visually representing a red object is to have a thought about the experience as representing that object qualitatively, that is, by way of its having some mental quality.

When one has a thought that one's own experience visually represents a red physical object, that thought need not be in any way consciously inferential or based on theory; it might well be independent of any inference of which one is conscious. From a first-person point of view, any such thought would seem unmediated and spontaneous. And it is the having of just such thoughts that makes one introspectively conscious of one's experiences. Such a thought, moreover, by representing the experience as itself visually representing a red physical object, makes one conscious of the experience as being of the type that qualitatively represents red objects. And being an experience of that type simply is having the relevant mental quality. So, being conscious of oneself as having a sensation of that type is automatically being conscious of oneself as having a sensation with the quality of mental red, and thus of the mental quality itself. One can be noninferentially, and therefore directly, conscious of the qualitative character of experiences themselves.

Introspection is the awareness of one's own mental states in a way that is deliberate, attentive, and reflective. When one consciously sees a red tomato, one is conscious of the tomato. And, since the experience involved in perceiving the tomato is a conscious experience, one is conscious also of that experience, though in a casual, fleeting, and inattentive way that normally escapes one's notice. It is ordinarily the tomato, not the experience, that could come to attract one's attention.

But one's attention can shift away from the tomato and onto the experience itself. Suppose, for whatever reason, that one concentrates not on the tomato but on one's experiencing of it and, in particular, on the sensory aspect of one's experiencing. Then one will be attentively, deliberately, and reflectively conscious of the qualitative character of one's experience in virtue of which one is experiencing the tomato. One thereby introspects one's experience of a tomato.

Shifts of attention are often a matter of what one looks at. That example of attention might suggest that, if one shifts one's attention from the tomato to introspecting one's experience of it, that shift must be a matter of casting one's gaze inside, as though looking at the sensation rather than the tomato. But shifts of attention are not always strictly perceptual; they are
often due to higher cortical processes, including shifts of what one's thoughts concentrate on, rather than what one perceptually focuses on. This is what happens when one shifts one's attention from the tomato to introspecting one's experience of it. One comes to introspect the experience by having a thought about it that occupies center stage in one's attention. One comes to be conscious of oneself as consciously focusing one's attention on the experience itself, and not the tomato.

This focusing on a mental state in virtue of which one perceives something, as against the thing one perceives, is the relatively unusual occurrence Husserl described as the bracketing of conscious states from the objects they represent. Such bracketing of the mental state from what it represents consists simply in one's focusing on it as such, thereby diverting attention from the object represented. One is still conscious in introspecting of both the mental state and the represented object, just as one is conscious of both when one consciously sees a tomato in the ordinary, unreflective way. Consciously seeing means having a conscious experience, and that means being conscious of the experience. But in ordinary, unreflective cases the experience one is conscious of attracts no attention, and one does not notice at all that one is conscious of it.

The difference between one's consciously seeing a tomato in that ordinary, unreflective way and one's introspecting one's sensation of the tomato is a matter of what one concentrates one's attention on. And that, in turn, is a matter of what thoughts dominate one's attention. No introspection occurs if one's dominant conscious thoughts are about the tomato. But if, instead, they are about one's having a particular type of visual experience, one is introspectively conscious of that experience.

When one shifts one's attention from the tomato to one's visual experience of it, it does not seem, subjectively, that some new quality arises in one's stream of consciousness. This doubtless seems to underwrite Harman's insistence that the only quality one is aware of in either case is that of the tomato. But that is too quick. As noted earlier, we can be conscious of a particular thing in various ways. When one sees a red tomato consciously but unreflectively, one conceptualizes the quality one is aware of as a property of the tomato. So that is how one is conscious of that quality. One thinks of the quality differently when one's attention has shifted from the tomato to one's experience of it. One then reconceptualizes the quality one is aware of as a property of the experience; one then becomes conscious of that quality as the qualitative aspect of an experience, in virtue of which that experience represents a red tomato. Whether one is conscious of the quality one is aware of as a physical property or as a mental property of an experience depends on how one's dominant conscious thoughts represent that quality.
When we consciously but nonintrospectively see a tomato, we conceptualize the quality we are aware of as a property of the tomato itself. Does that mean that we project onto the tomato a property of the qualitative state, a property which the tomato does not actually have? The projectivist view about color, recently defended by Paul A. Boghossian and J. David Velleman, among others, claims exactly that. According to them, “the intentional content of visual experience represents external objects as possessing colour qualities that belong, in fact, only to regions of the visual field.” Such projectivism holds that the content of our visual experiences is systematically in error with respect to the properties it attributes to perceived physical objects.

But conceptualizing the qualities we are aware of in visual experience as belonging to perceived objects involves no such projection. Nor is there any systematic error in thinking of those qualities as properties of perceived objects, though there is, as already noted, a systematic ambiguity we must watch for in the way we use color words. We use our color vocabulary to attribute physical properties to the objects we perceive, properties which those objects have in themselves independently of whether anybody perceives them. But we also use those very same color words to describe the visual experiences we have of those objects. Since color words ascribe two distinct types of property, one to perceived objects and the other to our visual experiences of them, there is no occasion to project the mental properties of visual experiences onto perceived physical objects.

We attribute the qualities we are aware of in nonintrospectively conscious experience to the physical objects we experience. So whether we project mental qualities onto physical objects will depend on just what qualities we are aware of when we see something consciously, but without introspecting the experience. It may be tempting to hold that the qualities we are aware of in such cases are the properties of the experiences themselves, since the qualities, as we experience them, occur only in perceiving. Indeed, we are perceptually aware of the independently occurring color properties of perceived objects only by way of the mental qualities of our experiences, in virtue of which the colors of perceived objects appear to us. The properties we perceive physical objects to have present themselves to experience by way of the mental qualities of those experiences.

But in nonintrospectively conscious experience, our conscious thoughts are not about those mental qualities, but about the properties we take the perceived objects to have, independently of whether anybody experiences those properties and, if so, in what way. And our conscious thoughts are about those independently occurring physical properties not as they appear to us, but as they occur independently of any perceptual process. So the properties we consciously attribute to perceived objects in these cases are
not the mental appearances of physical properties, but the physical properties themselves, as they occur independently of being perceived.

Only when we introspect our experiences do we have conscious thoughts about the mental qualities that are the appearances of the independently occurring physical properties. But in those cases we attribute the mental qualities we have conscious thoughts about to our qualitatively experiences, not to the objects we perceive. The idea that we project mental qualities onto perceived physical objects derives from conflating introspectively conscious experience, in which we attribute mental qualities, with nonintrospectively conscious cases, which are about perceived physical objects.

So, when we consciously but nonintrospectively see a red tomato and conceptualize the red quality we are aware of as a property of the tomato, we attribute to the tomato the quality as it is independent of perception, not as we experience it. Even if we are unclear about just what redness consists in independent of its being perceived, our visual experience attributes to the tomato only the independently occurring property, not the property as it appears in experience. There is no projection onto the tomato of the way that property appears to us. And, since we attribute to those objects only the properties themselves, not the face they present to perceptual experience, there is no systematic error in our interpreting the qualities we are aware of in experience as belonging to the objects we perceive.

When an experience is introspectively conscious, we have conscious thoughts about the experience itself. So we then conceptualize the quality we are aware of as a property of the experience itself. Even so, we still describe that quality in terms of what the experience represents. When we describe an experience as being red, we are describing it as being an experience of the sort that represents red physical objects. But as noted earlier, experiences represent things in virtue of the mental qualities they have; only thus could they represent things in the distinctively qualitative way that differs from just having a nonperceptual thought. So being aware of an experience as being the type that represents red objects is being aware of it as having the relevant mental quality. So, even when we are aware of these properties in representational terms, we are aware of them as mental qualities.

One might suppose that the property we are directly aware of is the red of the physical object, whereas we are only indirectly aware of the corresponding mental quality, since we conceive of that mental quality by reference to the perceptible physical property. Every mental quality is of a particular mental type in virtue of its resembling and differing from other qualities of that sensory modality in ways that parallel the ways the corresponding perceptible property resembles and differs from others in its perceptible family. So being aware of a mental quality as being of some particular
qualitative type is being aware of it in respect of those similarities and differences. But the directness that matters for being introspectively aware of mental states and their properties is subjective; it is just that nothing consciously mediates between our awareness and the quality we are aware of. And one need not be aware of any particular instance of a perceptible physical property to be aware of a mental quality in respect of such similarities and differences.

Whether one simply sees a tomato consciously or introspects one’s experience of the tomato is a function of how one’s dominant conscious thoughts represent the situation. Similar remarks apply to the introspecting of our intentional states. Suppose I am consciously thinking about a particular situation or a decision I made or the solution to a problem. If my conscious thoughts are just about the situation, decision, or solution, no introspecting occurs. I take no conscious note of the thoughts I have, only what those thoughts are about.

But if, for whatever reason, I shift my attention and come consciously to think about the thoughts I have about the situation, decision, or solution, I become introspectively aware of those thoughts. I may think about those thoughts for any number of reasons, perhaps because something about the thoughts surprises me, because they seem unclear, or because I come to wonder why I have those particular thoughts. I do not stop having the thoughts about the situation, decision, or solution; it is just that those thoughts no longer dominate my conscious thinking, but have become the subject matter of my dominant conscious thoughts. By consciously and deliberately focusing on those thoughts, rather than on what they are about, I come to be introspectively conscious of them. Introspection is not a special faculty, whose access to our mental states resembles our perceptual access to the objects around us. Rather, it is our thinking about our own mental states in a reflective, deliberate, and attentive manner.

VI. INTROSPECTION AS SELF-INTERPRETATION

Whenever one is conscious of something, one is conscious of that thing in some respect. This is so whether one is conscious of something perceptually or by having a thought about it as being present. One never perceives or thinks about anything in respect of every aspect of the thing.

This applies to being introspectively conscious of one’s mental states. A mental state’s being conscious, whether introspectively or not, consists in one’s being conscious of that state. So the respects in which one is conscious of it will determine how the state presents itself to consciousness. The state will be conscious in these respects and not in others.
Suppose I consciously see a red physical object. I may be conscious of the exact shade of red. But I may instead be conscious of the object simply as being of some indiscriminate shade of red. The object itself has, of course, some exact shade of red, but the way I am conscious of the object may or may not reflect that exact shade. I may be conscious of the object as having that exact shade or just as an object with some indiscriminate shade of red.

Similarly with the way we are conscious of our conscious states. I may be conscious of an experience of red in respect of its exact shade, but I may be conscious of it only as being of some indiscriminate shade. Whichever way I am conscious of the experience, that is what it’s like for me to have that experience. How I am conscious of a sensory experience determines what it’s like for me to be in it. And, even when I am conscious of the experience only as being of some indiscriminate shade, the experience itself, moreover, independently of my being conscious of it, has a mental quality with some exact shade. In most cases and perhaps in all, my attentively focusing on the experience would suffice to make me conscious of that exact shade, just as with the colors of physical objects.27

The way we are conscious of mental states and their properties in respect of different properties figured in the argument of section V. When one consciously sees a red tomato, one’s experience has a red qualitative aspect. But how one is conscious of that qualitative aspect depends on whether one is introspectively conscious of the experience. In the ordinary, unreflective case, one is aware of the experience only as representing the tomato; one’s dominant conscious thoughts represent the quality as a property of the tomato. But when, instead, one is introspectively conscious of the experience, one is conscious of the situation in respect of the properties of that experience. So one’s dominant conscious thoughts represent the very same qualitative aspect as a property of the experience itself.

All this accords with the HOT model of both introspective and nonintrospective consciousness. The difference between consciously seeing a red tomato in the ordinary, unreflective way and being introspectively aware of the experience is a matter of what HOTs occur in the two cases. When one consciously but unreflectively sees the tomato, one has a HOT about one’s experience, but that HOT is not attentive or deliberate. Indeed, that HOT is not even conscious; one’s having it wholly escapes one’s notice making one’s awareness of the experience inattentive, casual, and fleeting. When one introspects the experience, on the other hand, the HOT one has about that experience is both conscious and attentive. HOTs allow a straightforward and economical explanation of the difference between the two situations.

I argued in section II that even if we construe introspection as a kind of monitoring of our conscious states, we need not adopt a perceptual model of
how such monitoring occurs. It might instead be that introspection is the having of conscious HOTs. If these conscious HOTs were caused by target states by way of some suitable mental mechanism, they would be reliable indicators of the presence of those states. The resulting introspective awareness would be a kind of monitoring of those targets.

Some reliable causal mechanism is required for introspective awareness to constitute a kind of monitoring, whether that awareness is a matter of some process that resembles perceiving or a matter of having HOTs. It could be, as Lycan urges, that the “awareness is a product of attention mechanisms,” though the mechanisms could also be of some other sort. But whatever the case about that, introspective awareness is a kind of monitoring only if some mechanism normally leads from introspected states to introspective awareness.

Doubtless introspected states are sometimes causally implicated in our introspective awareness of them. But it is likely that introspective awareness sometimes arises uncaused by any target state and, more important, that no causal mechanism normally plays a role in leading from target state to one’s introspective awareness of it. This is best seen by noting a way in which introspective awareness is often in error. In well-known work on confabulatory introspective awareness, subjects report being in intentional states that there is convincing evidence do not even occur. Typically they report beliefs and desires that would make ex post facto sense of their behavior, often in ways that enable them to appear, both to themselves and to others, in some favorable light. Subjects literally confabulate stories not only about the causes of their being in particular intentional states, but actually about what intentional states they are in.

The subjects in these cases seem to invent states to be conscious of themselves as being in, states that fit with a particular picture they have either of their motivations and character or of some take on their social environment. It is not surprising that people sometimes invent stories about themselves in this way. What is striking is that subjects in these cases take themselves to be reporting on beliefs and desires to which they have direct, unmediated introspective access. The process of introspecting delivers erroneous results that conform to the way subjects want to see themselves.

It is difficult to see such confabulation as due to some failure of a monitoring mechanism. It is hard to believe that mishaps in such a mechanism lead to subjects’ being introspectively aware of themselves as having just those beliefs and desires which accord with the way they wish to see themselves. Far more likely, this confabulation of wished-for beliefs and desires is on a par with ordinary cases of self-deception. People interpret themselves in ways that fit with how they want to see themselves and the situations they are in; they become convinced of things about themselves that we have independent reason to doubt or disbelieve.
Such confabulation differs from ordinary self-deception because it results in introspective awareness. Subjects take themselves to be focusing consciously on the contents of their mental lives, and they take these introspective efforts to result in their being conscious of the confabulated states. The HOT model again allows a reasonable and economical explanation. Ordinary self-deception consists of having thoughts that result from one's desire to see things in a certain light. Confabulatory introspective awareness is just a special case of that. It is the case in which one has conscious self-deceptive thoughts about one's own mental states, thoughts whose inferential and motivational antecedents are not themselves conscious. One consciously interprets oneself as being in particular mental states and, because one is unaware of any inference or motivation leading to that self-interpretation, one is introspectively aware of oneself as being in those states.

Erroneous introspective reports are the handle we have on what underlies such introspective awareness, since self-deceptive self-interpretation sometimes leads us to be introspectively aware of ourselves as being in mental states we are not actually in. But many largely accurate cases of introspective awareness doubtless also result from such self-interpretation. People interpret themselves in the light of their situation and past experience, and some of these self-interpretations have to do with what mental states they are in. As long as one remains unaware of whatever inference and motivation leads to these self-interpretations about one's mental states, the self-interpretations will seem, from a first-person point of view, to be spontaneous and unmediated. They will seem to arise from just asking oneself what mental states one is in, from a deliberate decision to focus on the states in question by casting one's mental eye inward. But it is likely that such introspective awareness results in substantial measure from desires to see ourselves in a certain light. Introspection is often, if not always, a process of conscious self-interpretation.

Much interpretation of ourselves occurs without being at all conscious. What is special to introspective self-interpretation is that it is conscious and it pertains to one's own conscious states. Interpreting things in a particular way, whether consciously or not, typically results in those things' seeming to one to be that way. Similarly with self-interpretation in respect of one's own mental states. If one interprets oneself as believing something or wanting a particular thing, typically that is how one seems to oneself to be. Introspective self-interpretation is the conscious case, in which one's interpretation results in one's consciously seeing oneself as being in particular conscious states.

Erroneous introspective self-interpretation occurs not only in introspecting intentional states, but qualitative states as well. Being introspec-
tively aware of an experience of red as simply being of some indiscriminate shade is a case in point. Since there is every reason to suppose that the mental quality of the experience is of some determinate shade, our introspective awareness of the experience is in that respect erroneous. Our relative lack of attention to the state leads us to interpret ourselves as being in a state whose very nature is indeterminate.

A more dramatic case of mistaken introspective self-interpretation of qualitative states sometimes occurs in dental treatment. Patients occasionally seem, from a first-person point of view, to feel pain, although the relevant nerves are dead, missing, or anesthetized. The best explanation is that the patients are conscious of their sensations of vibration and fear as being sensations of pain. The apprehensive expectation of pain influences how these patients are conscious of sensations that would in ordinary circumstance be clear and unmistakable. Once this explanation is given and drilling resumes, patients are conscious of the sensations as being sensations of vibration and fear. Still, their memory of what it had been like for them to have the earlier experience remains unchanged, confirming that what it’s like for one to be in a qualitative state is a function of how one’s introspective self-interpretation represents the qualitative character of that state.30

Consider again the qualitative aspect I am conscious of when I see a red tomato. When I consciously but unreflectively see the tomato, my HOT is not conscious. So the only conscious thought I have relevant to my seeing the tomato is my thought about the tomato itself, which represents it as being red. My only relevant conscious thought interprets the quality I am conscious of as a perceptible physical property of the tomato. By contrast, when I introspect my experience of the tomato, my HOT about that experience is conscious, resulting in an additional conscious thought relevant to my experience of the tomato. And that conscious HOT construes the qualitative aspect of my stream of consciousness not as a property of the tomato, but as a mental property of my sensation. The qualitative aspect I had previously interpreted as a physical property of the tomato I now consciously reinterpret as a mental quality.

VII. PERSONHOOD AND THE UNITY OF CONSCIOUSNESS

Mental functioning is a necessary condition, but plainly not sufficient, for a creature to be a person. Many animals that fall short of being persons nonetheless function mentally, sometimes in fairly elaborate ways. Nor is being in mental states that are conscious by itself sufficient to be a person; it is likely that creatures such as higher mammals are in many conscious states without being persons. Human beings are the only terrestrial animals...
we know of that qualify as persons, but being a person is plainly not the same as being human. It is likely that creatures exist elsewhere that we would not hesitate to count as persons, and certainly such creatures could exist.

What, then, distinguishes persons from other creatures with conscious mental states? One central condition is the capacity not just to be in mental states that are conscious, but to be introspectively conscious of some of those states. Doubtless there is also an ethical dimension to our pretheoretic notion of what it is for a creature to be a person; perhaps, for example, one must be able to assume responsibility or see oneself and others in an ethical light. But whatever the case about that, part of what it is to be a person is having the kind of reflective consciousness that gives one a sense of oneself as a being with a reasonably coherent, unified mental life. Seeing one's mental life as coherent and unified is a crucial necessary condition for a creature to be a person.

A creature can have a greater or lesser capacity for introspective awareness and can vary in the degree to which that capacity results in its seeing itself as having a unified, coherent mental life. Indeed, it is unlikely that there is a single measure of coherence and unity of a mental life, so that a creature's sense of such unity and coherence might well be greater in one respect and less in another. It would be arbitrary to set any particular level or kind of introspective coherence as necessary for a creature to qualify as a person. So introspective unity and coherence, like any other mark of what it is to be a person, admits of degrees. The possibility of a creature's being a person to a greater or lesser extent conforms to our pretheoretic thinking about being a person. Though we count all human beings as persons, it is likely that distant ancestors of ours were persons to some degree, though not as fully as we are, and doubtless the same is true of other, nonterrestrial creatures as well.

Several factors result in the sense we have of our mental lives as unified. One kind of unity occurs independently of any introspective consciousness and contributes little if anything to our sense of mental unity. As already noted, our HOTs operate on many of our mental states not singly, but in large bunches. A vivid example of this occurs in the cocktail-party effect, in which one becomes suddenly conscious of hearing one's name in a conversation that one had until then consciously heard only as part of a background din. One must have been hearing the articulated words of the relevant conversation, though not consciously, since one's attention was drawn to the use of one's name. Indeed, one's name's occurring in any number of other conversations would have exactly the same effect, even though one was also conscious of those conversations only as part of the background noise. So one must have been hearing articulated words in all those
conversations, though not consciously. Though what it's like for one is just the hearing of a background din, one must nonconsciously be hearing very many individual words.

Once again, the HOT model offers an economical and credible explanation. One's HOTs group many auditory sensations together, making them conscious only as an unarticulated bunch. The same doubtless happens with many visual sensations that lie outside the area of foveation; one sees things in large bunches, though an item of special attention can visually jump out at one. Similarly as well as with the other sensory modalities. Doubtless such grouping of perceptual sensations occurs in many creatures other than humans, and occurs independently of any introspective awareness. Perhaps it contributes slightly to a creature's sense of mental unity, since it may produce the impression of being able to focus instantly anywhere in a unified field of sensory experiences. But it is unlikely that this kind of unity contributes anything to what it is for a creature to be a person, since such grouping of sensory experiences is gained at the expense of a drastic loss of conscious information.

There is another way in which we are conscious of our conscious states in bunches which does produce some sense of unity and coherence among those states. When qualitative states are conscious, we are typically conscious of those states not just individually, but in respect of their spatial relations to other states of the same sensory modality. The resulting unity of our sensory fields does not depend on introspective consciousness, nor is it special to persons; many creatures that are not persons have such unity.32

There is, however, yet another way in which we are conscious of our conscious states in groups, and this not only results in a distinctive sense of unity among those states. When we consciously reason, we are often conscious of one intentional state as leading to another. This is not just a matter of our being conscious of the various intentional states themselves; we are, in addition, conscious that these conscious intentional states exhibit a certain connectedness in our thinking. This sense of unity and coherence in one's reasoning is part of what it is to be a person. One could imagine this sense of unity occurring without introspective awareness; perhaps one could be conscious of one's intentional states as inferentially connected simply by having a single HOT about them all. But it is overwhelmingly likely that the awareness of such connections arises often, perhaps always, as a result of our being reflectively conscious of intentional states that are already individually conscious on their own.

Another factor that induces a sense of mental unity that is relevant to being a person is even more closely tied to introspective consciousness. Every HOT represents one as being in some particular mental state, since
each is a thought to the effect that one is, oneself, in that target state. So each HOT makes us conscious of its target as belonging to a self. Our HOTs do not involve any particular conception of the self to which they assign their targets. Indeed, the self that one is noninferentially conscious of mental states as belonging to is no more than a raw bearer of such states; one is not conscious of that self in any other way. And because one is not conscious of that bearer in respect of any other properties, one has a sense that all mental states of which one is noninferentially conscious belong to the same bearer. Since there is nothing that distinguishes the bearer to which one HOT assigns its target from the bearers to which others assign theirs, the HOTs seem to assign their targets all to the same self.

As long as one’s HOTs are not themselves conscious thoughts, one is not conscious of their seeming all to assign their targets to a single self. But, when one comes to be introspectively conscious of one’s mental states, one thereby becomes conscious of those HOTs, and therefore conscious of those HOTs’ seeming to assign their targets all to the same subject. Introspective consciousness leads to our being aware of the way our conscious states are represented as belonging to a single self.33

It may well be that the self we become conscious of our mental states as belonging to is merely notional; perhaps there is nothing that all one’s conscious states belong to other than the entire organism. But that does not matter for present purposes. Being a person is, at least in part, a matter of being conscious of oneself as having a reasonably unified, coherent mental life. And introspective consciousness results in a sense of one’s conscious states as all belonging to a single subject. Every HOT, even when not itself conscious, represents its target as belonging to a bearer indistinguishable from those to which other HOTs assign their targets. And introspective consciousness makes us aware of this feature of the way our HOTs represent their targets. HOTs in effect interpret the states they are about as all belonging to a single self whether or not any such self exists.34

Introspective awareness interacts with speech in a way that seems to underwrite the traditional idea of mental states as transparent to the self. Being able to express one’s introspective consciousness in speech amounts to being able to report the mental states one introspects. If, for example, one is introspectively aware of thinking that p, one can say that one thinks that p. Saying that one thinks that p is plainly not semantically the same as simply saying that p, since the two speech acts differ markedly in truth conditions. Still, the two types of speech act have roughly the same conditions of use, or performance conditions. So there is a tendency to regard the two speech acts as being in some way equivalent. This tendency has dramatic consequences.

Every speech act expresses an intentional state that has the same content and a mental attitude that corresponds to the illocutionary force of the
speech act. So, if one regards saying that one thinks that \( p \) as equivalent to saying that \( p \), one will extrapolate to the corresponding thoughts; one will also regard the thought that \( p \) as equivalent to the thought that one thinks that \( p \). But that would mean that the thought that \( p \) would not only have its ordinary content, but would also be literally about itself; the thought would thus constitute an awareness of itself. Seeing the speech acts of saying that \( p \) and saying that one thinks that \( p \) as equivalent gives rise to the traditional idea that intentional states are all transparent to the mind.

The temptation to assimilate the two types of speech act arises only when a creature has the ability to express its introspective consciousness in speech. Only then can one report one's intentional states noninferentially, and thereby treat the two speech acts as having roughly the same use. But, despite that rough performance-conditional equivalence, the two speech acts do differ semantically. The illusion of Cartesian transparency that comes with the ability to express one's introspective consciousness in speech results from failing to note that there is more to the role speech acts play than their conditions of use.35

NOTES

For work on the phenomenon of blindsight, see Lawrence Weiskrantz, Blindsight: A Case Study and Implications (Oxford: Oxford University Press, 1986), and Consciousness Lost and Found: A Neuropsychological Exploration (Oxford: Oxford University Press, 1997).


10. E.g., in Peter Carruthers, *Language, Thought, and Consciousness: An Essay in Philosophical Psychology* (Cambridge: Cambridge University Press, 1996). Carruthers’s actual formulation appeals to dispositional HOTs, but it is unclear what a dispositional thought could be other than a disposition for an episodic thought to occur.

Carruthers also argues that appeal to dispositions would allay a concern that there is insufficient computational capacity to accommodate HOTs for all our conscious states. But it is unclear why dispositions to have HOTs would take up less cortical space than the HOTs themselves. It is also unclear why, with our huge cortical resources, there is reason for such concern. But if there is, it may well be that many HOTs operate wholesale on bunches of conscious states. This is phenomenologically plausible, since our conscious states seem to be conscious in clusters, and focusing on a small area seems to withdraw consciousness from previously conscious peripheral states. I return to this possibility in sec. VII, in connection with the apparent unity of consciousness.

11. It is likely that such appeal to dispositions is, at bottom, a way to avoid confronting the popular intuition that thoughts are invariably conscious. Theorists tempted by the idea that mental states are all conscious sometimes appeal to dispositions to try to disarm arguments that nonconscious mental states actually occur. Dispositions to be in occurrent mental states play much the same causal roles as those played by the occurrent mental states themselves. So such theorists, while holding onto the claim that all occurrent mental states are conscious, allow for nonconscious dispositional states corresponding to the ordinary conscious versions. It is tempting to see Carruthers’s explanation of conscious states by way of dispositions to have HOTs as a way to avoid positing occurrent, but nonconscious, HOTs.

A well-known example of appeal to dispositions in place of nonconscious mental states can be found in Searle’s connection principle, on which all intentional states are potentially conscious. On that principle, nonconscious states can have only an ersatz intentionality derived from their connection with intentional states that are conscious. A state is never intentional without being conscious, according to Searle, since not being conscious would deprive it of its subjectivity. See John R. Searle, “Consciousness, Explanatory Inversion, and Cognitive Science,” *The Behavioral and Brain Sciences* 13, 4 (December 1990): 585–696. Searle’s argument for the connection principle relies on his claim that only conscious states can differ in the way they represent the same thing. But how a state represents something will have effects on behavior and other mental states whether or not the state in question is conscious.

For another example of holding that a mental phenomenon can fail to be conscious only if it is dispositional, see Norman Malcolm, “Thoughtless Brutes,” *Proceedings and Addresses of the American Philosophical Association*, 1972–73, XLVI (November 1973): 5–20, esp. sec. ii.

12. It is natural to construe the phrases ‘conscious of’ and ‘conscious that’ as factive, so that one is not conscious of something unless that thing exists, and not conscious that something is the case unless it is. But even if being conscious of something is factive in that way, being conscious of something as having a particular property does not imply that it has that property.
In any case, many verbs that are usually used factively also get used nonfactively; we speak of seeing pink elephants, as well as things about whose existence we are non-committal. It is likely that, whatever the ordinary usage of ‘conscious of’ and ‘conscious that’, we must treat these phrases as being nonfactive in developing a satisfactory account of what it is for a mental state to be conscious.

13. More precisely, though we may sometimes speculate about what led to our introspecting a state, we are never introspectively aware of the process leading to our introspective awareness.

It might seem more accurate to describe introspective focusing as voluntary, rather than deliberate. But this characterization is misleading, since introspective focusing is not properly an action at all.


15. Block’s apparent identification of monitoring with attentive introspection would also explain his claim that monitoring consciousness is somewhat intellectualized (234), as well as his denial that monitoring consciousness figures either in phenomenal or access consciousness (279, 280).


17. Dretske in effect concedes that this can happen, since he holds that being conscious of a difference always amounts to being conscious “that such a difference exists” (“Conscious Experience,” 275; cf. 266–67). And, as he notes, being aware of the thing in virtue of which the experiences differ does not mean being aware that they differ in that way (275).


20. John R. Searle, The Rediscovery of the Mind (Cambridge, Mass.: MIT Press, 1992), 97. Cf. 144: “In the case of vision, we have a clear distinction between the object seen and the visual experience . . . But we can’t make that distinction for the act of introspection of one’s own conscious mental states.” Cf. also 96.


I am grateful to Shoemaker, personal communication, for pressing various issues about representationalism.


25. Paul A. Boghossian and J. David Velleman, “Colour as a Secondary Quality,” *Mind* XCVIII, 389 (January 1989): 81–103, 95; reprinted in Alex Byrne and David Hilbert, eds., *Readings on Color*, volume 1: *The Philosophy of Color* (Cambridge, Mass.: MIT Press/Bradford Books, 1997), 81–103. They continue: “The most plausible hypothesis about what somebody means when he calls something red, in an everyday context, is that he is reporting what his eyes tell him. And according to our account, what his eyes tell him is that the thing has a particular visual quality, a quality that does not actually inhere in external objects but is a quality of his visual field” (100). Boghossian and Velleman claim Galileo, Locke, Newton, and Hume as early exponents of such projectivism about color (96–97, and esp. notes 15 and 16).

I am grateful to Christopher S. Hill for raising the issue about projectivism.

26. Pre-Galilean common sense acknowledges the distinction between perceptible properties as they are in themselves and as they appear to us. Thus, Aristotle distinguishes the potentiality of physical colors and other so-called proper sensibles, which is independent of their being perceived, from the way those perceptible properties are actualized in perception (e.g., *De Anima* III, 2, 425b26–426a1). So the post-Galilean temptation to deny color to physical objects does not result from any prior failure to countenance that difference. Rather, that temptation is due to another, independent Aristotelian claim, that the color properties actualized in physical objects are the very same properties as color qualities actualized in the soul (ibid.). And the post-Galilean insistence that physical reality is mathematically describable precludes those properties being the same. But the need to deny color properties to physical objects is removed once we recognize that the perceptible colors of physical objects are distinct kinds of property from the mental colors of visual experiences.

For more on these issues, see “Sensory Quality and the Relocation Story.”

27. This is not just a matter of focusing attentively on the exact shade of a physical object’s color; focusing attentively on hallucinatory sensory experiences is also enough to reveal the exact shade of the mental quality.

232
28. Consciousness and Experience, 30.


30. There are two places at which error might enter erroneous introspective awareness. It might be that one’s initial HOT is mistaken about what mental state one is in or that one’s introspective third-order thought is mistaken about what state one is conscious of oneself as being in. In the second case, one would have an accurate second-order thought but an erroneous third-order thought.

Perhaps confabulatory error can enter at the higher, introspective level. But it is not easy to see a reason to think that it happens there, as opposed to at the level of ordinary, unreflective consciousness. Nor it is easy to see how that question could be decisively settled short of isolating the different thought events neuroscientifically.

Eric Schwitzgebel has argued (personal communication) that the third-order level is more likely, but his argument seems to rest on his assumption that mental states, though not automatically subjects of introspective consciousness, are invariably conscious. See also his article in this issue of Philosophical Topics.

31. In connection with Carruthers’s dispositional version of the HOT hypothesis, n. 10.

32. Note that no sense of the unity or coherence of consciousness will result simply from one’s qualitative states’ standing in some spatial relations to one another. Such a sense arises only if one is conscious of those states as standing in those relations. Indeed, their actually standing in such relations is not even necessary for such a sense; all that matters is that they are conscious as standing in those relations. For problems about the way we are conscious of qualitative states as spatially unified within sensory fields, see my “Color, Mental Location, and the Visual Field,” Consciousness and Cognition 9, 4 (December 2000), sec. 4.

33. It is worth noting that Hume’s famous problem about the self resulted from his tacit adoption of a specifically perceptual model of introspecting; one cannot find a self when one seeks it perceptually. See David Hume, A Treatise of Human Nature, ed. L. A. Selby-Bigge (Oxford: Clarendon Press, 1888), Book I, Part IV, sec. vi, 252.


35. For more on the implications for consciousness of the distinction between these two types of speech act, see my “Thinking That One Thinks” and “Why Are Verbally Expressed Thoughts Conscious?”, forthcoming in my Consciousness and Mind (Oxford: Clarendon Press).