

EIGHT

Phenomenal Concepts and the Materialist Constraint

Joseph Levine

We begin with the following contrast. When considering the theoretical reduction of water to H_2O , we find the connection explanatory. We see why water is liquid at room temperature, transparent, turns to vapor under suitable circumstances, and the like, by appeal to its chemical structure and other facts expressible in the vocabulary of physics and chemistry. The connection posited between water and H_2O in no way appears arbitrary to us.

On the other hand, when we consider the theoretical reduction of a phenomenal state, such as a visual experience, something important seems to be left unexplained. Though appeal to the neurological structure of the state (together with an account of the overall physical structure of the relevant portions of the nervous system) explains a lot about how various stimuli cause the visual experience and how the visual experience interacts with other states to cause both behavior and other cognitive states, the qualitative character of the experience—what it's like to have the experience—does not seem to be explained. The connection between the neurological description and our first-person conception of what it's like seems totally arbitrary. One feels that this neurological configuration could just as easily have gone with a bluish visual experience as a reddish one. In fact, for all we can tell, it could just as easily have gone with a state that was like nothing at all for the subject.

It is this contrast—this sense of arbitrariness that attends the psychophysical reduction as opposed to the sense of intelligibility that attends other theoretical reductions—that is the core problem that goes by the name of “the explanatory gap.” Though there are many different ways to illustrate this contrast—for instance, by appeal to the conceivability argument, Frank Jackson's (1982) case of Mary, and the open question argument—it's important not to lose sight of the core idea itself, namely the contrast. It seems to me that often in the literature this core idea has been lost amid the finer details concerning questions of modality, derivability, and identity.

To illustrate what I mean, consider the issue of derivability. The question of derivability arises once we link the explanatory gap to the conceivability argument. The conceivability argument proceeds from the premise that the existence of

— -1
— 0
— +1

zombies cannot be ruled out a priori to the conclusion that they are metaphysically possible. The explanatory gap is implicated in the conceivability premise itself because, it is alleged, if we had an explanation of the phenomenal in terms of the physical, zombies wouldn't in fact be conceivable. Conceivability itself, as is clear from the statement of the argument above, has been interpreted as the absence of an a priori derivation of such statements as "This creature is having a reddish visual experience" from statements involving only physiological and functional vocabulary.

The issue of derivability has also come up in the discussion of Jackson's "knowledge argument." The punch line of the argument is that when Mary finally leaves what Perry (2001) calls the "Jackson room" and is exposed to a red surface, she exclaims, "Oh, so *that's* what it's like to see red." However, goes the argument, if phenomenal properties were really explicable in terms of their physiological correlates, she should have expected it to be like that. Based on all the theoretical knowledge she had accumulated in the Jackson room, she should have been able to predict what it would be like. This lack of ability to predict what it would be like is then interpreted to be a matter of her inability to come up with an a priori derivation of the relevant statement in "phenomenal" language from the theoretical descriptions couched in nonphenomenal language.

Once the question at issue has been transformed into one about a priori derivability, then the following dialectic takes hold. Defenders of materialism argue that in fact there is no relevant contrast between the standard cases of theoretical reduction and the psychophysical case. They argue that just as one cannot derive a priori statements of the form "This creature is having a reddish visual experience" from statements containing only nonphenomenal vocabulary, so too one cannot derive a priori statements such as "This cup is filled with water" from statements containing only vocabulary from chemistry and physics. If this absence of a priori derivability makes zombies conceivable, then so too is "zombie-H₂O" conceivable (i.e., H₂O that isn't water). Just as the conceivability of the latter does not throw any doubt on the metaphysical claim that water is H₂O—nor does it undermine the felt intelligibility and nonarbitrariness of the connection—so too the mere conceivability of zombies should throw no doubt on the claim that phenomenal properties are physical properties—or, more important, on the claim that the psychophysical link is fully intelligible.

Consider now the issue of identity. In presenting the core contrast underlying the claim that there is an explanatory gap between the phenomenal and the physical, I pointed to how we can explain many of the properties of water—such as its liquidity at room temperature—by appeal to its chemical structure (along with other relevant chemical facts, of course). I then noted that we seem unable to explain a phenomenal property, like the reddishness of a visual experience, by appeal to its underlying neurological structure. But of course there is an important difference here. In the water case, we look at some of water's properties and then seek an explanation in terms of the properties of H₂O. But in the psychophysical case, the explanandum at issue is the instantiation of the property that is alleged to be the very same property as that cited in the explanans: phenomenal reddishness, according to the materialist, *just is* the relevant neurological property. So of course there is no explanation, since identities are not really susceptible of explanation.

Just ask not why water is liquid at room temperature, but why water is H_2O , to see how misleading the initial presentation of the core contrast was. After all, what is there to say in answer to why water is H_2O but that that's just the way it is? A similar answer, it is alleged, is appropriate in the psychophysical case.

Though these two materialist replies that focus on the issues of derivability and identity contain important insights, I think they still miss the crucial point, which is the core contrast with which we began. Suppose I agree that in fact one cannot derive a priori "water" talk from chemical talk precisely because the difference in vocabulary blocks the derivation. Suppose I also agree that identity claims, strictly speaking, do not require explanations. Still, the following contrast continues to stare us in the face. If after learning all the relevant chemistry—after learning how the molecular structure H_2O is responsible for all the superficial properties by which we identify water—someone were nevertheless to assert that she still didn't see how water could be H_2O , that the connection between being H_2O and being water seemed quite arbitrary, I believe we literally wouldn't understand what she was talking about. Yet, after learning all that Mary supposedly learns in the Jackson room, we understand quite well the feeling Mary has when she says, "Oh, so *that's* what it's like to see red!" There is a clear sense of arbitrariness about it, a sense that it could easily have been some other way. If she were to follow up her exclamation with the question, "But why should it be like *that*?" we'd know what she means. The fact that we lack a priori derivability in both cases doesn't remove the fundamental epistemic or cognitive contrast.¹

One may reply here that by insisting on this core contrast, I am refusing to acknowledge the crucial point just mentioned concerning the inexplicability of identities. A cognitive sense of "arbitrariness" is a symptom of having hit explanatory bedrock. When confronted with a question such as why the fundamental laws of physics are as they are—why, say, light travels the speed it does—it's acceptable to answer that that's just the way the world is. Yes, it is arbitrary, because there is no more fundamental phenomenon by appeal to which we can explain it. In effect, this is the property dualist answer to the explanatory gap. Why, the dualist says, does it seem arbitrary that this phenomenal property is correlated with that physical property? Her answer is that this reflects a basic law of nature. Appeal to the basic nature of a connection is a satisfactory explanation of its apparent arbitrariness.

Of course the materialist can't appeal to a basic law connecting phenomenal and physical properties, for that would be to accept a form of dualism. But it appears she can still co-opt the "That's just the way it is" response by appealing to the basic

1. Some may question the claim that there is a fundamental contrast here. After all, given our ignorance with regard to the physical and functional basis of perceptual experience (we know a lot, but not what Mary supposedly knows), who knows how someone in Mary's position would react? Maybe our sense that she would say, "Oh, *that's* what it's like." is a projection from our relatively ignorant state. Well, I will just record here that I don't find this move compelling. Of course that isn't an argument, but my current project is not really aimed at responding to this challenge. Rather, I'm interested in addressing a materialist who grants this basic point about what Mary would say but claims it in no way threatens her doctrine.

or brute nature of identity. When we reach identity claims, it is argued, we have also reached explanatory bedrock. Now I grant this, to an extent, but it doesn't address the problem presented by the core contrast. After all, in both the water-H₂O and psychophysical cases we are dealing with identity claims, yet our cognitive assessment of the arbitrariness of the links is quite different. We need an account of this difference.

Perhaps we can put the matter this way. We begin by identifying certain phenomenal states, such as visual experiences, with certain neurological states. We find that by doing so we can explain many of the properties of the phenomenal state, as mentioned above. We can explain, basically, its causal role. We then seek an explanation of its qualitative or phenomenal character. Why, we ask, is it like this to see red? Appeal to neurological properties doesn't seem to answer the question. The neurological properties seem apt only for explaining causes and effects. But then the materialist makes the next move. You see, she says, the qualitative character *just is* one of the neurological properties. At this point, we intuitively balk. How can that be, we retort? They certainly don't *seem* to be the same property. The idea seems unintelligible, in a way that the identity of water with H₂O doesn't. So now we face, rather than the explanatory question, the non-identity question. Why, in this case, does it seem so bizarre to consider what is picked out by the one vocabulary to be the very same thing as what is picked out by the other vocabulary, when no such sense of bizarreness attends other theoretical reductions? Whether we think of it as an explanatory gap or a distinctness gap, the problem is really the same.

Before turning finally to a discussion of phenomenal concepts, I want to pick up one loose thread from above. I noted earlier that the explanatory gap has been connected to the conceivability argument, to the knowledge argument, and also to the open question argument. I think a consideration of this last link will reinforce my argument concerning the relevance of the core contrast. By the "open question" argument, I have in mind the following. Suppose we are confronted with an alien species or an advanced robot. We know everything there is to know about its internal workings. It turns out that its functional organization is, down to a fairly low level of implementation, very much like our own, though the physical mechanisms are different. Now we ask, is it conscious? And, if so, is what it is like for this creature to see red the same as what it is like for us?

Contrast this case with the famous case of Twin Earth. On Twin Earth, we find a substance that behaves, at the macro level, exactly the way water behaves on Earth. We are happy to consider it water. However, after we conduct chemical tests on it, we see that its underlying structure, XYZ, is quite different from H₂O. Is it water nevertheless? Of course, those of us growing up in philosophy after the Kripke-Putnam revolution are prone to respond immediately that it isn't water. On the other hand, one can see how one might make the argument that it is water, that there are two kinds of water. Some might compromise and claim that it's really not determined in advance, that we just have to decide how we want to use the term "water" and whether or not we want to include XYZ in its extension.

Any of these three positions makes sense, I contend. We might have good reasons for picking one above the others as a better semantic hypothesis, but one can imagine arguments for all three. The point is, what all three positions have in

common is the view that, after all the chemistry is done, what is left is a semantic issue. There is no real question of nonsemantic fact left open. All the relevant nonsemantic facts are revealed. This situation stands in stark contrast to how the psychophysical case described above at least appears. Whether or not the newly encountered creatures possess consciousness, and, if so, what it's like for them, does not seem at all a semantic question. What is left open, it seems quite clear, is a matter of genuine, nonsemantic fact, one that we haven't a clue how to go about determining. After all, what would it mean to "just decide" to call these creatures "conscious" or to call their phenomenal character when looking at red "reddish"?

I think the presence of an open nonsemantic question in the psychophysical case and its absence in the water case beautifully captures the core contrast. That there seems to be a genuine open question reflects the strong intuition that we are dealing with two properties here, the phenomenal one and the physical one, and that there is at best a brute, arbitrary connection between them. Though I don't think the falsity of the materialist identity claim is *entailed* by this intuition of distinctness, I do believe the materialist has a burden to explain its persistence. This is precisely what some materialists have attempted recently by appealing to the special properties of what are called "phenomenal concepts."

Most materialist attempts to explain the existence of an explanatory gap revolve around the idea of a phenomenal concept. Phenomenal concepts are concepts of phenomenal properties, the ones employed in standard first-person thoughts about one's conscious experience. So, for instance, when I wonder how the reddishness of my visual experience is explained by appeal to its physical properties, the concept of reddishness that is a constituent of that thought is a phenomenal concept. Though other concepts—those, say, that are expressed by vocabulary from neuroscience or computational psychology—might pick out the phenomenal property of reddishness, they don't qualify as phenomenal concepts. Phenomenal concepts are quite special ways of conceiving of our sensory experiences, proprietary to the first-person point of view.

The general materialist strategy then is this. The initial puzzle concerns a certain cognitive state, our state of wondering how the physical properties of our sensory states explain their phenomenal properties. That we have such cognitive states is *prima facie* puzzling because the phenomenal properties in question just are the physical properties in question, so what's to explain? Obviously, they *seem* to be different properties, and the explanatory question makes sense to us. The answer to the puzzle presented by these cognitive states is to note that they involve, as constituents, two radically different kinds of concept: phenomenal concepts and, for want of a better term, nonphenomenal concepts. It is because we are conceiving of phenomenal properties via these two different kinds of concept that the explanatory question makes sense, that they seem so strongly to be about distinct properties. But the distinctness is all in the concepts, not the properties the concepts are concepts of.

Notice that the strategy just outlined does not merely appeal to the distinctness of the concepts involved, but to the alleged radical difference in kind between phenomenal and nonphenomenal concepts. This is crucial because we are able to see,

with no special cognitive difficulty, how numerous properties and individuals that are picked out by distinct concepts might be the same thing. We started, after all, by noting such contrasts. So the critical move on behalf of the materialist is to provide an account of phenomenal concepts—and what's involved when we bring them together with nonphenomenal concepts within the same thoughts—that explains the unique cognitive features of this case. The attempts to provide just such an account are what I want to investigate in this chapter.

In the following sections, I will look at various ways of characterizing phenomenal concepts that are supposed to explain the presence of an explanatory gap. Though I base these characterizations loosely on the literature, none of them precisely corresponds to any one particular theorist, and some may not correspond to any. My purpose is to systematically survey the options for explaining the explanatory gap by appeal to the peculiar features of phenomenal concepts, not to take issue with any one such account.²

Before beginning this survey of the options, I want to emphasize that any materialist proposal for explaining the gap must meet a condition I will call the Materialist Constraint: namely, that no appeal be made in the explanation to any mental property or relation that is basic. For instance, suppose a materialist argued that she could explain why there should be an explanatory gap (even though phenomenal properties were constituted by physical-functional properties) by appeal to some basic mental relation like acquaintance that held between subjects of experience and their brain states when conceiving of those states via phenomenal concepts. Acquaintance itself is not given a materialist explanation, but appealing to it, let us say, removes the mystery of the gap with respect to phenomenal properties. How appeal to acquaintance might do the job is not important for now (I will return to this later). The point is that it does the materialist no good to explain away the gap by violating her own doctrine—that is, by admitting into her ontology a mental relation that is basic. Thus, in our examination of the various proposals to follow, it will be crucial to note that violation of the Materialist Constraint immediately disqualifies a proposal because it ceases to be a *materialist* explanation of the gap. Whatever it is that makes phenomenal concepts special, it must be possible to see how this feature can be implemented in a physical system by physical mechanisms.

I begin with perhaps the simplest idea. Phenomenal concepts are representational primitives, and therefore thoughts containing them cannot be derived from thoughts that do not contain them. The mere difference in mental “vocabulary” between phenomenal language and nonphenomenal language explains why no purported explanation containing only nonphenomenal language in the explanans and phenomenal language in the explanandum can succeed.

This account is based on two principal ideas: that the explanatory gap is primarily a derivability gap, and that most ordinary lexicalized concepts possess

2. My discussion is based on the following works: Balog 2002; Block, chap. 12, this volume; Hill and McLaughlin 1999; Levin, chap. 6, this volume; Loar 1990/97; Papineau 2002; Perry 2001; Sturgeon 2000, chap. 2; and Tye 2000, chap. 2.

analyses that in principle allow their elimination. As can be seen from the discussion above, both ideas are necessary to make this move work. I noted that one common reply to the conceivability argument is to maintain that metaphysical necessity can exist even in the absence of a priori derivability, and to point to the standard theoretical reductions as evidence of this. My counter-reply was to emphasize that we still need to explain the core contrast between the psychophysical case and the standard cases. Clearly, absence of derivability can't be the distinguishing factor, since we lack upward derivability in both cases. For these materialists, then, appeal to the fact that phenomenal concepts are representational primitives doesn't help, since it doesn't distinguish them from many nonphenomenal concepts.

However, I am now imagining a materialist who agrees that in general, most macro-level descriptions can be derived from the relevant micro-level descriptions. We do have upward derivability in the case of water, heat, and all the other standard cases of theoretical reduction. By admitting that upward derivability is normally present when there is upward metaphysical necessitation, the materialist can pin the core contrast between the psychophysical case and the standard cases on the presence or absence of upward derivability. Then, contrary to the antimaterialist's insistence that this lack of derivability reflects a metaphysical distinction in properties, the materialist maintains that it only reflects the primitive nature of phenomenal concepts. They *seem* to be about distinct properties, but that is only because of the distinctive character of the concepts by which we conceive them.

I present this option only to set it aside for now. I have two reasons. First, this position seems quite implausible on its face, and few materialists would endorse it. One has to buy the idea that only phenomenal concepts (along with logical, mathematical, and indexical concepts, and perhaps the concept of causation) are primitives, and that all of our other concepts are ultimately definable in terms of them. This sure sounds like the discredited doctrine of phenomenalism. Second, even if one did bite this bullet, I think there are independent reasons for thinking that mere lack of derivability is not the principal issue here. But to make that case we need to survey some of the other options first.

The first move was to account for the special character of phenomenal concepts by appealing to their status as conceptual primitives. This didn't work because many nonphenomenal concepts are primitive as well. The next move also involves imputing a kind of primitiveness to phenomenal concepts, but this time it is not conceptual or representational but rather epistemic or judgmental. The idea is this. For most concepts, when we apply them in judgment—for example, judging that there's a dog in front of me, there's water in that glass, and so on—our application of the concept to an object depends on the application of other concepts. These other judgments serve as evidence for the judgment in question. So I judge that there's a dog in front of me because it appears to me that there's a dog in front of me; similarly for the judgment that there's water in that glass. Of course I needn't consciously go through this inference. But if someone were to challenge my claim about the dog or the water, my justification would certainly involve mentioning how things visually appear to me.

Now, when it comes to phenomenal judgments—say, I’m having a reddish visual experience, I’m having a headache—there don’t seem to be any epistemic liaisons of this sort to serve as evidence. I judge that I’m having a reddish visual experience because I am; the same for the headache. After all, what else could I point to? In these cases, the phenomenal states themselves seem directly to give rise to the judgments with no evidentiary intermediate.³ Notice that the phenomenal states do not themselves serve as evidence, since evidence already presupposes conceptualization. Rather, it’s just that we are set up, when things work normally, to (be in a cognitive position to) judge that we are having a certain experience whenever we are.

Let’s suppose that epistemic primitiveness really is a distinctive feature of phenomenal concepts. We still need to know why the fact that phenomenal concepts are applied in this epistemically primitive manner should give rise to the explanatory gap. How is that account supposed to go?

Perhaps the idea is this. The fact that most ordinary concepts possess epistemic liaisons of a type that phenomenal concepts lack helps to explain why accounts of underlying mechanisms in these cases yield satisfying explanations. For instance, when I identify something as water, I do this on the basis of its taste, visual appearance, feel, and the like. The chemical account of water not only provides a candidate for water’s identity but also links that candidate to the mechanisms responsible for these other features by means of which I normally identify water. By thus embedding the concept of H₂O in a story that connects not just to the concept of water but also to all the other concepts of its epistemic liaisons, the sense that we have a genuinely explanatory account of water is generated.

But if this is how the appeal to epistemic primitiveness is supposed to work, it doesn’t in fact succeed. Though it may be true that phenomenal concepts can be applied in judgment without the application of other concepts in an evidentiary manner, this doesn’t mean that phenomenal concepts are bereft of relevant links to other concepts. In particular, an account of the mechanisms underlying the production and interaction of phenomenal states would link up with many other concepts, especially nonphenomenal ones. We have a rich body of beliefs concerning the causes and effects of phenomenal states—composed of both phenomenal and nonphenomenal concepts in the very same cognitive states—and they constitute a ready set of explananda for theoretical accounts of phenomenal experience. Phenomenal concepts are well integrated with other concepts, as the very facts mentioned in the paragraph above attest. After all, what are these other concepts that are connected to our application of the concept of water in judgment if not phenomenal concepts concerning the way water affects our sensory experience? So the mere fact that we are built to apply these representations without evidential intermediaries, and that they are activated only after the instantiation of the relevant phenomenal states, doesn’t explain why thinking of phenomenal states by way of

3. This doesn’t mean the judgments are incorrigible—though they may be. It’s just that I do not normally rely on any other judgments as evidence in order to make them. Whether they could still turn out to be false is another question.

them should make these states seem so arbitrarily connected to their neurological correlates.

I think there is a lesson to learn from the inadequacy of these first two accounts. What we're trying to explain is this. When we entertain these two different concepts of what is supposed to be the same property, we can't resist thinking of them as picking out distinct properties. We were trying to explain this phenomenon by appealing to a kind of cognitive isolation distinctive of one of the two concepts, the phenomenal one. But we saw that phenomenal concepts maintained links to nonphenomenal concepts, and though they may be primitive and unanalyzable, that didn't distinguish them from many nonphenomenal concepts. So it seems as if cognitive isolation isn't really the issue.

Perhaps the answer lies in the relation between phenomenal concepts and their objects, phenomenal properties, rather than in the relation between phenomenal and nonphenomenal concepts. That is, maybe there's something special in the way phenomenal concepts represent phenomenal properties that accounts for this strong sense we have that when entertaining both a phenomenal and a nonphenomenal concept, we are dealing with distinct phenomena. Of course any model of that special relation between phenomenal concept and phenomenal property must respect the Materialist Constraint if it is going to defend materialism against the challenge of the explanatory gap.

Further support for the idea that what's special about phenomenal concepts is the way they relate to their objects comes from another significant feature of phenomenal concepts that calls out for explanation: the fact that only subjects who have actually experienced the relevant phenomenal properties are capable of possessing the corresponding phenomenal concepts. For example, one of the main ideas underlying the Mary case is that one cannot employ a phenomenal concept unless one has personally instantiated the property it is a concept of. The reason Mary supposedly learns something new, is able only upon leaving the Jackson room to judge "So *that's* what it's like to see red," is that one can't employ, or even possess, phenomenal concepts until one has experienced the corresponding phenomenal states. Whereas one doesn't need to have instantiated constricting blood vessels in order to have a concept of them (the relevant nonphenomenal concept, that is), one does need to have experienced a headache to have that special first-person phenomenal concept of a headache. Let us refer to this feature of phenomenal concepts as the etiological constraint. Clearly there is something special in the way phenomenal concepts and their corresponding properties are related that explains the etiological constraint, and it stands to reason that whatever this special feature is, it also explains why there is an explanatory gap as well. In the next section we'll explore one way of trying to capture this special nature of the phenomenal concept-phenomenal property relation.

The model to be explored in this section incorporates the features of conceptual and epistemic primitiveness, but it adds a crucial element: phenomenal concepts are taken to involve an essential demonstrative component. I am going to use John Perry's (2001) discussion of Jackson's knowledge argument as my target here. He doesn't explicitly address the issue of the explanatory gap, but, as discussed earlier, the same

issues involved in the problem of the explanatory gap come out in the knowledge argument. When necessary I'll make explicit the connection between the two.

Before presenting Perry's response to the knowledge argument, and in view of our discussion of the etiological constraint above, there is one preliminary point worth making. When considering Mary's new knowledge, it might be tempting to try to dismiss the problem immediately as follows. Look, one might say, of course Mary couldn't predict what it would be like to see red from what she knew of the physics and physiology of vision while in the Jackson room. After all, in order to possess the relevant concept—the phenomenal concept—with which to frame the relevant judgment concerning what it is like, one must first have the experience of seeing red oneself, and Mary hadn't done that yet. Her new knowledge upon emerging from the room is merely a matter of acquiring a new concept.

Though in fact the accounts we are going to investigate incorporate some aspects of this response, it's important to see from the outset that saying this much alone is clearly not sufficient. Nida-Rümelin (1995) has made the case quite convincingly. She asks us to consider Marianna, who starts out just like Mary, but instead of seeing a ripe tomato upon release, is ushered into a room with many abstract colored shapes hung on the walls, with no labels to say which color is which. She is asked if she can tell which of these is the color of the sky, of ripe tomatoes, and so on. It seems pretty clear that Marianna would not be in a position to know. When she is told that this one is red, she now learns what it is like to see red (or see ripe tomatoes). The knowledge seems new, despite the fact that she already possesses the first-person phenomenal concept of what it is like. Thus her inability to predict what it would be like, or to explain what it is like, given the conceptual resources available to her in the room is not merely a matter of her lacking the relevant experience. Coming to have the experience alone, and thereby acquiring the phenomenal concept, is not sufficient. The right sort of connection must also be made, and it is her inability to automatically establish that cognitive link between her previously acquired physical concepts and her newly acquired phenomenal concept that must be explained.

This is where the demonstrative account comes in. Perry asks us to consider the following situation. He has long admired Fred Dretske's work and knows it fairly well. In particular, he knows that Dretske is the author of *Knowledge and the Flow of Information*. However, he has never met Dretske and doesn't know what he looks like. At a party he meets a man he doesn't recognize and falls into a conversation about the topic of knowledge and information. He suggests to his conversational partner that he read *Knowledge and the Flow of Information*, proceeding then to present the main argument of the book. Much to his embarrassment, the man whom he's lecturing on the book informs him that he in fact wrote *Knowledge and the Flow of Information*.

So let's consider the following two statements:

1. Dretske wrote *Knowledge and the Flow of Information*.
2. This man wrote *Knowledge and the Flow of Information*.

Perry clearly knew (1) before meeting Dretske at the party. But (2) seems to be something he learned only after Dretske informed him of it in the middle of their

conversation. Before that moment, while conversing with Dretske (but not knowing his name), Perry presumably would have doubted (2) (otherwise Perry wouldn't have presumed to recommend that he read the book). So (2) seems to be a bit of new knowledge. Yet if we take "this man" to directly refer to Dretske, statements (1) and (2) express the same proposition, describe the same fact. Furthermore, though Perry doesn't himself say this, there doesn't seem to be any explanatory gap here. That is, we don't find ourselves puzzling about how this man could be Dretske.

The moral for the case of Mary is supposed to be straightforward. Let Q_r stand for Mary's concept of the qualitative character the average human being experiences when seeing a ripe tomato in normal light. In the Jackson room, where Mary learns all the relevant physical and functional facts about visual experience, Mary learns in particular that

3. When Sally (a normal perceiver) sees a ripe tomato (in normal light), she occupies a state of type Q_r .

After emerging from the Jackson room, while herself looking at a ripe tomato, Mary now comes to believe (she knows enough about her own brain and visual system to know that she herself is a normal perceiver):

4. When Sally sees a ripe tomato she occupies a state with *this* qualitative character (demonstrating the qualitative character of her own state).

Since, according to the materialist, the qualitative character she's picking out with her demonstrative is in fact Q_r , (3) and (4) express the same proposition, describe the same fact. The problem with saying this is supposed to be that (4) seems to be a new belief, a new piece of knowledge. However, just as the epistemic novelty of (2) in no way impugns the identity of this man with Dretske, so too, argues Perry, the epistemic novelty of (4) in no way impugns the identity of this qualitative character with Q_r . Furthermore, just as we have no problem understanding how this man could be Dretske, we should have no problem understanding how this qualitative character should be Q_r .

Perry's analysis of what's going on in the party situation is instructive. He asks us to imagine that the mind is like a three-story building. On the first story are perceptual buffers, where files are opened for objects currently being perceived. Various features concerning the look of these objects are stored in these files. On the third floor are what he calls "detached" files, relatively permanent files for various objects that are not attached to any current perceptions. They store all sorts of information from all sorts of sources, including memories of past perceptual encounters if there were any and information gleaned from reading and reports from others. The second story is dedicated to connecting files from stories one and three. Sockets hang down from story three, and plugs lead up from story one. When an object is recognized or when one learns the identity of an object currently perceived for the first time, the plug from the first-story file connects to the socket from the appropriate third-story file. So, in the party case, when Dretske told Perry who he was, the plug from the perceptual buffer file dubbed "this man" connected to the socket hanging down from the upper-story file named "Dretske," and this

allowed information to flow freely back and forth between the two files. Both files, of course, were about, or “of,” the same man all along.

As an armchair first approximation to what goes on in perceptual identification and recognition, this story seems pretty good. Let’s suppose it’s right. So when Perry finally learns that he’s talking to Dretske himself—when the plug enters the socket—some genuinely new information does become available to Perry: namely, what Dretske looks like. “This man,” directly referential though it may be, does bring along with it the information that the object referred to is currently being perceived to have various sensible properties. That Dretske is currently being perceived by Perry to have these visual properties is genuinely new information, information not contained in the detached “Dretske” file he maintained up until now in his third-story file drawer. Though we do identify this man with Dretske, we don’t identify any of the properties represented in the perceptual buffer file with any of the properties represented in the previously detached third-story file. Thus the sense that (2) expresses in some way a new piece of knowledge is easily explained by the fact that Perry now knows that the object presenting certain visual properties to him is the same object that wrote *Knowledge and the Flow of Information*. The object is the same, but not the properties associated with the two ways of picking it out.

But now, if we apply this model to Mary’s case, the original problem reemerges. The idea is supposed to be that in formulating the thought expressed by (4), Mary is demonstrating her experience in much the same way that Perry is demonstrating Dretske in (2). Presumably this is not supposed to be perception in the normal sense, but it is supposed to bear a strong resemblance to it. In particular, we are supposed to think of the new knowledge as a matter of linking plug and socket, as in the Dretske case. But there is a problem with this. Remember that *Qr* stands for Mary’s previously detached concept of the relevant qualitative character. The file contains all sorts of neurophysiological, computational, and optical information. “This qualitative character” stands for the buffer file. But what does it contain? Well, it surely seems to contain some substantive idea of what it’s like. The problem, then, is this. If we push the analogy with the Dretske case, then we can see easily enough how the two files can be “of” the same object—or, in this case, state. But an integral part of the Dretske case was that the information contained *within* the two files was different. If we maintain that aspect of the analogy in the Mary case, however, we end up reintroducing a new property, and we’re back to where we started.

Well, maybe we’re not supposed to take the analogy that literally. Maybe the feature of the Dretske case that involved the file folder containing perceptually derived information was inessential to the story. Perhaps we should think of the “buffer” in the Mary case as containing merely generic, topic-neutral information. That is, her first-person phenomenal representation has something like the form of “this sensory state,” or “this qualitative character,” so that the specific character is not itself represented in the file folder itself. In this way it would be quite different from the sort of file we’re imagining in the standard perceptual buffer case. Or why not go further and think of the relevant file folder as just empty? Perhaps now we have a model of what’s going on with Mary that explains her new “knowledge” in a benign way.

A moment's reflection, however, reveals that the "empty file folder" idea won't work, for once we drop the assumption that the buffer file folder contains at least minimal information, we lose any grip on what's going on. Demonstratives, after all, are almost always associated with perceptual contents of some sort. I pick out something I see, hear, or touch with the expression, or thought, "*that F*" (where *F* represents some sortal or other), or even just "*that*." Even in the case where no sortal is employed, it's still clear that the demonstration is anchored to some perceptual representation. The whole idea of buffers containing files really only makes sense when there's something to put in the files. So although I can perfectly well make sense of the idea of demonstrating my current experience, we still need a model of what fills the role of the content material that anchors the demonstration: what's inside the buffer's file folder. The demonstrative itself cannot do all the work that's required here.

Demonstratives, then, seem to require some associated content material in order to lock onto their objects (whether they be objects or properties, particulars or universals). But there's still the option that that content material is quite thin and generic, just a matter of providing a sortal like "type of qualitative character" or "type of sensory state." So the idea is this. When Mary entertains (4), she is identifying the same property, or type of state, that is picked out by her rich scientific description with the one picked out by her demonstration of "this state," where the only substantive characterization available in the demonstrative file is that it is a state she is in. No qualitative or descriptive content is associated with this way of picking it out.

Returning to the Dretske case, the better analogy would be the following scenario. I've met Fred Dretske before, know perfectly well what he looks like, sounds like, and so on. However, it turns out that he has a twin brother who looks exactly like him, talks like him, and holds the same views on knowledge and information. In fact, his name is "Fred" as well. Of course, Fred₂ (as I'll call him, not what he calls himself) didn't write *Knowledge and the Flow of Information*. So when speaking with someone whom I know to be one of the Freds at a party, I might wonder which one it is. Then, after Fred tells me he wrote *Knowledge and the Flow of Information*, I will have learned something new, which I can express as (2).

This situation is now quite similar to the one we're imagining Mary is in when she expresses (4). It's not quite the same, since, as in Perry's scenario, it is still the case that the perceptual buffer's file folder contains quite a bit of information concerning the look and sound of the man I'm speaking to. However, because of my previously established familiarity with Fred, none of this information (or, at least, none of it that remains with me, or achieves salience) contained within the buffer's file is new. The file's plug connecting to the socket from above achieves no information flow over and above the mere fact that *this* is the one that *that* upper-story file is about (as opposed to the one I have under the name "Fred₂").

However, I think it's clear that this modified Dretske scenario is not a good model for what's going on in the Mary case, precisely because, on this scenario, no real new information is introduced via the demonstrative presentation. Mary doesn't seem to learn just that the state she can describe in such rich theoretical vocabulary is happening here and now; that it's *this* one. She forms a new

conception, one with substantive and determinate content, of this state. The situation is much more like Perry's original Dretske case, where he learns what the author of *Knowledge and the Flow of Information* looks like. That is, in the Mary case, just as in Perry's version of the Dretske case, the new information is not constituted only by the mere linking of buffer file with detached file, but also by the fact that there is something new in that buffer file itself: a new way of representing this particular type of qualitative character. It is to this new representation, contained within the buffer file, that we must look for the phenomenal concept we're after.

Before turning to our next option, I want to return to a thread I earlier left hanging. Recall that I described a materialist who accepted the view that most ordinary nonphenomenal concepts were analyzable in such a way as to permit upward derivations from the relevant microlevel descriptions, but who maintained that phenomenal concepts were primitives and therefore were not subject to upward derivations. On this view, it was the presence of upward derivability in the standard nonphenomenal cases and the absence of it in the phenomenal case that explained the core contrast. However, our investigation of the demonstrative model shows that mere absence of upward derivability cannot be the whole story. Descriptions that essentially involve indexicals or demonstratives cannot be derived from descriptions lacking them. Yet, as we see from the modified Dretske case, no sense of substantively new information, or a distinct property, is automatically engendered. There must be something about the way that phenomenal concepts afford a grasp of their objects—afford a kind of “cognitive presence” to phenomenal properties—that explains why they seem distinct from anything conceived by another method. Providing a model of this relation, of what this cognitive presence amounts to, that accords with the Materialist Constraint is the challenge.

I used the term “cognitive presence” to try to express the unique relation that phenomenal concepts seem to bear to what they represent, but I might just as easily have used the traditional Russellian term “acquaintance.” Russell (1912, chap. 5) divided the objects of knowledge and thought into those that were known by acquaintance and those that were known by description. Among the objects knowable by acquaintance were the immediate contents of sensory experience. Most ordinary objects were known only by description, where the descriptions in question contained logical terms and those representing items with which we were acquainted. Thus for Russell, all epistemic access and reference bottomed out with acquaintance.

Let's put aside the foundationalist epistemology and theory of reference, allowing that terms representing items with which we are not acquainted might pick out their objects, and afford epistemic access to them, without employing modes of presentation that are ultimately constructed from objects with which we are acquainted. One might still find the distinction between items (whether they be objects or properties) with which we are acquainted and those with which we are not acquainted useful; though, if we are eschewing Russell's foundationalism, the latter category shouldn't be characterized as those items we know “by description.” Let's allow, instead, that there are two forms of direct reference: one involving acquaintance and one not.

The essential idea behind Russell's distinction still remains. When it comes to the properties of our immediate experience, we stand in a kind of epistemic relation to them that is more intimate, more substantive, than the kind of relation that obtains between our minds and other items. The properties of experience are, to use my other phrase, cognitively present to us. The idea, then, is to explain why it so strongly seems to be the case that what is presented by way of phenomenal concepts is distinct from what is presented by nonphenomenal concepts by appeal to the distinction between acquaintance and other forms of representation. Because phenomenal concepts afford acquaintance, whereas nonphenomenal concepts do not, even if they in fact pick out the very same properties, we find it cognitively difficult to see how this can be.

Perhaps we can put it this way. Nonphenomenal concepts either pick out their objects by way of substantive modes of presentation constructed out of other concepts, or by direct labeling that involves no substantive mode of presentation at all—merely, say, a causal relation of some sort between the concept and what it's a concept of. By contrast, phenomenal concepts employ a substantive mode of presentation and use what they are about—phenomenal properties—as the modes of presentation.⁴ Since phenomenal properties are themselves involved in the very mode of presentation when conceived via phenomenal concepts, but not when conceived via nonphenomenal concepts, the referents of the two sorts of concepts will present themselves to us as distinct, even though they are identical.

I realize that the notions of acquaintance, of cognitive presence, and of phenomenal properties being their own modes of presentation are all still too metaphorical. But let's suppose we have enough of a handle on them to proceed. What we're looking for is some model of this special cognitive relation that supposedly obtains between a phenomenal concept and its corresponding phenomenal property that simultaneously satisfies the Materialist Constraint. This means that whatever acquaintance is, it can't be a basic relation; it must be constructible out of other, nonmental relations.

In this section, I want to investigate the possibilities for essentially providing a materialist model of acquaintance. There are two approaches I want to look at, and they share a crucial feature: both try to incorporate an instance of the phenomenal property into the phenomenal concept itself. In this way, they attempt to capture the idea that the phenomenal property (or an instance of it) serves as its own mode of presentation, which we can for now take to be the essential element in the relation of acquaintance.

The first approach is really a modification of the demonstrative model. In standard cases of demonstrative representations, there are three elements in play: the demonstrative component itself, a perceptual representation, and the object demonstrated. Notice that the demonstrative representation, or concept, itself contains only the first two elements mentioned. The object is not itself part of the

4. Loar has explicitly mentioned this idea of phenomenal properties being their own modes of presentation.

representation; it is merely what is represented. So, for instance, in Perry's Dretske case, his thought "that man" picks out Dretske by way of demonstrating whatever it is that presents the particular appearance he is currently perceptually aware of. In terms of his file model, the perceptual information is contained in the file, while the file itself plays the role of the demonstrative element. "That man" refers to whoever this perceptual information is about.

But suppose that, instead of treating the demonstrative representation as composed of the first two elements, which together pick out the third, we collapse the second two elements into one. That is, the demonstrative representation in this case is taken to consist of the demonstrative element together with what is demonstrated, the phenomenal state (or, to be more precise, the property it is currently instantiating). If we like, in keeping with Perry's metaphorical architecture, we can picture this as putting the phenomenal state itself into the demonstrative file folder. A phenomenal concept, then, is a complex state consisting of a demonstrative together with the state demonstrated, interpreted to represent the relevant property instantiated by the demonstrated state. To be acquainted with a property is to demonstrate an instance of it in a state that includes the instance as a component of the representation.

The second approach is to forget about the demonstrative element altogether, and just let tokens of phenomenal states themselves serve as representations of the phenomenal properties they instantiate. In other words, phenomenal concepts are tokened by the very same states that serve to instantiate the properties they represent. Again, this is a very graphic way to implement the notion that a phenomenal property serves as its own mode of presentation.

As a first step in evaluating these two approaches, I want to argue that in fact there really is no significant difference between them; they amount to the same model in the end. The argument takes the form of showing how each version presupposes, or incorporates, the other one. Let's start with how the demonstrative version incorporates the self-representation version. Notice that the idea of putting the phenomenal state itself—the demonstratum (or the current instantiation of the demonstratum)—into the demonstrative file itself, as playing the role of mode of presentation, is really quite odd. In the standard file story, there are two ways of looking at the contents of the file. For instance, in Perry's original Dretske case, the demonstrative buffer file contains the perceptual information concerning the appearance of the man with whom Perry is speaking. This perceptual information, however, can be thought of as a physical token of some sort—imagine that it is literally a picture stored in a file—or it can be thought of as a representation of an appearance. Clearly in Perry's story about meeting Dretske, it is the second way of thinking about the contents of the file that is relevant. The intrinsic features of the vehicle are irrelevant.

The point is that the contents of the file, which serve as the mode of presentation, bring us into contact with the object of the file by representing certain of its properties. It is a total distortion of the "file" model to stick the object of the file itself into the file and say that it serves as its own mode of presentation, unless, of course, we are already taking for granted that it is representing itself. But then we might as well do away with the file altogether. The entire burden of representing the

phenomenal property is borne by the phenomenal state itself. To let the object itself serve as the mode of presentation of the demonstrative representation that picks it out is already to presume the kind of self-representation involved in the second model.

Now let's take it in the other direction. The idea is supposed to be that when a phenomenal property is instantiated by a token phenomenal state, it is thereby represented by this token. The question that immediately arises, of course, is how the representation relation is established in this case. After all, normally, instantiating a property is neither necessary nor sufficient for representing it. So what is it in this case that makes the instantiation of the property also a representation of it?

Clearly, the answer must lie with the functional role occupied by phenomenal states. Most representations represent what they do by virtue of some causal/informational link with their referents. In some cases, however, semantic significance can arise purely as a matter of functional role. A good example of this might be the logical constants. It's hard to see how a symbol meaning conjunction could acquire that meaning by virtue of some causal/informational link with the relevant truth-function: what would that mean? However, one could see how a symbol might count as representing conjunction by virtue of its interactions with other symbols. Using this analogy, one might imagine that by virtue of some particular pattern of causal interactions with other representational states, phenomenal states could acquire interpretations that involved reference to the very phenomenal properties they instantiate.

Let's assume, then, that the representation of phenomenal properties is effected in some way by functional role. There are two ways to interpret this appeal to functional role. The first is to imagine that for each phenomenal property, there is a unique functional role that serves to designate it (or, to be more precise, is such that by virtue of playing that role, a state represents it). There are two reasons this couldn't be right. First, there are just too many distinguishable phenomenal properties for there to be a distinct type of functional role corresponding to each. It's also unclear how a functional role could even serve this function of picking out a particular phenomenal property. But even if this problem could be overcome, this way of interpreting the appeal to functional role doesn't really capture the whole idea of using a phenomenal state as its own mode of presentation, of realizing the relation of acquaintance. After all, if playing functional role *Fr* is what makes a state represent phenomenal property *R*, then the fact that it's an instance of *R* that is playing the role representing *R* would be beside the point. It could just as easily have been some other state. What's needed, then, is a way of interpreting the appeal to functional role that takes directly into account the identity of the role filler.

The second way of interpreting the appeal to functional role is to imagine a single type of role covering all phenomenal states/properties, one that serves to pick out whatever phenomenal property it is that is instanced by the token filling the role. This would make the identity of the token phenomenal state an essential part of the representation. In this case, we would have a genuine case of self-representation. Phenomenal states, by virtue of playing this particular functional

role, would be interpreted as “saying” something like “the phenomenal property I am currently instantiating” or just “this phenomenal property.”⁵

At this point it should be clear why the self-representation model really amounts to the same thing as the demonstrative model. The functional role by virtue of which the phenomenal state is representing the phenomenal property it instantiates serves as a demonstrative, or indexical. How, after all, would one implement such a role? One obvious way would be to use location. Symbols occupying certain locations would be interpreted as referring to themselves, or something like that. But then the location is really just like a pointer; it has the same significance as a symbol saying “this one.”⁶

The crucial feature of the model under consideration—whether or not an explicitly demonstrative element is present—is the presence of the phenomenal state itself within the implementation of the corresponding phenomenal concept. The physical presence of an instance of the phenomenal property is thereby supposed to explain the especially intimate cognitive relation afforded by phenomenal concepts. In other words, a phenomenal concept affords acquaintance with the relevant phenomenal property by containing an instance of that property within it.

It’s hard to imagine how else a physicalist could capture cognitive immediacy, or acquaintance. If physically placing the relevant state right into the structure that realizes a phenomenal concept doesn’t do it, what could? Yet, it seems to me that putting the matter this starkly merely highlights the model’s inadequacy. Acquaintance, or cognitive presence, or whatever it is that is supposed to constitute the especially immediate and intimate cognitive relation between phenomenal concepts and their objects, is just that: a *cognitive* relation. It is not at all clear why, or how, *physical* presence translates into cognitive presence. In general, when considering the cognitive properties of a representational system, the physical identities of the implementing tokens are irrelevant. What matters is how the various tokens relate to each other. Their relations to their objects matter only to the extent that it is necessary to determine from those relations what they represent. But once that is determined, it is unclear how differences in the mechanisms of the representation relation are supposed to explain differences in cognitive significance.

We must remember here what we’re after. The bottom line is that we want an explanation of the existence of the explanatory gap that satisfies the Materialist Constraint. Why should the phenomenal character of a sensory experience seem inexplicable in terms of the corresponding physiological states? In particular, how could there even be a question of explaining the phenomenal character if it *just is* a certain physiological property? Furthermore, as we saw earlier, when considering creatures physically different from us, we can’t help feeling that there is a

5. Two advocates of this view, Katalin Balog (2002) and David Papineau (2002), have, independently, drawn an analogy to quotation. The “quoted material” is the state itself—that is, the property instantiated by the state—and the quoting mechanism is the functional role.

6. The analogy with quotation makes this indexical feature especially apparent. Quotation could just as easily have been implemented by prefixing a word with “this word”—e.g., this word (dog), instead of “dog.”

substantive, nonsemantic open question regarding the nature and/or existence of their conscious experience. But this must be an illusion if the materialist identity theory is correct. Clearly something about the way phenomenal states present themselves to us in the first-person mode make them seem distinct from whatever is represented by their third-person descriptions. This is the source of the idea that the explanatory gap arises here, in the psychophysical case, but not in other cases of theoretical reductions, because of the peculiar concepts by which we represent phenomenal properties in standard first-person access.

After surveying various candidates for the peculiar feature on which to pin the blame, we saw that it isn't merely a matter of representational or epistemic primitiveness. The problem isn't that the conception of phenomenal properties afforded by phenomenal concepts is too thin, lacking connections to other concepts. If anything, the problem seems to be quite the opposite. The first-person access we have to the properties of experience seems quite rich; we are afforded a very substantive and determinate conception of a reddish experience merely by having it. The idea, then, is that it makes sense that we would find it hard to see how that with which we were acquainted in this way could be the very same thing as that which is picked out without the benefit of acquaintance.⁷ But if this is what we're after, then it doesn't help merely to find some functional difference between phenomenal concepts and others. We need to find a difference that plausibly reconstructs the acquaintance/non-acquaintance distinction.

The proposal under consideration is supposed to do just that. Acquaintance is explained by the physical presence of the represented within the representation itself. But, I ask again: How is physical presence an explanation of cognitive presence? Sure, when we think of a relation such as acquaintance, with its sense of immediacy, metaphorical language about "sticking the object right in there" irresistibly comes to mind. But this language is metaphorical. The current proposal is to solve the problem by taking it literally.

Perhaps, in the end, this is the right way to go. But we are still owed an account of how physical presence alone is responsible for cognitive presence. That is, how does the presence of the relevant state within the physical implementation of the representation become something of which we are aware? It still smacks of that famous cartoon of the physicist scribbling all these formulas on the blackboard with this one circle in the middle that says "and then a miracle occurs." The transition from physical containment to awareness—the special kind allegedly afforded by phenomenal concepts—is still an inexplicable transition. It is subject to its own explanatory gap, just as much as is the original relation between phenomenal properties and their physical correlates.⁸

7. Notice again that I don't say, "by description." I leave open the possibility that whether a representation directly refers is not, contra Russell, determined by the presence or absence of the acquaintance relation.

8. In note 5, I mentioned Papineau's (2002) quotation-like view of phenomenal concepts. For him, however, the explanatory punch of appeal to this mechanism has nothing to do with an acquaintance-like relation, so a word about his view is in order.

There are two points I want to address in conclusion. First, someone might object to my entire argument as follows.⁹ What you are asking for, goes the objection, is, in effect, to bridge the explanatory gap. But this is precisely what we materialists admit can't be done, as one could predict from the various models surveyed here. So you are asking for the impossible, and the inability to do the impossible does not count against any theory.

My response is that I'm only requiring of the various models surveyed that they accomplish what they are advertised to accomplish: namely, explain why there is an explanatory gap. Granted, I'm asking for an explanation. But remember the original gap separated phenomenal properties from their underlying physical mechanisms. This gap I am not asking to be bridged. However, materialists claim that though they cannot explain phenomenal properties in terms of physical properties, they *can* explain why they can't explain it. The various models of how to

According to Papineau, the "intuition of distinctness" arises as follows. During the exercise of phenomenal concepts, actual instances of a phenomenal property are involved, whereas during the exercise of nonphenomenal concepts of the same properties, such instances are not involved. As he puts it:

The experience itself is in a sense being *used* in our thinking, and so is present in us. For this reason exercising a phenomenal concept will *feel* like having the experience itself. . . . So there is an intuitive sense in which exercises of material concepts "leave out" the experience at issue. They . . . don't activate or involve these experiences. (170)

The idea seems to be that because we're actually having the experience in the one case and not in the other, we believe that only the former is really about the experience. Our intuition of distinctness is merely a matter of the involvement of the phenomenal state itself in the one case and its absence in the other.

However, this appearance is deceiving. To begin with, why should the presence of the phenomenal state itself lead to an intuition of distinctness? After all, demonstratives rely on the presence of their objects and yet they bring about no special intuition of distinctness when flanking a nondemonstrative on the other side of an identity sign. One might reply that in the case of phenomenal concepts, the phenomenal states are not merely there to be demonstrated, but are actually components of the representations themselves. But it's unclear what the force of this distinction is unless we appeal to an acquaintance-like relation.

Here's what I have in mind. I've argued here that there is no substantive difference between the self-representation and demonstrative models. In the demonstrative model, one symbol functions as a demonstrative and has the demonstratum as its semantic content. In the self-representation model, the referent itself serves as the symbol that picks it out. But if we think of self-representation on analogy with quotation, self-representation has two components: the quotation mechanism and what appears between the quotation marks. How is this substantively different from a demonstrative picking out an object that is present, though off to the side? How does incorporation of the referent itself into the representation in this way make any real difference?

This is where acquaintance comes in. An excerpt from the Papineau quote above is noteworthy: "For this reason exercising a phenomenal concept will *feel* like having the experience itself." Of course, in one sense of "feel like," this is trivial: "feeling" the experience is having it. But that isn't enough, for we are constantly occupying all sorts of states of which we are unaware and that have no cognitive significance to us. Rather, for the presence of the experience when exercising a phenomenal concept to make a difference, "feeling" it must itself carry with it awareness; it must itself rise to the level of cognitive significance. But this is precisely what we don't know how to implement in a purely physical system. This is precisely the kind of cognitive presence, or immediacy, that is constitutive of an acquaintance-like relation, and for which we have no model that satisfies the Materialist Constraint.

9. This is inspired by remarks of both Barry Loewer and Janet Levin, though I cannot guarantee I'm properly capturing their concern.

implement phenomenal concepts is supposed to accomplish precisely that explanatory task. So, since the challenge that materialists say they can meet is itself to provide an explanation—in this case, of the fact that there is an explanatory gap between the physical and the phenomenal—it's a perfectly legitimate objection to point out that they haven't provided the requisite explanation. One might say that there now is a second explanatory gap: between implementations of cognitive architecture and whatever it is about phenomenal concepts—in my terms, that they afford genuine cognitive presence to phenomenal properties—that is responsible for the original explanatory gap. If one thought the original explanatory gap was a problem and needed to be explained away, then one ought to be bothered by this one as well.

Notice that nothing I've argued is intended to show that phenomenal properties aren't physical properties in the end. For that matter, phenomenal concepts may indeed be physically realized in one of the ways described above. The problem is that we don't understand how either story could be true, how the features we encounter in experience, or the encountering relation itself, could turn out to be a neural mechanism. This is indeed a situation materialists should find troubling.

Finally, I want to end with a bit of speculation. Suppose I'm right that we can't now imagine how a materialist story of phenomenal concepts would go. No mere physical-causal mechanism can provide the kind of cognitive presence we seem to enjoy with respect to our phenomenal experience. So what is it we need? It seems to me that we need something like the old-fashioned relation of acquaintance. We are acquainted with our experience, and as acquaintance *presents* properties, not merely represents them, we find it difficult to integrate what is presented with what is only represented in a way that allows the latter to explain the former. If acquaintance itself cannot be explained in terms of physical-causal mechanisms, as I claim (at least so far) it can't, then we have to contemplate the possibility that it is a brute relation. If so, then the Materialist Constraint is violated, and materialism is false.

It could turn out, then, that materialism is false not because phenomenal properties themselves are not physical—they may yet be for all we know. Rather, it would be false, on this view, because somehow we embody a relation to them that is itself brute and irreducible to physical relations. Is this a coherent position? Could phenomenal properties be physical while acquaintance is not? I don't know, but the question, to my mind, deserves exploration.

But haven't we already done away with acquaintance, sense data, and the "myth of the given"? One answer is to just say, "maybe not." On the other hand, nothing in what I've said entails that there are sense data or that acquaintance plays the same foundational epistemic role assigned to it by Russell and the positivists. However, I do want to acknowledge that there is much to puzzle about in allowing a relation of acquaintance. Does acquaintance entail "revelation," the doctrine that the essential nature of that with which we are acquainted is revealed thereby? Indeed, is the notion even coherent in the end? That is, if we abandon materialism for its inability to explain phenomenal experience, do we then flirt with outright incoherence instead? These, too, are questions to which I have no answer at present. But, again, I think they deserve exploration.

ACKNOWLEDGMENTS

Earlier versions of this chapter were presented to the NEH Institute on Consciousness and Intentionality, University of California—Santa Cruz (2002); the Jowett Society at Oxford University; the University of St. Andrews; Ohio State University; Bowling Green University; the conference on “Consciousness: Conceptual and Explanatory Issues” in Magdeburg, Germany; the University of Canterbury, Christchurch, New Zealand; the Australian National University; and the University of Otago, Dunedin, New Zealand. I want to thank all the members of these audiences for their helpful comments and criticisms. I especially want to thank Katalin Balog, Janet Levin, and Susanna Siegel for their comments on earlier drafts.

REFERENCES

- Balog, K. (2002). The “Quotational Account” of Phenomenal Concepts. Unpublished.
- Hill, C. S., and McLaughlin, B. P. (1999). There Are Fewer Things in Reality Than Are Dreamt of in Chalmers’ Philosophy. *Philosophy and Phenomenological Research* 59: 445–54.
- Jackson, F. (1982). Epiphenomenal Qualia. *Philosophical Quarterly* 32: 127–36.
- Loar, B. (1990/97). Phenomenal States. In *Philosophical Perspectives 4: Action Theory and Philosophy of Mind*, ed. J. Tomberlin: 81–108. Atascadero, Calif: Ridgeview. Revised version in *The Nature of Consciousness*, ed. N. Block, O. Flanagan, and G. Güzeldere: 597–616. Cambridge: MIT Press, 1997.
- Nida-Rümelin, M. (1995). What Mary Couldn’t Know: Belief about Phenomenal States. In *Conscious Experience*, ed. T. Metzinger: 219–42. Paderborn: Schöningh/Imprint Academic.
- Papineau, D. (2002). *Thinking about Consciousness*. Oxford: Oxford University Press.
- Perry, J. (2001). *Knowledge, Possibility, and Consciousness*. Cambridge: MIT Press.
- Russell, B. (1912). *The Problems of Philosophy*. New York: Oxford University Press.
- Sturgeon, S. (2000). *Matters of Mind*. London: Routledge.
- Tye, M. (2000). *Consciousness, Color, and Content*. Cambridge: MIT Press.