

Joseph Levine

1 Consciousness certainly is connected with awareness. In fact, some people would say the two terms are synonymous. To be conscious of something is to be aware of it. Conscious mental states are those we are aware of. From these simple platitudes comes the motivation, or intuitive support for theories of consciousness built on the notion of representation, whether it be representation of the conscious states themselves or of their objects. Of course a crucial element linking the platitudes about awareness with representationalist theories is the thought that awareness can be captured or understood in terms of representation. I don't deny that awareness—conscious awareness, that is—entails representation; how could it not? What would it mean to be aware of something without somehow representing it? What I am suspicious of, however, is the idea that awareness is exhausted by representation. It seems to me that there is something going on with awareness that takes us beyond mere representation—or maybe it's a matter of a special kind of representation—that leaves every version of representationalist theory I know of inadequate as a theory of consciousness.

The plan for the essay is as follows. In sections 2 and 3 I will characterize how the notion of representation has been employed in theories of consciousness, identifying two sources of motivation for appealing to a notion of self-representation in the process. In section 4 I will describe three levels at which conscious awareness operates, and speculate on ways of formulating an integrated account of all three involving the notion of self-representation. In sections 5 and 6, I'll try to vindicate the suspicion mentioned in the paragraph above.

2 The simplest way of reducing awareness to representation is by what can be called "externalist representationalism." Consider a paradigm conscious mental state, such as seeing a ripe tomato sitting on the kitchen counter. One natural way to characterize the state is this: one is (visually) conscious of the tomato. What this amounts to on the representationalist theory is that one's

visual system is representing (something to the effect) that there is a red figure of a certain shape and texture in one's visual field.

Now there are two features of conscious sensory states that require theoretical elucidation: *qualitative character* and *subjectivity*. The former is implicated in the distinction between sensory states and nonsensory cognitive states like thoughts and beliefs, while the latter is implicated in the distinction between conscious and unconscious mental states. In the case at hand, seeing a ripe tomato, there is both a distinctive qualitative character to be reckoned with and also the fact that the state is conscious—"for the subject"—in a way that unconscious states are not.

Representationalism tries to solve the problem of qualitative character by putting the qualia out in the world, so to speak. For one's experience to have a reddish qualitative character is just for one's experience to be representing red (or some specific shade of red). To distinguish seeing red from thinking about it, which also involves representing it, certain further constraints on the format and content of sensory representations are imposed. A popular move is to claim that sensory representations involve so-called nonconceptual content, and this is what distinguishes them from thoughts, which have conceptual contents. Sometimes this idea is joined with the idea that sensory representations employ a more "pictorial" format than the more "symbolic" format of higher-level cognitive capacities (Dretske 1995a; Tye 1995—for criticism of their externalist representationalism, see Levine 2003).

On the question of subjectivity, what distinguishes conscious from unconscious states, externalist representationalists differ. Those who eschew any type of second-order representationalism opt for some condition on the relative availability of the sensory representations to higher-level cognitive processes, particularly those involving planning, action, and speech (e.g., Tye 1995). So, for instance, what makes the visual experience of the tomato conscious, and, say, the apprehension of an X in the blind visual field of a blindsight patient unconscious, is the fact that one can spontaneously report on the tomato, use the information in the visual sensation to plan one's lunch, and the like. In this sense the information about the tomato is there for the subject in a way that the information about the X is not in the blindsight case.

Many philosophers feel that the mere availability to higher cognitive processes cannot properly capture the sense in which a visual experience such as seeing the tomato on the counter is a conscious state. They feel that aside from the dispositional property of being, as it were, ready to hand for planning and the like, there is a more active, or categorical sense in which the state is a part of one's current awareness. Conscious states, on this view, are

states one is conscious of, not merely states through which one is conscious of objects and events like the tomato on the counter.

There are many versions of higher-order views, but one significant division is between those that posit a distinct second-order state that represents the first-order state and those that maintain that the conscious state somehow represents itself. On both views, the consciousness, or awareness, is constituted by the representation of the target state. To be conscious of seeing the tomato is to represent that one is seeing it. The difference, again, has to do with whether this representation of the fact that one is seeing it takes place in a separate state or is a feature of the visual state itself.

Two-state higher-order views have been subjected to extensive critique in the literature. One reason for exploring the viability of a one-state, self-representational higher-order view is that it may seem to overcome some of the problems attending the two-state view (see Kriegel 2003b and this volume). One advantage in particular is that it seems to better capture the phenomenological situation.

Consider again the simple example above. I consciously see a ripe tomato on the kitchen counter. Clearly the primary object of my conscious state is what I'm seeing, the scene on the counter. But it also seems to be the case that my very awareness of the tomato on the counter includes within it somehow an apprehension that I am seeing what I'm seeing. That certainly seems to be part of what it means to say this visual state is conscious. Granted, both objects of awareness, the external situation and the visual state itself, are made objects of awareness by the two-state view. However, what isn't captured by that view is the way these two acts of apprehension seem intimately connected within a single state of mind. Only the one-state, self-representational view seems to capture this.

3 So far we've been discussing the relation between awareness and representation in the context of attempts to provide a theory of consciousness itself—of what makes a mental state a conscious state. But there is another dialectical route by which one can arrive at the idea that self-representation is involved in our awareness of our conscious states. It's worthwhile to incorporate this other route into our discussion, since in the literature on consciousness these two sources of motivation for appealing to self-representation have been important. It would be interesting to see how the two concerns relate to each other.

The other route has to do with what are called "phenomenal concepts." We begin with the problem of the explanatory gap, or, what for our purposes amounts to the same thing, the "intuition of distinctness" (see Levine 1983; Papineau 2002). When we consider the qualitative character of a visual color

sensation, such as seeing that ripe tomato on the kitchen counter, and compare that property of our experience with a third-person characterization of what is going on in the relevant part of our brain's visual system at the time, the two seem utterly disparate. The idea that we are talking about the same state, or property, or that one can explain the reddish qualitative character by appeal to what is going on in the brain, just seems bizarre. Since materialists are committed to the claim that these are the same state or property, or that the qualitative property is realized in the brain property (and so therefore should be explicable in terms of it), they need to explain away this strong reaction against the relevant identity or explanatory claims.

In developing this critique of materialism it is common to contrast the qualia-brain state case with other instances of theoretical identifications. There doesn't seem to be any special problem swallowing the idea that water just is H_2O , or in explaining any of the superficial properties of water in terms of its molecular structure (together with other chemical facts and laws). So, the argument goes, what accounts for this special intuitive resistance when it comes to this theoretical identification/reduction? Dualists have a ready response: it's because qualitative mental properties really are distinct and irreducible to physical or functional properties, connected to them only by brute, fundamental laws, that the former cannot be explained in terms of the latter. The burden is then on the materialist to come up with another explanation.

This is where the appeal to phenomenal concepts comes in (Loar 1990; Tye 2000). Qualia, it is argued, are nothing over and above physical (or functional) states of the brain. However, there is something special about them; it's a matter of our cognitive access to them. Unlike other objects and properties, we have special first-person access to qualia, which means that the concepts we form of them through this route—phenomenal concepts (which means here concepts of phenomenal properties, not concepts that are themselves phenomenal—though wait a minute)—are different in important ways from other concepts. This difference in kind is then used to explain our cognitive inability to appreciate the relevant psycho-physical identity/reduction.

Earlier I said that there were two aspects to the problem of consciousness: qualitative character and subjectivity. Insofar as versions of representationalism have been employed to address both aspects, we see a close connection between them. But now we see another close connection. The problem of qualia immediately leads to the problem of subjectivity, in that we find ourselves in need of an account of the latter to explain away certain problems

with the former. Higher-order theories, in attempting to provide an account of the distinction between conscious and unconscious mental states, addressed subjectivity directly through the device of representation. But even those theorists who eschew the higher-order account of what makes a state conscious find themselves in need of a theory of higher-order representation to overcome the intuition of distinctness and the explanatory gap. In fact, both higher-order theories of consciousness and theories of phenomenal concepts face the same issue: how to understand the special cognitive relation that attends the first-person point of view.

One aspect of this special, first-person relation that seems crucial is something we might call, for lack of a better term, "cognitive immediacy." There seems to be a more intimate cognitive connection between the subject and what she is conscious of, or the consciousness itself, than is present in other circumstances. Of course "immediacy" and "intimacy" are both metaphors, and it is notoriously difficult to characterize this feature in an illuminating way. Yet it does seem to be the core feature that distinguishes first-person access from other forms of cognitive access.

In the case of higher-order theories of consciousness, one way the issue of immediacy arises is this: it's clear that not just any representation of a mental state makes it conscious. The standard example is one's coming to believe that one has repressed desires in therapy, even though the desires themselves are still unconscious. The kind of representation of a mental state that constitutes being consciously aware of it has a special immediacy about it. It is thus a burden of the higher-order theory to provide an illuminating account of what this cognitive immediacy comes to.

In the case of phenomenal concepts there is also a need for an account of the special first-person cognitive immediacy. The problem is that merely appealing to a difference in the concepts by which one apprehends a phenomenon like one's occupying a certain sensory state is not sufficient to explain the particular cognitive resistance we feel to reduction and identity claims in the psycho-physical case. After all, the way we access the substance we call both "water" and " H_2O " also involves two very different concepts. Why should it be so hard to see how a phenomenal property is identical or explanatorily reducible to a physical property yet not at all hard to see how water could be H_2O ? The answer purportedly lies somewhere in the idea that first-person access, by way of phenomenal concepts, provides an especially immediate form of access to sensory states that thereby makes them seem so different in kind from the sorts of states we conceptualize in third-person theoretical terms. (Whether or not this really answers the question is a topic to which I will return in section 5.)

Bringing these two concerns together—the need for an account of immediacy both for higher-order theories of consciousness itself and for phenomenal concepts—we converge on the idea of self-representation. Perhaps what is special about the kind of representation involved in being aware of one's sensory states is that it is that very state that is representing itself, not a distinct state as in standard versions of higher-order theory. Similarly, the sort of first-person access involved in apprehension of a qualitative state by way of a phenomenal concept might involve using that very qualitative state to represent itself. Certainly it's hard to see how one can get more "immediate" a form of cognitive access than having a state represent itself. Perhaps then the grand unifying theory of consciousness and the explanatory gap can be built around the notion of self-representation. Well, let's see.

4 Before embarking on an evaluation of self-representation theory, I want to address another question. We identified two points at which cognitive immediacy becomes an issue: at the awareness of one's experience that makes it conscious in the first place and at the awareness that is constitutive of possessing a phenomenal concept. How do these two forms of self-representation relate to each other? Are they the same thing?

On the face of it, one would think not. On one theory, self-representation is what distinguishes conscious states from nonconscious states. Having a conscious experience is itself a matter of occupying a state that reflexively represents itself. On the other theory, self-representation is used to distinguish phenomenal concepts from other ways of representing conscious experiences. On this view, there isn't anything particularly reflexive about a conscious experience *per se*. Rather, it's when we adopt an explicitly introspective attitude toward the experience that self-representation comes into play.

These two theories answer to different concerns as well. The first type of self-representational theory is attempting to analyze what it is to be conscious. It is motivated by the relational approach of higher-order theory, but departs from the standard version in locating the representation relation that is constitutive of conscious awareness inside a single state. The theory of phenomenal concepts proposes no analysis of conscious awareness itself. Conscious awareness is taken to be identical with some physical or functional property, whatever that may turn out to be. The entire appeal to self-representation has to do with explaining why there is cognitive resistance to accepting the identity of conscious awareness with whatever it turns out to be, described in whatever theoretical vocabulary turns out to be appropriate to it. So why would one expect a single account to satisfy both of these quite different motivations? Again, on the surface it doesn't seem likely.

Yet, despite these differences, it seems to me that there ought to be a way to integrate them. As I pointed out above, both seem to appeal to self-representation as a way to realize a kind of cognitive immediacy that is supposedly distinctive of the first-person point of view. It would be extremely odd if these two domains in which this distinctive cognitive immediacy were present weren't intimately connected. In fact, whether or not one accepts self-representation as an account of first-person cognitive immediacy, the two phenomena to which the two self-representation theories are addressed ought to be capable of integration. The fact that our access to our conscious experiences is special when we introspectively reflect on them must have something to do with the fact that consciousness involves a kind of self-awareness in the first place.

Just to complicate the story a bit more, let me add a third kind of cognitive immediacy that seems involved in conscious experience. Consider again our example of a typical conscious experience, my seeing a ripe tomato on the kitchen counter. We have focused so far on the fact that there is a qualitative character to that experience, and that this character seems to be something of which I'm aware. But another feature of the experience is the way that the ripe tomato seems immediately present to me in the experience. I am not in any way aware of any cognitive distance between me and the scene in front of me; the fact that what I'm doing is representing the world is clearly not itself part of the experience. The world is just there.

This sense of the immediacy afforded by sensory experience is part of what fuels the transparency argument employed by externalist representationalists about qualia. Look, they say, the redness isn't experienced as a feature of your visual state, but rather as a feature of the tomato. It's all out there. Isn't that what you find when you reflect on your experience? they argue—persuasively, to my mind. I don't intend to engage externalist representationalism here (see Levine 2003); I just wanted to note that transparency and the immediacy of the objects of conscious awareness seem to be part of the same phenomenon. When we encounter the world in perception, it doesn't seem to be merely represented by us, but *presented* to us.

So we've got three levels of representation in play; let's just call them levels one, two, and three. Level one is the representation of the world outside by a visual experience; the spatial layout, textures, and colors constitutive of the scene involving the ripe tomato on the counter—call it "spatial layout L." Level two is that awareness of the experience that seems inseparable from any experience, whether made the focus of attention or not. Level three is the representation of the experience when it is made the object of explicit introspective attention, as in the sorts of judgments concerning identity and

reduction for which appeal to phenomenal concepts is made. Can we integrate them all into an account of conscious cognitive immediacy that is based on self-representation?

Here's my quite tentative proposal. To begin with, it's important to remember that any claim to immediacy in the level-one content has got to be tempered by the recognition that perception cannot really afford immediate, or unmediated access to the world. After all, there are hallucinations and illusions. It can look to me for all the world as if there is a ripe tomato on the counter in front of me even when there isn't—even when I'm sitting in a completely dark room, so long as my brain occupies the relevant state. No doubt perceptual experience *seems* to deliver the world directly to me in this very hard-to-characterize way, but it can't be just as it seems, or hallucinations and illusions wouldn't be possible.

The point is that it isn't immediacy itself, perhaps, that requires explanation, but the *appearance* of it. So how does this help? How does the fact that the explanandum is now the appearance of immediacy at level one make it easier to see how appeal to the self-representational content alongside the externally directed content serves to explain it? Well, it might be that this appearance of immediacy between subject and external object is a kind of cognitive illusion engendered by the *actual* immediacy possessed by the self-representational content itself.

To see what I mean, let's turn now to level two. As I look at the ripe tomato on the counter—the whole scene, spatial layout L, being the level-one content of my experience—I am simultaneously aware, alongside the primary awareness of L, of the fact that I am now having a visual experience as of L. Consider just this secondary self-awareness for a moment. This level-two representation doesn't seem to involve making the visual experience a full-fledged object of awareness in its own right, the way that the experience does with the spatial layout L or the level-three introspective state does with the level-one state. Yet, it does seem to involve mentally registering, in some way, that one is having the experience.

One feature that seems to go along with this intermediate character—that is, that it involves a kind of registering yet without focal attention—is the idea that the awareness of the experience is somehow intrinsic to the experience itself. Here we have the most intimate form of cognitive immediacy in the entire three-level structure. Conscious visual experience seems to be just the type of state that must, by its very nature, be also apprehended as it affords apprehension of the external scene. The idea that this secondary, level-two kind of representation should be self-representation—the experience representing itself—certainly seems inviting.

But of course appeal to self-representation here still only serves to explain, if anything, the immediacy and intrinsicity of the level-two form of awareness itself. How would this help with level one, the awareness of the tomato on the counter? Consider what it is we are aware of at level two. We said that the representational content at issue is something to the effect that I am having (or it is now occurring—whether explicit reference to a subject is included is not clear) a visual experience as of spatial layout L. But this means that it's the content of the level-one representation that is the object of the self-representation. This fact might be the key to the apparent immediacy of level one itself.

Perhaps what is going on here is a kind of transparency phenomenon, a connection already noted above. That is, the self-awareness of level two involves “seeing through” its object, the level-one content, to what that content is about—but only apparently, of course. In our apprehension of the visual experience, a relation that itself has genuine cognitive immediacy, we somehow transfer that sense of immediacy to the external scene, the spatial layout L represented by the level-one content of the experience. One way to put it is this: what is genuinely immediately “open” to us, in the context of a visual experience, is the intentional object corresponding to layout L. We interpret this, in the experience, to mean that L itself is immediately open to us. But this, it turns out, is only apparent, a kind of cognitive illusion.

The foregoing account is wildly speculative of course. My goal here is to explore how far we can go in explaining the peculiar nature and structure of a conscious experience by appeal to the notion of self-representation. Since something we've been calling, for lack of a better name, “cognitive immediacy,” seems to be one of the crucial features at each of the three levels of representation associated with a conscious experience, it would make sense that whatever form of self-representation posited as essential to consciousness should somehow unify these phenomena. The story above is an attempt to show how this might be done at least for the first two levels.

Let's turn now to the connection between levels two and three. We're assuming, at least for the sake of argument, that some appeal to the immediacy of the relation between phenomenal concept and phenomenal state would explain the cognitive resistance that is naturally felt when considering certain psycho-physical identity statements or explanatory reductions. Some philosophers then go on to appeal to the idea of self-representation to explain the relevant sense of cognitive immediacy (Block 2002; Papineau 2002). The question I want to consider now is whether or not the self-representation at issue here is plausibly identified with the self-representation involved with the first two levels.

Here's roughly what they have in mind. In contexts where I find myself puzzling over how this qualitative state could be a neurophysiological state (or something along these lines), I am actually using a token of the very qualitative state type at issue to represent that type (or the relevant property). Papineau (2002) compares what's going on here to quotation, where a token of the type is used to refer to the type itself. So the question is what relation this form of self-representation—involving in the exercise of a phenomenal concept—bears to the self-representation involved in the mere having of a conscious experience.

Before we address that question, we need to be somewhat clearer about how to understand the self-representation theory of phenomenal concepts. When I entertain the thought that reddish experiences might be identical to (or realized by) neural states, do I literally have to be having those very experiences? Right now I am perfectly capable of wondering about this question with respect to the look of ripe tomatoes, but I'm not in fact now looking at a ripe tomato. But if I were employing a token of that experience type to represent it in thought, wouldn't I have to be right now having the experience?

Of course sometimes I am having the experience when I think about it. Often it is the case that it's as I'm staring at that ripe tomato that I find myself puzzling about the experience's alleged physical nature. So perhaps in those cases the theory makes sense. But what about the other times, like now, when I can find myself having gappy worries about visual experiences of tomatoes without actually looking at one? Well, it's plausible that even now—and in my own case I find it borne out—when not actually visually experiencing the ripe tomato, what's going on is that I'm calling up an image of a ripe tomato and focusing on that. If this phenomenon is general, if in order to get into the relevant puzzling state of mind, to have gappy worries, it's necessary that one at least imagine the relevant experience, then perhaps that will do. We can say that the image token is representing a certain type that is shared by the full-fledged experience. There may be problems with this move, but let's accept it for the moment.

The question before us now is how to understand the relation between the self-representational nature of conscious states themselves and that of phenomenal concepts. One possibility is that the very same representation that captures the awareness inherent in a conscious state—*qua* being conscious—is itself the form of representation employed when thinking about a conscious state through a phenomenal concept: that is, perhaps levels two and three collapse. If one could make a case for this proposal,

it would certainly unify the various phenomena at issue and to that extent provide a more satisfying theory.

Offhand this proposal doesn't seem promising though. What is supposedly special about level two is that it doesn't involve the kind of focused attention that is clearly involved when expressing gappy worries. How could that, as it were, off-to-the-side, by-the-way kind of awareness that seems to attend every conscious experience be the very same form of representation that is involved when I think hard and explicitly about an experience? The two kinds of awareness seem so different in character that it's hard to see how the very same representational means could underlie their exercise.

Yet, it really would be odd if these two kinds of awareness, both possessing that kind of immediacy that suggests self-representation, weren't at root the same. I am tempted again to speculate, as follows. The basic phenomenon here is the awareness/self-representation of level two. This is where whatever is peculiar to conscious experience seems to reside. This is the level where subjectivity, the experience being "for me," has its home. Somehow, when I focus on my experience in an explicitly introspective way, that very awareness which is only of the inattentive sort when my attention is directed outward, is built into a full-fledged explicit awareness on that occasion.

What I imagine, and again this is quite speculative, is something like the following. We begin with a state that has two contents, the external content and the reflexive content (corresponding to awareness levels one and two). Given the nonfocal nature of level-two awareness, it makes sense to characterize the reflexive content as a secondary content, with the primary content being the external one. What "primary" and "secondary" amounts to would have to be spelled out in functional terms. That is, it would be a matter of the role the state played that determines which of its two contents is primary. In the case of normal perceptual experience, the relevant state is primarily being used for the information it's conveying concerning matters external.

However, on occasion, one is concerned mostly about the state itself. I'm having an experience and it's telling me about the external world, but my interest is in the experience, not in the situation it's telling me about. Suppose this functional shift involved treating the reflexive content as primary and the external content as secondary. Level-three awareness, then, need not involve the introduction of a new representational vehicle, but rather a change in the functional role of the vehicle that is already there. This would be as tight a connection between levels two and three as you could get, and therefore an especially appealing way of integrating the concerns that motivate positing these two levels in the first place.

Let me summarize. Conscious experience seems to involve a special, immediate cognitive relation on three levels. First, there is a relation between the subject and the primary object of her conscious state, whatever state of the external world is presented to her in the experience. Second, in the very act of being conscious of some state of her external environment, she is simultaneously aware of her conscious experience itself. Third and finally, when explicitly contemplating her conscious experience, her cognitive relation to the experience seems to partake of the very same sort of immediacy as manifested at the other two levels.

I have speculated that the fundamental form of immediate awareness is that of the second sort, level two. The immediacy that apparently attaches to the first level is really a kind of mistaken projection from the second level. We are immediately aware of our own experience, but since what this experience consists in is a representation of the world around us, that content is presented as if it too were immediately present. That this is only apparent is demonstrated by the possibility of hallucination and illusion.

On the other side, the idea is that explicit second-order forms of awareness (which we've called level-three representation), the kind involved in full-fledged introspection of a conscious experience, is somehow constructed out of the awareness already present at level two. This would explain why level-three awareness would also possess the immediacy needed to distinguish it from other forms of cognitive access. The fundamental principle then is that the awareness that makes a state conscious, a state for the subject, is what gives it its presentational character *vis-à-vis* the external world and also its puzzling character *vis-à-vis* physicalist theories. What endows a conscious state with this fundamental sort of awareness, its subjectivity? The hypothesis on the floor is that what makes it conscious, subjective, in this way is that the state in question, in addition to representing whatever external content it represents, also represents itself.

5. According to the account above, phenomenal concepts achieve their special cognitive immediacy through the mechanism of self-representation. It's worth asking, however, just what work cognitive immediacy, implemented through self-representation, is supposed to do here. Sometimes it's said that qualia serve as their own modes of presentation (Loar 1990). Others emphasize how getting oneself into the gappy state of mind, and thus employing phenomenal concepts, usually involves either currently having the relevant experience, or conjuring up an image that shares certain qualitative features with the original experience (Papineau 1995), so that one is in effect saying, "but how could *that* be a physical/functional state/property?" as one contemplates the experience (or the image) itself.

But how is this connection with the conscious state itself supposed to explain why conceiving of it through this means produces cognitive resistance to the idea that it is identical to, or explicable by reference to, a physical state? After all, suppose I look at a glass of water and consider the claim that it is identical to H₂O. It doesn't seem as if the fact that I'm currently encountering it through one of its superficial properties, its appearance, makes it especially difficult to accept the idea that it is composed of invisible molecules two parts hydrogen and one part oxygen. Why is the special connection between a phenomenal concept and an instance of the experience itself, or an image qualitatively similar to it, of particular significance in explaining cognitive resistance to both explanatory and identity claims?

Of course, one explanation for the significance of immediacy in accounting for the relevant cognitive resistance is not friendly to materialists. One is tempted to say, as we reflect both on an experience we're having and its neurological or functional description, that the first-person way of conceiving it provides us a glimpse of something left out of the third-person theoretical description. We see it as it really is—capture its essence, as it were—in a way that can't be accomplished otherwise. But this account implies that some feature of the experience really is left out of the theoretical description and can be cognized only through the conscious experience itself. This can't be what the materialist has in mind.

Some philosophers (Perry 2001a) have compared the way we conceive of our conscious states in the first-person mode to indexical representations. This of course would fit the self-representation model very well, since self-representation is a kind of indexical representation. But, as I've argued elsewhere (most recently, Levine forthcoming), this doesn't help much either. For one thing, we have all sorts of informative identity claims involving indexicals that do not cause the kind of cognitive resistance we seem to face with psycho-physical identity/explanatory reduction claims. "Where are we?" the lost driver asks, of course knowing that she's "here." She is then informed by the passenger that they've just crossed a certain intersection, say Main Street and 9th Avenue. There isn't any problem identifying "here" with the nonindexical description in terms of street names. Imagine the puzzled reaction she'd get if she complained, "but how could Main and 9th be *here*?" (unless, of course, she had some reason to think she was quite some distance from that intersection, which is a different kind of puzzlement entirely).

There is another problem with the appeal to indexicals, which relates to a matter that will come up again later. Indexicals and demonstratives are inherently subject to ambiguity. When you point at an object and say "this," what is picked out is not a function only of the demonstrative and the

pointing, but also of the intention behind the utterance. With reference to individuals, the ambiguity can probably be resolved by the addition of a sortal, such as in "this coffee mug." But when dealing with what Loar calls a "type demonstrative," where you are picking out a kind, or type, by virtue of demonstrating an exemplar, or token, the problem of ambiguity becomes much more intractable. Each token is a token of an indefinite number of types. Which one are you demonstrating? Even if you add an experience type, that won't pin it down sufficiently since each token experience is a token of many different experience types. The more constraining information you load into the intention with which the demonstrative is employed to cut down on the possible ambiguity, the less work is done by the demonstrative itself. My bet is there's no way to successfully refer to the type intended by "this (type of) experience" without already having the type in mind. But then the demonstrative does no work at all.

Somehow, the very fact that entertaining a phenomenal concept involves occupying the phenomenal state itself is supposed to engender a special cognitive relation, one that prevents establishing the kind of conceptual link that comes with identity and explanatory reduction claims. Why should that be, if it isn't, as the dualist claims, that first-person modes of presentation actually present something we can't get any other way? I don't pretend to have a good answer to this question. But having pressed it so far, let me take the following account as at least a first step toward an answer. What I will argue is that even if we find this account satisfactory, any form of self-representation that is plausibly naturalizable won't really fill the requisite role.

What we want to avoid clearly is the dualist challenge that first-person access puts us in touch with a feature of our mental state that cannot be accessed in any other way. We don't want anything genuinely new to pop up there. So if what's represented isn't new, the phenomenon causing the relevant cognitive resistance must reside in the mode of access itself. Perhaps there is some unique cognitive relation that is created when a state is used to represent itself. Let's call this relation, for lack of a better word, and for obvious historical reasons, "acquaintance." The claim then is that we have two modes of access to phenomena: acquaintance and (well, why not?) description. It is simply a fact about ourselves that we cannot conceptually link what we are acquainted with to what we describe even if the two are really the same thing.

As I said, I don't pretend to know quite how this story is supposed to go. But however it goes, I don't see how any standardly naturalizable account of self-representation could implement it. There are three ways to naturalize

representation that I know of: causal history, nomic dependence (with perhaps some historical-teleological elements added), and functional role. The first two are nonstarters for self-representation (but see below). Causal history is a mechanism that works most naturally for singular terms. Current uses of "Joe Levine" refer to me because they have a causal ancestry leading back to an original baptism, or something like that. Nomic dependence grounds information that, when appropriate bells and whistles are added, yields representation of properties. One type of bell or whistle is to throw in some causal history via evolution. So a current state's representational content is that property that nomically covaried with states of its type in ancestors of this organism and because of which those states contributed to the organism's survival. It seems pretty clear that since everything covaries with itself, and states don't cause themselves, that neither nomic dependence nor causal history can serve as the mechanism of self-representation.

What's left, then, is functional role. Again, I won't pretend to know just how a state's functional role determines its content, but for our purposes the case of indexicals serves as a model. It has long been thought that what makes a term an indexical must be something special about its particular role in reasoning and planning action. So, for instance, one crucial difference between conceiving of myself in the first person, as "I" or "me," and conceiving of myself via a description that picks me out, or even my name, is that only the first-person way hooks directly into my action planning. If I think I'm about to be hit by a car, that will immediately cause me to (attempt to) get out of the way; whereas thinking that that guy is about to be hit, while pointing at what is (unbeknownst to me) my reflection in the mirror, will cause me to shout "get out of the way." Let's assume then, for the sake of argument, that there is some functional role such that playing it is sufficient for having a reflexive content of the sort claimed to be involved in phenomenal concepts.

But now here's the problem. Normally the way a functional state is physically implemented is irrelevant to the role, except that the implementer must meet the role's description. But other than that, we don't pay attention to the way the role is implemented: what's of psychological significance is the role itself. That's what functionalism is all about. That's why it's not supposed to matter whether a subject is made out of brain cells, computer chips, or Swiss cheese. Nothing changes when the functional roles involve representation. Which physical tokens are used doesn't matter, so long as the appropriate pattern of interactions among the tokens is maintained.

Of course one could grant this general point about functional roles and yet still maintain that in this case the nature of the role-implementer relation

is of psychological significance. Why? Since the role in this case is self-representation, then the identity of the implementer provides the identity of what's represented. I grant this, but now it seems as if all we have is indexicals again. The place picked out by "here," or its mental representation equivalent, and the time picked out by "now," are whatever place and time happen to be where and when the relevant tokens occur.² But again, this fact about indexicals leads to no general cognitive mystery about the nature of the present moment or location.

Why should the fact that we use a token of that very type to represent it make any difference? Why should it cause us to be unable to understand how an identity could be true when one side of the identity is represented in that way and the other side by symbols not of that type? I think the reason this sort of account has any intuitive appeal is that we aren't thinking of it in this way really; that is, we aren't thinking of a functional system that uses a reflexive device to refer to a type that that very token instantiates. Rather we're thinking of the fact that on one side of the identity is a representation that somehow involves a full-fledged conscious experience whereas the other side does not. Yes, if we're pointing at an experience—a cognitively rich and special phenomenon—and saying that *that* is identical to what is picked out by this other description, one can see perhaps why that would be hard to accept, why it would meet the sort of cognitive resistance we're trying to explain. But merely to point out that a physical state type is represented reflexively by a token of that type doesn't seem to really explain that cognitive resistance. The latter explanation, as far as I can see, borrows all its intuitive plausibility from the former. However, once we're already appealing to what is distinctive about experience in the explanation, we're taking what most needs to be explained for granted.

Let me put this another way. Here I am contemplating this visual experience of the red tomato. I'm told that having the experience is the same thing as (or to be explained by appeal to) my brain's being in a certain neural state. I find this unintelligible, mysterious. Along comes the materialist and explains my reaction this way. "You see," she says, "you are representing this state in two very different ways. On the one hand, you are describing it in terms of its physical properties. On the other hand, you are representing it by experiencing it. Your being in the state is itself your way of representing it. Therefore, since one side of the proposed identity involves your actually being in the state and the other doesn't, you can't find the idea that it's really the same thing intelligible."

But your being in the state as your way of representing it seems to make a difference precisely because being in the state is a matter of having an

experience. However, if all being an experience amounts to—all that distinguishes it for these purposes from nonexperiential mental states—is being a state that is used to represent itself, then we can't appeal to our intuitive notion of experiential character to explain the cognitive resistance to the identity claim; we can only appeal to the self-representational character. If that's all that's involved, though, it's hard to see why it gives rise to any mystery. Why should it matter whether an indexical picks out itself or something else? Why is that of such psychological moment?

6 The argument of the last section was aimed at the self-representation associated with level three. When we explicitly reflect on our conscious experience, we encounter cognitive resistance to proposed explanatory reductions and identity claims couched in physical or functional terms. It was supposed to be the self-representational character of phenomenal concepts, the form of representation employed when we introspect on our experience, that accounted for this cognitive resistance. If what I've argued above (and in Levine forthcoming) is right, then this account is inadequate. But how does this affect the claim that self-representation provides a decent account of level-two conscious awareness?

In one sense it affects it only indirectly. What I mean is this. One problem faced by the self-representation theory that is shared by every broadly functionalist theory of conscious experience is the explanatory gap/intuition of distinctness. I'm looking again at that ripe tomato on the counter (it's *very* ripe by now). I reflect on this visual experience, and consider the proposition that what it is a brain state that is representing itself. Really? Well it certainly seems as if I can coherently conceive of a device that is occupying a state that meets whatever are the appropriate functional conditions for self-representation and yet isn't having a conscious experience at all. It really doesn't seem at all like that is an explanation of what is going on with me now. Is that a problem?

Well, if one were looking to the self-representation theory itself to solve the problem of the explanatory gap, then it would be. But even if that isn't the goal, one still has to deal with it, and, if the argument of section 5 is right, one can't appeal to the special (perhaps self-representational) character of phenomenal concepts to do the job. So in this sense, the fact that self-representation theory doesn't succeed as a theory of what is special about phenomenal concepts is a problem for self-representationalism about conscious experience itself. But this problem isn't specific to self-representationalism; it affects any broadly functionalist theory, as I said.

As mentioned above, the problem just presented only indirectly affects the account of level-two awareness. Is there something in the very considerations

I used to undermine the self-representational account of phenomenal concepts that would affect the account of level-two awareness as well? I think there is, and what unites these two objections is that they both stem from the inadequacy of the self-representational account as an explanation of cognitive immediacy. Before I go into this, however, I want to consider another question, one that will help clarify some aspects of the self-representational account. The question is whether the self-representational view really does overcome some of the standard objections that have been pushed against the two-state higher-order theory.

Two objections in particular seem relevant (and are explicitly discussed in Kriegel 2003a). The first was briefly mentioned above. On the standard two-state view, conscious awareness is a relation between two states, the higher-order state that, as it were, is the consciousness of a particular mental state and the target state of which one is conscious. On this theory the higher-order state is not itself conscious.³ But this seems phenomenologically bizarre. The consciousness of the experience of seeing the ripe tomato seems as much a matter of which we are conscious as the ripe tomato itself. How can we say that the consciousness itself is not something we are aware of from within the first-person point of view?

The second objection has to do with the possibility that one might have the higher-order state in the absence of the first-order state. After all, we have hallucinations, which are representations of scenes outside us that don't really obtain. Why couldn't our internal monitoring system suffer a similar hallucination? If it did, would we be having a conscious experience? It seems that the higher-order theorist would have to say we were having a conscious experience, or at least it would be for us just as if we were. But if it is just as if we were having a conscious experience, what more do you need to call it an actual conscious experience? If one grants that this would count as an experience, however, it seems to totally undermine the relational character of the higher-order theory. All that seems to matter resides now in one state, the higher-order state.

It should be clear how adopting the one-state view at least appears to overcome these objections. With regard to the first objection, since this one state that is representing both the ripe tomato on the counter and itself is conscious, we aren't leaving a crucial piece of consciousness outside itself, as it were. All is enclosed in one state and part of our phenomenology. With regard to the second objection, the imagined kind of internal hallucination isn't possible. Since it's one state that is representing itself, if the representor is there, so is the represented.

Before we accept this appearance of overcoming the objections, however, we need to get clearer about how to understand the one-state view. There are of course a number of different ways of distinguishing versions of the view (see Kriegel this volume), but I'm interested in one way in particular. Let's call them the one-vehicle model and the two-vehicle model. On the one-vehicle model, what we have is one representational vehicle with two contents, one content directed outward and the other reflexively directed on itself. On the two-vehicle model, the one state contains two representational vehicles, one directed outward, and the other directed at the first. On the latter model, the two vehicles constitute distinct parts of a single state.

When talking about individuating vehicles (as Kriegel 2003b does), it's crucial to be clear about the level at which the individuation is taking place. If we're individuating by physical types, then probably any state that is suitable for realizing a psychological state will be one that could be broken down into parts in some way. But this fact, that the physical state in question can be seen as having parts, is not relevant to our purposes. The relevant question for distinguishing between the two models just described is this: are the two representational contents that both models posit (the outwardly directed one and the reflexive one) attributed to two separate physical states/mechanisms that constitute distinct parts of a larger physical state, or does the very same physical state/mechanism express both representational contents? So long as there isn't any principled way of associating one of the contents with one of the physical state's parts and the other with a different part, even if at the physical level the state clearly has parts, this will still count as a "seamless" one-state vehicle. So the question is, which of these two models does the advocate of self-representationalism have in mind? (Note that both one-vehicle and two-vehicle models are models of the one-state view, as the two vehicles in the latter model are characterized as parts of a single state.)

Kriegel (2003b) argues for the second model, on the grounds that it can utilize a causal theory of content. One part of a state can have a causal impact on the other, though a single state can't cause itself. In section 5 I was assuming the one-vehicle model when I argued that both nomic dependence and causal history were nonstarters for implementing self-representation. I allowed that perhaps there was going to be a functional-role account, but admittedly it's not easy to see what that would be.

Despite the fact that the two-vehicle model will have an easier time with finding ways to implement the relevant representation relation, I think the one-vehicle model better captures the spirit of self-representationalism, and better overcomes the objections that attend the two-state higher-order theory.

Take the second objection, for instance: the problem was that it seemed possible to have the higher-order state without its target state. But why can't that occur on the two-vehicle one-state model? Just because the two parts of the state count, for certain taxonomic purposes, as one state, that doesn't mean that one part of the state can't occur without the other.

Consider now the first objection to the two-state view, that the awareness itself is left out of consciousness. Well, if one part of a state is representing the other, then it doesn't really capture the way in which consciousness of the experience is inherently part of the experience itself. If A and B are distinct, B with its outward content and A referring to B, then even though we're calling the joint state A-B a single state, it still seems as if A isn't itself something of which we're consciously aware, which seems to conflict with the intuitive advantage the one-state view was supposed to have over the two-state view.

As mentioned above, state individuation is very sensitive to considerations of level of analysis. What may be one item from one perspective, is more than one from another. In his attempt to show that there can be sound empirical reasons for individuating the relevant states one way or another, Kriegel (2003b) points to the synchronization phenomena discussed by Crick and Koch (1990). He imagines two events, N1, representing the sound of a bagpipe, and N2, representing N1. So far we have a standard two-state, higher-order theory. But now suppose that the firing rates involved in the two events are appropriately synchronized, so they are "bound" to each other. Now, we have a principled basis for calling the joint state, N1-N2, a single state, and empirical vindication of the one-state (two-vehicle) view.

But what matters is not merely whether N1 and N2 are physically bound through their synchronized firing rates, but whether they are psychologically bound.⁴ That is, does the synchronization of their firing rates effect a psychological integration of the two states? In particular, does it bind them in the sense that together they express the content or contents attributable to them? The point is, so long as we think of the two representational contents as still belonging to N1 and N2, with N2 having N1 as its content, and N1 having the sound of the bagpipe as its content, I don't see a principled difference between this position and the standard two-state higher-order theory, despite the synchronization of their realizers. On the other hand, if their synchronization yields a joint state possessing two contents that are, as it were, diffused throughout N1-N2, representing both the sound of the bagpipe and the representational state itself, with neither physical part bearing one of these two contents on its own, then it does make a difference. But then we have the one-vehicle model.

Let's assume then that we are dealing with one integrated representational vehicle that possesses two contents. One problem is of course the one mentioned above, that it's not easy to see how this state that represents some outward scene comes to also possess its reflexive content. Carruthers (2000) has an account, but I have argued that it doesn't really motivate the assignment of the secondary, reflexive content (Levine 2001b). I don't want to repeat that argument here, and of course he has a reply (Carruthers 2001). Let's just leave it that finding a basis for attributing the reflexive content is an issue this theory must face.

However, lest one wonder why in section 5 I was quite content to assume some functional account or other would suffice and here I am questioning the basis for the self-representational content, let me point out that there is an important difference between level-two and level-three self-representation. At level three, the level at which phenomenal concepts operate, we are imagining a mechanism that takes a state and uses it to refer to itself. In the context of that employment of the state, it isn't being used to represent what it normally represents. Rather, we're imagining some indexical-style device that picks out that state itself. On level two, however, we're imagining that the target state retains its primary representational function of providing information concerning the immediately surrounding environment. Level-two awareness is a secondary, added-on representational content. In that situation it's harder to see what the requisite grounds for attributing the reflexive content might be.

But let's suppose that there is some functional role played by the state such that it grounds this secondary, reflexive content. I grant that the one-vehicle one-state view does overcome the two objections cited earlier to the standard two-state higher-order view, so it certainly passes that test. So why is there still an explanatory gap here? I think that though self-representation is probably part of what it is for a mental state to be conscious, the problem is that the notion of representation employed by the theory, the kind that could plausibly be implemented through some sort of functional role, isn't adequate to capture the kind of representation involved in conscious awareness. The problem, in other words, is that conscious awareness seems to be a *sui generis* form of representation, and not merely because it's reflexive. Something about the representation relation itself—that it affords acquaintance, and not just representation—is such as to yield a mystery concerning its possible physical realization. In the remainder of this section I'll try to elaborate on this idea.

I think there are two ways in which the immediacy of conscious awareness seems to elude explanation by appeal to standard, "functionalizable"

representation (by which I mean any notion of representation that is broadly functionalist, including nomic/causal relations under that rubric). The first has to do with the subjectivity of conscious experience. Subjectivity, as I described it earlier, is that feature of a mental state by virtue of which it is of significance for the subject; not merely something happening within her, but “for her.” The self-representation thesis aims to explicate that sense of significance for the subject through the fact that the state is being represented.

But now, what makes that representation itself of significance for the subject, and thus conscious? Remember, it was the felt need to endow the awareness of the visual state with consciousness itself that motivated the move from the two-state higher-order view to the one-state (and one-vehicle) view in the first place. So how is this done?

Well, it’s supposed to be taken care of by the fact that the very state that bears this content is also its representational object. But that isn’t sufficient, since we want to know how it’s being the object of its own content is itself a matter of significance for the subject. In the end, the answer for the (reductive) representationalist has to be that it’s a matter of the state’s functional role (either that or yet another representation, which of course leads to regress). That is, the content counts as something of which the subject is consciously aware because it maintains a set of access relations to other states, playing the requisite roles in planning behavior, available for verbal report, and the like. But if this is what it comes to, then couldn’t we have been satisfied with the first-order representationalist’s accessibility account of conscious awareness to begin with? That is, if a higher-order representation’s being conscious is a matter of its functional role, then why couldn’t that be the case for a first-order representation?

What distinguishes a subject for whom there is something it is like to occupy her states from one for whom there is nothing it is like is not the presence or absence of representational states; it is how those states constitute representations *for* the subjects in question. First-order representationalist theories of consciousness, as described earlier, try to make that significance for the subject out of availability to other processes. Standard two-state higher-order theories are motivated by the insight that that is inadequate, so instead seek to capture the in-the-moment significance through representation by another state. The one-state higher-order view is motivated by the insight that that too is inadequate, since the representing that constitutes the conscious awareness is left outside the conscious state. But now I’m claiming that even the one-state view is inadequate on this score. It isn’t enough to stick the higher-order representation into the state it’s a representation of.

We need for that higher-order representation itself to be of the right sort of significance *for* the subject, and merely being a representation playing its requisite functional role doesn’t seem to cut it.

I think the moral of the story is that we can’t pin the kind of significance for the subject that distinguishes a moment of conscious awareness from a mere representation by just piling on more representations, for it doesn’t yield genuine explanatory payoff. Somehow, what we have in conscious states are representations that are intrinsically of subjective significance, “animated” as it were, and I maintain that we really don’t understand how that is possible. It doesn’t seem to be a matter of more of the same—more representation of the same kind—but rather representation of a different kind altogether.⁵

I said there were two ways that conscious awareness seems to outstrip mere representation. The first had to do with significance for the subject. The second is closely linked to that. It’s a matter of subjective significance as well, though not of the awareness itself, but of what one is aware of. In a way, this is just the problem I pointed out for level-three awareness concerning the immediacy of qualitative character reappearing at level two. The way in which we are aware of what is going on with ourselves when having a visual experience, say, seems to bring us into immediate cognitive contact with it. One-state higher-order theory, the self-representation theory, supposedly accounts for that by having the representor be identical with the represented, a pretty immediate relation one has to admit. But as I mentioned in section 5, it isn’t clear why this physical immediacy should have anything to do with the kind of cognitive immediacy we’re trying to explain.

Again, consider our typical conscious experience, my seeing the ripe tomato on the counter. The way the tomato appears with respect to color is part of what it is like for me to have that visual experience. What’s supposed to capture that, there being a way it’s like for me, on the self-representational theory is the secondary reflexive content possessed by the relevant perceptual state. But how does this make the color appearance available to me? It does so either by re-representing the color within the secondary content or by demonstrating it.

On the first model, the secondary content is of the form (I am now representing that there is a red tomato in front of me) (or (there is now occurring a representing of a red tomato at such-and-such location)). On the second model, it is more like (I am now representing this), pointing at itself, as it were. Neither model seems to get at the way it is like something for me to have this experience. On the first model, it really does seem irrelevant that the secondary content is attached to the very same state as the first. We get

no more cognitive immediacy than we would get from a two-state theory. True, the secondary content is realized by the same state as the first, but again this is an implementation matter, not a psychological matter.

The second model might seem to do better. But on this model, on which it does seem to at least make a representational difference which vehicle is carrying the content, the secondary content doesn't contain a representation of the quale in question, the reddish appearance—it just points to it. In a sense, from the point of view of the secondary representation, it doesn't matter what's on the other end of the pointer. Note again the relevance of the consideration mentioned above concerning the inherent ambiguity of demonstration. To pick out the relevant phenomenal type, one must be capable of genuinely thinking about it, not merely pointing at it. We can put it this way: the relevant phenomenal type must be “in mind” in a contentful way, not merely “in the mind” to exemplify what's being demonstrated.

7 Several decades ago, when the computational model of mind was the fresh new idea, some philosophers objected on the grounds that if one posited representations in the brain, then one had to posit little men (in those days no one posited little women) for whom they were representations. Representations, so the argument went, required subjects that understood them, or they weren't really representations after all. But of course this way lies infinite regress, since understanding, being a mental process itself, must then be a computational process involving the manipulation of representations as well: little people inside little people.

The computationalist's response was to note the power of functional role to implement understanding without requiring a separate process of understanding. In this sense computational devices were provided with what were sometimes called “self-understanding” representations—representations that meant what they did by virtue of their causal role within the system, together with their causal connection to the world around them.

In a way, my complaint about the attempt to construct conscious awareness out of a functionally characterizable self-representation is similar in spirit to that original objection to computational representationalism way back when. True, the computationalist response did dispense with that objection by grounding—or, better, providing the conceptual space for eventually grounding—intentionality, or representational content, in the causal network that constituted the machine. But although this method of implementing representation seems sufficient to make the representation count as a representation for the machine, it doesn't seem up to the task of making a representation “for” a conscious subject. So we still need to know, whether it's plain representation or the fancier self-representation: what makes it for me?

What makes it the case that I am consciously aware? How, with the causal materials at hand, do we turn the light on?⁶

Notes

1. I'm indebted here to an argument of Jerry Fodor's conveyed to me informally. That argument had nothing to do with this issue, and I make no claim to be presenting his argument accurately.
2. I don't mention “I” in this connection because it's plausible that its referent is the subject of conscious experience, and so this case is infected with the question at issue.
3. Of course it would be conscious if it were the target of yet another state that represents it, but this isn't the typical case; and besides, this process of having representations of representations couldn't go on forever.
4. Of course the original “binding problem” for which this synchronization phenomenon was proposed as a solution is a matter of psychology. But it isn't clear how that notion of binding is supposed to apply to this case.
5. In conversation, Uriah Kriegel brought up the following response. Suppose the self-representation included explicit reference to the subject, so that it had the form “I am now experiencing such-and-such.” Wouldn't the presence of “I” secure that sense of being “for the subject” that we're looking for? But given what we've said already about what makes a representation an “I” representation in the first place—not merely its reference to the subject, but its role in initiating action (and the like)—it's not clear how this helps. We still have no sense of a representation being intrinsically for the subject. The “animation” is what we need to explain, and having a functionally specified “I” doesn't do it.
6. I'd like to thank audiences at New York University, the University of Arizona, and the University of Massachusetts at Amherst for helpful discussion on the topic of this essay. I'm also especially indebted to Uriah Kriegel for extensive comments on an earlier draft.