

# Phenomenal Concepts and the Explanatory Gap

---

**David J. Chalmers**  
**Philosophy Program**  
**Research School of Social Sciences**  
**Australian National University**

Confronted with the apparent explanatory gap between physical processes and consciousness, philosophers have reacted in many different ways. Some deny that any explanatory gap exists at all. Some hold that there is an explanatory gap for now, but that it will eventually be closed. Some hold that the explanatory gap corresponds to an ontological gap in nature.

In this paper, I want to explore another reaction to the explanatory gap. Those who react in this way agree that there is an explanatory gap, but they hold that it stems from the way we *think* about consciousness. In particular, this view locates the gap in the relationship between our *concepts* of physical processes and our *concepts* of consciousness, rather than in the relationship between physical processes and consciousness themselves.

Following Stoljar (2005), we can call this the *phenomenal concept strategy*. Proponents of this strategy argue that phenomenal concepts—our concepts of conscious states—have a certain special nature. Proponents suggest that given this special nature, it is predictable that we will find an explanatory gap between physical processes conceived under physical concepts, and conscious states conceived under phenomenal concepts. At the same time, they argue that our possession of concepts with this special nature can itself be explained in physical terms.

---

Published in (T. Alter and S. Walter. eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (Oxford University Press, 2006). This chapter was greatly influenced by a round-table discussion at the 2002 NEH Summer Institute on Consciousness and Intentionality in which Kati Balog, Ned Block, John Hawthorne, Joe Levine, and Scott Sturgeon, and others took part. I was also influenced by Levine's "Phenomenal Concepts and the Materialist Constraint". I first formally presented this chapter at a session on David Papineau's book *Thinking about Consciousness* at the 2003 Pacific APA meeting, and has also been presented at ANU, Arizona, Birkbeck College, Sydney, Texas A&M, Texas Tech, Victoria, and Wisconsin, as well as at workshops in Buenos Aires and Copenhagen. Thanks to all those present on those occasions for very useful reactions.

If this is right, then we may not have a straightforward physical explanation of consciousness, but we have the next best thing: a physical explanation of why we find an explanatory gap. From here, proponents infer that the existence of the explanatory gap is entirely compatible with the truth of physicalism. From there, they infer that there can be no sound argument from the existence of the explanatory gap to the falsity of physicalism.

In addition, proponents often use this strategy to deflate other intuitions that lead some to reject physicalism about consciousness: intuitions about conceivability and about knowledge, for example. They suggest that these intuitions are consequences of the special nature of phenomenal concepts (which, again, can itself be explained in physical terms). They conclude that these intuitions cannot give us conclusive reason to reject physicalism.

This extremely interesting strategy is perhaps the most attractive option for a physicalist to take in responding to the problem of consciousness. If it succeeded, the strategy would respect both the reality of consciousness and the epistemic intuitions that generate the puzzle of consciousness while explaining why these phenomena are entirely compatible with physicalism.

I think that the strategy cannot succeed. On close examination, we can see that no account of phenomenal concepts is both powerful enough to explain our epistemic situation with regard to consciousness and tame enough to be explained in physical terms. That is, if the relevant features of phenomenal concepts can be explained in physical terms, the features cannot explain the explanatory gap. And if the features can explain the explanatory gap, they cannot themselves be explained in physical terms. In what follows I will explain why.

## **1 Epistemic Gaps and Ontological Gaps**

Let  $P$  be the complete microphysical truth about the universe: a long conjunctive sentence detailing the fundamental microphysical properties of every fundamental microphysical entity across space and time. Let  $Q$  be an arbitrary truth about phenomenal consciousness: for example, the truth that somebody is phenomenally conscious (that is, that there is something it is like to be that person) or that I am experiencing a certain shade of phenomenal blueness.

Many puzzles of consciousness start from the observation that there is an apparent *epistemic gap* between  $P$  and  $Q$ : a gap between knowledge of  $P$  and knowledge of  $Q$ , or between our conception of  $P$  and our conception of  $Q$ .

Take Frank Jackson's case of Mary in the black-and-white room, who knows all the microphysical facts but who still does not know what it is like to see red. It appears that Mary may know  $P$  and may have no limitations on powers of a priori reasoning, but may still fail to know  $Q$  (where here  $Q$  is a truth about what it is like for ordinary people to see red things). This suggests that the truth of  $Q$  is not deducible by a priori reasoning from the truth of  $P$ . More specifically, it suggests that the material conditional  $P \supset Q$  is not knowable a priori.

Or take the conceivability of zombies. A zombie is a hypothetical creature that is physically identical to a conscious being but is not conscious at all. Many people hold that zombies are conceivable in principle, and they hold further that in principle one could conceive of a zombie world: one that is physically identical to ours, but without consciousness. Many people also hold that we can conceive of an *inverted world*: one that is physically identical to ours, but in which some conscious states differ from the corresponding states in our world. If this is right, then there is a gap between conceiving of  $P$  and conceiving of  $Q$ . It appears that  $P \& \sim Q$  is conceivable, where  $Q$  is a truth such as 'Someone is phenomenally conscious' (in the first case), or a truth specifying a particular state of phenomenal consciousness (in the second).

(I will not say much about exactly what conceivability involves, because most of what I say will be compatible with various understandings of conceivability. But at minimum, we can say that the conceivability of  $S$  requires that the truth of  $S$  cannot be ruled out a priori. This is the notion that I have elsewhere called *negative* conceivability (strictly: ideal primary negative conceivability). One may also suggest that the conceivability of  $S$  requires that one can clearly and distinctly imagine a situation in which  $S$  is the case. This is the notion that I have elsewhere called *positive* conceivability (strictly: ideal primary positive conceivability). I think that positive conceivability is the canonical notion of conceivability, but for the most part, the arguments in this chapter can operate with either notion. In those cases in which the distinction is relevant, I will make it explicit. For much more on these notions of conceivability, see Chalmers 2002.)

Many hold further that these epistemic gaps go along with an explanatory gap between  $P$  and  $Q$ . The explanatory gap comes from considering the question, Why,

given that  $P$  is the case, is  $Q$  the case? (Why, given that  $P$  is the case, is there phenomenal consciousness? And why are there the specific conscious states that there are?) The gap is grounded in part in the apparent inability to deduce  $Q$  from  $P$ : if one cannot deduce that  $Q$  is the case from the information that  $P$  is the case, then it is hard to see how one could explain the truth of  $Q$  *wholly* in terms of the truth of  $P$ . It is grounded even more strongly in the conceivability of  $P$  without  $Q$ . If one can conceive of a world that is physically just like this one but without consciousness, then it seems that one has to add something more to  $P$  to explain why there is consciousness in our world. And if one can conceive of a world that is physically just like this one but with different states of consciousness, then it seems that one has to add something more to  $P$  to explain why conscious states are the way they are in our world.

From these epistemic gaps, some infer an ontological gap. One may infer this ontological gap directly from the explanatory gap: if we cannot explain consciousness in terms of physical processes, then consciousness cannot be a physical process. Or one may infer it from one of the other epistemic gaps. For example, one may infer from the claim that  $P \& \sim Q$  is conceivable that  $P \& \sim Q$  is metaphysically possible, and conclude that physicalism is false. If there is a possible world physically just like this one but without consciousness, then the existence of consciousness is an ontologically further fact about our world.

At this point, materialists typically respond in one of two ways. Type-A materialists deny the epistemic gap. Paradigmatic type-A materialists deny there is any factual knowledge that Mary lacks inside her black-and-white room; they deny that zombies are conceivable, at least on ideal reflection; and they deny that there is an explanatory gap that survives reflection. Type-A materialism is an important view, but proponents of the phenomenal concept strategy reject type-A materialism, so I will not discuss it further here.

Type-B materialists accept that there is an epistemic gap but deny the inference to an ontological gap. Paradigmatic type-B materialists hold that Mary lacks knowledge, but not of ontologically distinct facts about the world; they hold that zombies are conceivable but not metaphysically possible; and that although there may be no satisfying explanation of consciousness in physical terms, consciousness is a physical process all the same.

Type-B materialists typically embrace *conceptual dualism* combined with *ontological monism*. They hold that phenomenal *concepts* are distinct from any physical or functional concepts. But they hold that phenomenal *properties* are identical to certain physical or functional properties, or at least that they are constituted by these properties in such a way that they supervene on them with metaphysical necessity. In this view, conceptual dualism gives rise to the explanatory gap, whereas ontological monism avoids any ontological gap.

Here type-B materialists often appeal to analogies with other cases in which distinct concepts refer to the same property. ‘Heat’ and ‘molecular motion’ express distinct concepts, for example, but many hold that they refer to the same property. By analogy, some type-B materialists suggest that a phenomenal term (e.g., ‘pain’) and a physical term (e.g., ‘C-fiber firing’) may express distinct concepts but pick out the same property. More generally, type-B materialists typically hold that the material conditional ‘ $P \supset Q$ ’ is an instance of Kripke’s necessary a posteriori: like ‘water is  $H_2O$ ,’ the conditional is not knowable a priori, but it is true in all possible worlds. If successful, these analogies would reconcile the epistemic gap with ontological monism.

However, the success of these analogies is widely disputed. Kripke himself argued that the relation between mental and physical expressions is different in kind from the relation between ‘heat’ and ‘the motion of molecules,’ or that between ‘water’ and ‘ $H_2O$ ,’ so that the grounds for a posteriori identities or necessities in these standard cases are not present in the mental-physical case. Since then, many opponents and even proponents of type-B materialism have argued that mental and physical properties are not analogous. Some (e.g., White 1986, Loar 1990/97) argue that in the standard cases, the distinct concepts (e.g., ‘heat’ and ‘the motion of molecules’) are associated with distinct properties at least as modes of presentation of their referent, if not as their actual referent. Some (e.g., Chalmers 1996, 2002) argue that the standard cases are all compatible with an attenuated link between conceivability and possibility, expressible using two-dimensional semantics. Some (e.g., Jackson 1998) argue that the standard cases are all compatible with the thesis that physicalism requires a priori entailment of all truths by physical truths. Some (e.g., Levine 2001) argue that the physical-phenomenal case involves a ‘thick’ explanatory gap that is unlike those present in the standard cases.

These differences strongly suggest that the standard way of reconciling conceptual dualism with ontological monism does not apply to the conceptual dualism between the physical and the phenomenal. If the principles that hold in the standard cases applied here, then the conceptual dualism would lead to an ontological dualism. For example, we would expect distinct properties to serve as modes of presentation for physical and phenomenal concepts; and from here one can reason to an underlying ontological dualism at the level of these properties. Likewise, we would expect there to be some metaphysically possible world in the vicinity of what we conceive when we conceive of zombies and inverts; and from here one can reason to a failure of metaphysical supervenience of everything on the physical. If so, then the epistemic gap will once again lead to an ontological gap.

## 2 The Phenomenal Concept Strategy

Partly to avoid these problems, many type-B materialists have turned to a different strategy for reconciling conceptual dualism and ontological monism. Instead of focusing on quite general features of a posteriori identities and necessities, this strategy focuses on features that are specific to phenomenal concepts. Proponents of the phenomenal concept strategy typically allow that we are faced with a distinctive epistemic gap and the physical-phenomenal case, one that is in certain respects unlike the epistemic gaps that one finds in the standard cases. But they hold that this distinctive epistemic gap can be explained in terms of certain distinctive features of phenomenal concepts. And they hold that these distinctive features are themselves compatible with an underlying ontological monism.

*Recognitional concepts:* The *locus classicus* for the phenomenal concept strategy is Brian Loar's paper 'Phenomenal States' (1990/97), in which he suggests that phenomenal concepts are recognitional concepts that pick out their objects via noncontingent modes of presentation. (Related proposals involving recognitional concepts are made by Carruthers 2004, Tye 2003 and Levin 2006.) Recognitional concepts are concepts deployed when we recognize an object as being one of *those*, without relying on theoretical knowledge or other background knowledge. For example, we may have a recognitional concept of a certain sort of cactus. One may also have a theoretical concept of that sort of cactus, so that there are two concepts referring to the same sort of entity. In standard cases, these two concepts will be

associated with distinct properties as modes of presentation (for example, one's recognitional concept of a cactus may be associated with the property *typically causes such-and-such experience*), so this will not ground a full-scale ontological monism. But Loar suggests that phenomenal concepts are special recognitional concepts because the property that is the referent also serves as a mode of presentation. He argues that this special character of phenomenal concepts explains the distinctive epistemic gap in a manner that is compatible with ontological monism.

*Distinct conceptual roles:* Developing a suggestion by Nagel (1974), Hill (1997; see also Hill and McLaughlin 1999) suggests that phenomenal concepts and physical concepts are associated with distinct faculties and modes of reasoning, and that they play very different conceptual roles. Hill argues that the distinctive epistemic gaps between the physical and phenomenal are explained by this distinctness in conceptual roles, and he suggests that we should expect the epistemic gaps to be present even if the distinct concepts refer to the same property.

*Indexical concepts:* A number of philosophers (including Ismael 1999, O'Dea 2002, and Perry 2001) have suggested that phenomenal concepts are a sort of indexical concept, analogous to *I* and *now*. There are familiar epistemic gaps between objective and indexical concepts, noted by Perry (1977) and many others. For example, even given complete objective knowledge of the world, one might not be able to know what time it is now, or where one is located. Proponents of the indexical concept strategy suggest that the epistemic gap between the physical and phenomenal has a similar character. On this view, just as 'now' picks out a certain objective time under an indexical mode of presentation, phenomenal concepts pick out states of the brain under an indexical mode of presentation.

*Quotational concepts:* Finally, some philosophers have suggested that phenomenal concepts are special because their referents—phenomenal states—serve as constituents of the concepts themselves (or as constituents of the corresponding mental representations). Sometimes this view of phenomenal concepts is put forward without any associated ambition to support type-B materialism (e.g., Chalmers 2003a). But some, such as Papineau (2006) and Block (2006), suggest that this view of phenomenal concepts can explain the epistemic gap in terms acceptable to a materialist. For example, Papineau sees phenomenal concepts as *quotational concepts*, which represent their referent as *That state: —*, where the blank space is filled by an embedded phenomenal state in a way loosely analogous to the way that a

word might be embedded between quotation marks. Papineau suggests that even if the embedded state is a neural state, this quotational structure will still give rise to the familiar epistemic gaps.

Other proponents of the phenomenal concept strategy include Sturgeon (1994), who proposes that the explanatory gap is grounded in the fact that phenomenal states serve as their own canonical evidence; Levine (2001), who suggests that phenomenal concepts may crucially involve a nonascriptive mode of presentation of their referent; and Aydede and Güzeldere (2005), who give an information-theoretic analysis of the special relation between phenomenal concepts and perceptual concepts.

I have discussed many of these views elsewhere. (See Chalmers 1999 for discussion of the first two views, and Chalmers 2003a for discussion of the third and fourth.) Here I will focus instead on what is common to all the views, arguing on quite general grounds that no instance of the phenomenal concept strategy can succeed in grounding a type-B materialist view of the phenomenal. Later I will apply this general argument to some specific views.

The general structure of the phenomenal concept strategy can be represented as follows. Proponents put forward a thesis *C* attributing certain psychological features—call these the key features—to human beings. They argue (1) that *C* is true: humans actually have the key features; (2) that *C* explains our epistemic situation with regard to consciousness: *C* explains why we are confronted with the relevant distinctive epistemic gaps; and (3) that *C* itself can be explained in physical terms: one can (at least in principle) give a materialistically acceptable explanation of how it is that humans have the key features.

This is a powerful strategy. If it is successful, we may not have a direct physical explanation of consciousness, but we will have the next best thing: a physical explanation of the explanatory gap. One might plausibly hold that if we have a physical explanation of all the epistemic data that generate arguments for dualism, then the force of these arguments will be undercut. I think this matter is not completely obvious—one might hold that the residual first-order explanatory gap still poses a problem for physicalism—but I will concede the point for the purposes of this paper. There is no question that a physical explanation of the relevant epistemic gaps would at least carry considerable force in favor of physicalism.

Note that for the strategy to work, all three components are essential. If (1) or (2) fail, then the presence of the relevant epistemic gaps in us will not be explained. If (3)

fails, on the other hand, then although thesis *C* may help us understand the conceptual structure of the epistemic gap, it will carry no weight in deflating the gap. If the epistemic gap is grounded in special features of phenomenal concepts that are not physically explainable, then these features will generate a gap of their own. Opponents of the strategy will then argue that the special features themselves require nonphysical explanation, and may plausibly suggest that the special features themselves reflect the presence of irreducible phenomenal experience. If so, the phenomenal concept strategy will do little to support physicalism.

It should be noted that not all proponents of the phenomenal concept strategy are explicitly committed to (3), the thesis that the relevant features of phenomenal concepts must be physically explicable. Some proponents, such as Loar and Sturgeon, are silent on the matter. Almost all of them, however, use the phenomenal concept strategy to resist the inference from the epistemic gap to an ontological gap. I will argue later that without (3), the phenomenal concept strategy has no force in resisting this inference.

There is a related strategy that I will *not* discuss here. This is the type-A materialist strategy of appealing to psychological features to explain why we have false beliefs or mistaken epistemic intuitions about consciousness (see, for example, Dennett 1981; and Jackson 2003). In its most extreme form, this strategy may involve an attempted psychological explanation of why we think we are conscious, when in fact we are not. In a less extreme form, the strategy may involve an attempted psychological explanation of why we think there is an epistemic gap between physical and phenomenal truths when in fact there is not. For example, it may attempt to explain why we think Mary gains new knowledge when in fact she does not, or why we think zombies are conceivable when in fact they are not. This is an important and interesting strategy, but it is not my target here. My target is, rather, a type-B materialist who accepts that we are phenomenally conscious and that there is an epistemic gap between physical and phenomenal truths, and who aims to give a psychological explanation of the existence of this epistemic gap.

### **3 A Master Argument**

I will argue that no account can simultaneously satisfy (2) and (3). For any candidate thesis *C* about psychological features of human beings, then either

(1) *C* is not physically explicable

or

(2) *C* does not explain our epistemic situation with regard to consciousness.

Here the key question will be: is  $P \& \sim C$  conceivable? That is, can we conceive of beings physically identical to us (in physically identical environments, if necessary) that do *not* have the psychological features attributed by thesis *C*?

One might approach this question by asking, Would zombies have the key features attributed by thesis *C*? Or at least by asking, Is it conceivable that zombies lack the key features? Note that neither question assumes that zombies are metaphysically possible. We simply need the assumption that zombies are conceivable, an assumption that type-B materialists typically grant.

One can also approach the question by considering a scenario closer to home. Instead of considering physically identical zombies, we can consider functionally identical zombies: say, functionally identical creatures that have silicon chips where we have neurons and that lack consciousness. Most type-B materialists allow that it is at least an open epistemic possibility that silicon functional isomorphs in the actual world would lack consciousness. We can then ask, Assuming that these functional isomorphs lack consciousness, do they also lack the key features attributed by thesis *C*? If it is conceivable that a functional isomorph lacks these features, then it will almost certainly be conceivable that a physical isomorph lacks these features.

In any case, either physical duplicates that lack the key features are conceivable or they are not. This allows us to set up a master argument against the phenomenal concept strategy, in the form of a dilemma:

- (1) If  $P \& \sim C$  is conceivable, then  $C$  is not physically explicable.
- (2) If  $P \& \sim C$  is not conceivable, then  $C$  cannot explain our epistemic situation.

- 
- (3) Either  $C$  is not physically explicable, or  $C$  cannot explain our epistemic situation.

The argument is valid. It has the form of a dilemma, with each premise representing one of the horns. In what follows I will discuss each horn in turn, arguing for the corresponding premise.

### 3.1 First Horn: $P \& \sim C$ Is Conceivable

Premise (1) says that if  $P \& \sim C$  is conceivable, then  $C$  is not physically explicable. The argument for this premise is straightforward. It parallels the original reasoning from the claim that  $P \& \sim Q$  is conceivable to the claim that  $Q$  is not physically explicable. If one can conceive of physical duplicates that lack the key features attributed by thesis  $C$ , then there will be an explanatory gap between  $P$  and  $C$ . That is, there will be no wholly physical explanation that makes transparent why thesis  $C$  is true. To explain why, in the actual world, creatures with the relevant physical structure satisfy thesis  $C$ , we will need additional explanatory materials, just as we need such principles to explain why actual creatures with this physical structure are conscious.

Here, again, we are assuming nothing about the relationship between conceivability and possibility. It may be that creatures satisfying  $P \& \sim C$  are metaphysically impossible. We are simply assuming a connection between conceivability and explanation. More precisely, we are assuming a connection between conceivability and a certain sort of reductive explanation, the sort that is relevant here: explanation that makes transparent why some high-level truth obtains, given that certain low-level truths obtain. If it is conceivable that the low-level truths obtain without the high-level truths obtaining, then this sort of transparent explanation will fail. The original explanatory gap between consciousness and the physical turns on the absence of just this sort of transparent explanation. If it is conceivable that  $P$

obtains without *C* obtaining, then we will have just the same sort of explanatory gap between physical processes and the relevant features of phenomenal concepts.

Type-B materialists typically accept this connection between conceivability and transparent explanation, even though they reject the connection between conceivability and possibility. So for now, I will take the connection between conceivability and explanation for granted. Later I will argue that even rejecting the connection will not remove the dilemma for the type-B materialist.

One might think that a proponent of the phenomenal concept strategy *must* take this first horn of the dilemma, as thesis *C* will be a thesis about *phenomenal* concepts. If thesis *C* explicitly requires the existence of phenomenal concepts, and if phenomenal concepts require the existence of phenomenal states, then it is out of the question that zombies could have the features attributed by thesis *C*. If *C* builds in the truth of *Q*, and *P*&~*Q* is conceivable, then *P*&~*C* will automatically be conceivable. A physical explanation of the truth of thesis *C* would then be ruled out.

We can avoid this problem by stipulating that thesis *C* should be cast in *topic-neutral* terms: terms that do not explicitly attribute phenomenal states or concepts that refer to them. The restriction to topic-neutral terms allows that thesis *C* may include psychological or epistemological vocabulary, in addition to physical and functional vocabulary. But phenomenal vocabulary is barred. For example, instead of casting thesis *C* as a thesis explicitly about phenomenal concepts, one can cast it as a thesis about *quasi-phenomenal* concepts, where these can be understood as concepts deployed in certain circumstances that are associated with certain sorts of perceptual and introspective processes, and so on. Phenomenal concepts will be quasi-phenomenal concepts, but now it is not out of the question that zombies might have quasi-phenomenal concepts too.

Formulated this way, thesis *C* will then say that quasi-phenomenal concepts have certain properties, such as being recognitional concepts without contingent modes of presentation. We can likewise appeal to quasi-phenomenal concepts in characterizing our epistemic situation with regard to consciousness. This allows the possibility that even if consciousness cannot be physically explained, we might be able to physically explain the key psychological features and our epistemic situation. If we could physically explain why we are in such an epistemic situation, we would have done the crucial work in physically explaining the existence of an explanatory gap.

Henceforth, I will take it for granted that thesis  $C$  should be cast in topic-neutral terms. The same goes for the characterization of our epistemic situation. Understood this way, it is by no means out of the question that zombies would have quasi-phenomenal concepts with the properties in question, and that  $P \& \sim C$  is not conceivable, leading to the second horn of the dilemma. That question is no longer prejudiced by building in theses about phenomenology. Rather, the question will turn on the character of the psychological features themselves.

(This point is relevant to a discussion of this argument by David Papineau [2006], where he suggests that he might take both horns of the dilemma. Once one stipulates that  $C$  is cast in topic-neutral terms, it is clear that Papineau takes the second horn.)

Of course, it remains possible that even when thesis  $C$  is understood in topic-neutral terms, the character of the psychological features involved in  $C$  is such that  $P \& \sim C$  is conceivable. If so, then the first horn of the dilemma is raised as strongly as ever. On this horn, the relevant psychological features will raise just as much of an explanatory gap as consciousness itself, and an appeal to these features can do little to deflate the explanatory gap.

### 3.2 Second Horn: $P \& \sim C$ Is Not Conceivable

Premise 2 says that if  $P \& \sim C$  is not conceivable, then  $C$  cannot explain our epistemic situation. The case for this premise is not quite as straightforward as the case for premise 1. One can put the case informally as follows:

- (4) If  $P \& \sim C$  is not conceivable, then zombies satisfy  $C$ .
- (5) Zombies do not share our epistemic situation.
- (6) If zombies satisfy  $C$  but do not share our epistemic situation, then  $C$  cannot explain our epistemic situation.

- 
- (7) If  $P \& \sim C$  is not conceivable, then  $C$  cannot explain our epistemic situation.

Strictly speaking, the references to zombies should be put within the scope of a conceivability operator. One can formalize the argument in this fashion, but for now I will use the informal version for ease of discussion.

(A formalized version might run as follows, where  $c$  is a conceivability operator and  $E$  represents our epistemic situation: (4) If  $\sim c(P \& \sim C) \& c(P \& \sim E)$ , then  $c(P \& C \& \sim E)$ ; (5)  $c(P \& \sim E)$ ; (6) If  $c(P \& C \& \sim E)$ , then  $C$  cannot explain  $E$ ; so (7) If  $\sim c(P \& \sim C)$ , then  $C$  cannot explain  $E$ . Premise (4) is slightly more complicated in this version than in the informal version, in order to capture the crucial claim that the specific zombie relevant to (5) satisfies  $C$ . This formalization will not play a central role, as its force can be captured by the simpler formalization of the overall argument summarized in the next section.)

Here, premise (6) is simply another application of the connection between conceivability and explanation. Premise (4) might be derived from a principle of completeness about the conceivable (if  $R$  is conceivable, then for arbitrary  $S$ , either  $R \& S$  is conceivable, or  $R \& \sim S$  is conceivable). But in this context, one can also defend (4) more straightforwardly by noting that if the truth of  $C$  is transparently explained by  $P$ , as the first horn requires, then if we specify that  $P$  holds in a conceivable situation, it will follow transparently that  $C$  holds in that situation.

The real work in this argument is done by premise (5). This premise amounts to the claim that  $P \& \sim E$  is conceivable, where  $E$  characterizes our epistemic situation. To clarify this premise further, one needs to clarify the notion of our epistemic situation.

I will take it that the epistemic situation of an individual includes the truth-values of their beliefs and the epistemic status of their beliefs (as justified or unjustified, and as cognitively significant or insignificant). As before, an epistemic situation (and a sentence  $E$  characterizing it) should be understood in topic-neutral terms, so that it does not build in claims about the presence of phenomenal states or phenomenal concepts. We can say that two individuals share their epistemic situation when they have corresponding beliefs, all of which have corresponding truth-value and epistemic status.

A zombie will share the epistemic situation of a conscious being if the zombie and the conscious being have corresponding beliefs, all of which have corresponding truth-values and epistemic status. Here, I assume an intuitive notion of correspondence between the beliefs of a conscious being and the beliefs (if any) of its zombie twin. For example, corresponding utterances by a conscious being and its zombie twin will express corresponding beliefs. It is important to note that this notion of correspondence does not require that corresponding beliefs have the same *content*. It is plausible that a nonconscious being such as a zombie cannot have beliefs with

exactly the same content as our beliefs about consciousness. But we can nevertheless talk of the zombie's *corresponding* beliefs. So the claim that a zombie and a conscious being share their epistemic situation does not require that their beliefs have the same content. This mirrors the general requirement that epistemic situations be understood in topic-neutral terms.

I will assume here, at least for the sake of argument, that zombies can have beliefs (that is, that it is conceivable that zombies have beliefs). This is by no means obvious. But if zombies cannot have beliefs, then the phenomenal concept strategy cannot get off the ground. If zombies cannot have beliefs, then presumably they cannot possess concepts either, so there will be an explanatory gap between physical processes and the possession of concepts. If so, then there will be an explanatory gap between physical processes and the key features of phenomenal concepts, leading to the first horn of the dilemma. And even if zombies can have concepts with the key features, then as long as they cannot have beliefs, the key features cannot explain our epistemic situation, leading to the second horn of the dilemma. So the assumption that zombies can have beliefs should be seen as a concession to the type-B materialist for the sake of argument.

For a given conscious being with a given epistemic situation as understood above,  $E$  will be a sentence asserting the existence of a being with that epistemic situation. This sentence will be made true by that being in its original epistemic situation, and it will be made true by any being that shares this epistemic situation in the sense specified above. Premise 5, the claim that zombies do not share our epistemic situation, can be understood as the claim that  $P \& \sim E$  is conceivable, where  $E$  characterizes the epistemic situation of an actual conscious being. That is, it is the claim that (it is conceivable that) zombies' beliefs differ in their truth-value or their epistemic status from the corresponding beliefs of their actual conscious twins.

Why think that zombies do not share our epistemic situation? The first reason for this is intuitive. On the face of it, zombies have a much less accurate self-conception than conscious beings do. I believe that I am conscious, that I have states with remarkable qualitative character available to introspection, that these states resist transparent reductive explanation, and so on. My zombie twin has corresponding beliefs. It is not straightforward to determine just what content these beliefs might possess. But there is a strong intuition that these beliefs are false, or at least that they are less justified than my beliefs.

One can develop this intuitive consideration by considering a zombie's utterances of sentences such as 'I am phenomenally conscious.' It is not clear exactly what a zombie asserts in asserting this sentence. But it is plausible that the zombie does not assert a truth.

Balog (1999) suggests that the zombie does assert a truth, as its term 'phenomenal consciousness' will refer to a brain state. This seems to give implausible results, however. We can imagine a debate in a zombie world between a zombie eliminativist and a zombie realist:

Zombie Eliminativist: 'There's no such thing as phenomenal consciousness.'

Zombie Realist: 'Yes, there is.'

Zombie Eliminativist: 'We are conscious insofar as 'consciousness' is a functional concept, but we are not conscious in any further sense.'

Zombie Realist: 'No, we are conscious in a sense that is not functionally analyzable.'

When such a debate is held in the actual world, the type-B materialist and the property dualist agree that the zombie realist is right, and the zombie eliminativist is wrong. But it is plausible that in a zombie scenario, the zombie realist would be wrong, and the zombie eliminativist would be right. If so, then where we have true beliefs about consciousness, some corresponding beliefs of our zombie twins are false, so that zombies do not share our epistemic situation.

Still, because judgments about the truth-value of a zombie's judgments are disputed, we can also appeal to a different strategy, one that focuses on the nature of our knowledge compared to a zombie's knowledge. Let us focus on the epistemic situation of Mary, upon seeing red for the first time. Here, Mary gains cognitively significant knowledge of what it is like to see red, knowledge that could not be inferred from physical knowledge. What about Mary's zombie twin, Zombie Mary? What sort of knowledge does Zombie Mary gain when she emerges from the black-and-white room?

It is plausible that Zombie Mary at least gains certain abilities. For example, upon seeing a red thing, she will gain the ability to perceptually classify red things together. It is also reasonable to suppose that Zombie Mary will gain certain *indexical* knowledge, of the form *I am in this state now*, where *this state* functions indexically

to pick out whatever state she is in. But this knowledge is analogous to trivial indexical knowledge of the form *It is this time now*, and is equally cognitively insignificant. There is no reason to believe that Zombie Mary will gain cognitively significant introspective knowledge, analogous to the cognitively significant knowledge that Mary gains. On the face of it, there is nothing for Zombie Mary to gain knowledge of. For Zombie Mary, all is dark inside, so even confronting her with a new sort of stimulus will not bring about new significant introspective knowledge.

If this is right, then Zombie Mary does not share Mary's epistemic situation. In addition to her abilities and her indexical beliefs, Mary has significant knowledge of what it is like to see red, knowledge not inferable from her physical knowledge. But Zombie Mary does not have significant non-indexical knowledge that corresponds to Mary's knowledge. If so, then Zombie Mary does not share Mary's epistemic situation.

One can also bring out the contrast by considering a case somewhat closer to home. Balog (1999) appeals to hypothetical conscious humans called 'Yogis,' who have the ability to refer directly to their brain states by deploying direct recognitional concepts of those states, even when those states have no associated phenomenal quality. She suggests that zombies likewise might have direct recognitional knowledge of their brain states by deploying a recognitional concept analogous to a Yogi's.

Even if Yogi concepts like this are possible, however, it is clear that they are nothing like phenomenal concepts. A Yogi going into a new brain state for the first time *might* sometimes acquire a new recognitional concept associated with that state. But a Yogi will not acquire new cognitively significant knowledge that is analogous to Mary's phenomenal knowledge. At best, a Yogi will acquire trivial knowledge, which we might express roughly as 'that sort of brain state is that sort of brain state.' So even if Zombie Mary can have a recognitional concept like this, she will still not have an epistemic situation like Mary's.

(I think a Yogi's concept is probably best understood as a response-dependent concept: if the concept is *flurg*, it is priori for the Yogi that a flurg is whatever brain state normally triggers flurg-judgments. Once a Yogi discovers that brain state *B* triggers these judgments, he will know that a flurg is an instance of *B*, and there will be no further question about flurges. This contrasts with a phenomenal concept: once we discover that our phenomenal redness judgments are typically triggered by brain

state  $B$ , we will still regard the question of the nature of phenomenal redness as wide open. This difference between response-dependent concepts and phenomenal concepts tends to further undercut Balog's suggestion that Yogi's concepts are just like phenomenal concepts.)

If the above is correct, then  $P \& \sim E$  is conceivable, and premise 5 is correct. When this is combined with premises 4 and 6, the conclusion follows. That is: if  $P \& \sim C$  is not conceivable, then Zombie Mary has the psychological features attributed by  $C$ , but she does not share Mary's epistemic situation. So the psychological features attributed by  $C$  cannot explain Mary's epistemic situation, and more generally, cannot explain our epistemic situation with respect to consciousness.

### 3.3 Summary

We can summarize the arguments above more briefly as follows.

- (1)  $P \& \sim E$  is conceivable
- (2) If  $P \& \sim E$  is conceivable, then  $P \& \sim C$  is conceivable or  $C \& \sim E$  is conceivable.
- (3) If  $P \& \sim C$  is conceivable,  $P$  cannot explain  $C$ .
- (4) If  $C \& \sim E$  is conceivable,  $C$  cannot explain  $E$ .

- 
- (5)  $P$  cannot explain  $C$  or  $C$  cannot explain  $E$ .

Premise (1) is supported by the considerations about Zombie Mary above. Premise (2) is a plausible consequence of the logic of conceivability. Premises (3) and (4) are applications of the connection between conceivability and explanation. The conclusion says that  $C$  cannot satisfy the constraints laid out in the general requirements for the phenomenal concept strategy. The argument here is general, applying to any candidate for  $C$ . It follows that the phenomenal concept strategy cannot succeed: no psychological features are simultaneously physically explicable and able to explain the distinctive epistemic gaps in the phenomenal domain.

## 4 Reactions

Proponents of the phenomenal concept strategy may react to this argument in one of four ways. First, they may accept that  $P$  cannot explain  $C$  but hold that the phenomenal concept strategy still has force. Second, they may accept that  $C$  cannot explain  $E$  (at least as I have construed  $E$ ) but hold that the phenomenal concept strategy still has force. Third, they may deny that  $P \& \sim E$  is conceivable and hold that Zombie Mary shares the same epistemic situation as Mary. Fourth, they may deny the connection between conceivability and explanation. (Each of these reactions has been suggested in discussions I have had with type-B materialists, with the first and third reactions being more common than the second and fourth.) In what follows, I will discuss each of the reactions in turn.

### **Option 1: Accept That $P$ Cannot Explain $C$**

The first response adopts what we might call the ‘thick phenomenal concepts’ strategy. On this approach, proponents appeal to features of phenomenal concepts that are thick enough to explain our distinctive epistemic situation with respect to consciousness but are too thick to be physically explained.

An example of such an approach may be the proposal that phenomenal concepts involve a direct acquaintance with their referent of a sort that discloses an aspect of their referent’s intrinsic nature. Such a proposal may well help to explain the distinctive epistemic progress that Mary makes and that Zombie Mary does not make: Mary has concepts that involve direct acquaintance with their referents, whereas Zombie Mary does not. But the very fact that Mary has such concepts and that Zombie Mary does not suggests that this feature of phenomenal concepts cannot be physically explained. The proposal requires a special psychological feature (acquaintance) whose existence one would not predict from just the physical/functional structure of the brain.

The obvious problem here is the problem mentioned before. On this account, even if there is a sort of explanation of the explanatory gap in terms of features of phenomenal concepts, the explanatory gap recurs just as strongly in the explanation of phenomenal concepts themselves. Because of this, the strategy may make some progress in *diagnosing* the explanatory gap, but it will do little to deflate the gap.

A proponent may suggest that just as the first-order explanatory gap can be explained in terms of second-order features of phenomenal concepts, the second-order explanatory gap concerning phenomenal concepts can be explained in terms of third-order features of our concepts of phenomenal concepts, and so on. Alternatively, an opponent may suggest that the second-order explanatory gap can be explained in terms of the same second-order features of phenomenal concepts that explain the first-order explanatory gap. The first move here obviously leads to a regress of explanation, and the second move leads to a circular explanation. Explanatory structures of this sort can be informative, but again they will do nothing to deflate the explanatory gap unless the chain of explanation is at some point grounded in physical explanation.

A proponent may also suggest that to require that the key psychological features be physically explicable is to set the bar too high. On this view, all that is needed is a psychological explanation of the epistemic gap that is *compatible* with the truth of physicalism, not one that is itself transparently explainable in physical terms. However, an opponent will now question the compatibility of the account with the truth of physicalism. Just as the original explanatory gap gave reason to think that consciousness is not wholly physical, the new explanatory gap gives reason to think that phenomenal concepts are not wholly physical.

At this point, the proponent may respond by saying that ontological physicalism is compatible with the existence of explanatory gaps. But now we are back where we started, before the phenomenal concept strategy came in. Antiphysicalists argue from an epistemic gap to an ontological gap. The phenomenal concept strategy as outlined earlier was supposed to ground the rejection of this inference by showing how such epistemic gaps can arise in a purely physical system. If successful, the strategy would help to *justify* the claim that the epistemic gap is compatible with ontological physicalism, and so would lend significant support to type-B materialism. But the weaker version of the strategy outlined above can give no such support. On this version, the proponent needs *independent* grounds to reject the inference from an explanatory gap to an ontological gap. If the proponent has no such grounds, then the phenomenal concept strategy does nothing to provide them. An opponent will simply say that the explanatory gap between physical processes and phenomenal concepts provides all the more reason to reject physicalism. If the proponent already has such grounds, on the other hand, then the phenomenal concept strategy is rendered

redundant. Either way, the strategy will play no role in supporting type-B materialism against the antiphysicist.

This limitation does not entail that the limited version of the phenomenal concept strategy is without interest. Even if it does not *support* a type-B materialist view, we can see this sort of account of phenomenal concepts as helping to *flesh out* a type-B materialist view by giving an account of what phenomenal concepts might be like under the assumption that type-B materialism is true. If we have independent reasons to be type-B materialists, we may then have reason to suppose that phenomenal concepts work as the account suggested. And if we have some independent method of deflating the original explanatory gap, then presumably this method may also apply to the new explanatory gap. For example, if a type-B materialist accepts an explanatorily primitive identity between certain physical/functional properties and phenomenal properties, she may also accept an explanatorily primitive identity between certain physical/functional properties and the properties of phenomenal concepts. But insofar as one has reasons to reject type-B materialism, the phenomenal concept strategy will do nothing to undermine these reasons.

(Note that I am not arguing in this paper that type-B materialism is false. I have done that elsewhere. Here I am simply arguing that the phenomenal concept strategy provides no support for type-B materialism and provides no grounds for rejecting arguments from the epistemic gap to an ontological gap.)

Overall, I think that accepting an explanatory gap between physical processes and phenomenal concepts is the most reasonable reaction to the arguments above for a type-B materialist. To accept such a gap does not immediately rule out the truth of type-B materialism, and the account of phenomenal concepts may help in elaborating the position. But now the phenomenal concept strategy does nothing to *support* type-B materialism against the antimaterialist. To resist antimaterialist arguments, and to deflate the significance of the explanatory gap, the type-B materialist must look elsewhere.

### **Option 2: Accept That *C* Does Not Explain *E***

The second possible reaction for a type-B materialist is to embrace the second horn of the dilemma, accepting that the key psychological features that they appeal to do not explain our epistemic situation, at least as I have construed that epistemic situation. We might think of this as a ‘thin phenomenal concepts’ strategy. Here, the

psychological features in question are tame enough to be physically explained, but they are not powerful enough to explain the full-blown epistemic gaps associated with consciousness.

The problem with this strategy is the same as the problem for the first strategy. Because it leaves a residual explanatory gap, it does little to close the original explanatory gap. The issues that come up here are similar to the issues under the first reaction, so I will not go over them again. If anything, this reaction is less attractive than the first reaction because an account of phenomenal concepts that cannot explain our epistemic situation with regard to consciousness would seem to have very little to recommend it.

There is a version of this reaction that is worth attending to, however. This version concedes that the key psychological features in question cannot explain our full epistemic situation *as I have defined it*, but asserts that the features can explain our epistemic situation in a narrower sense, where it is this sense that is crucial to explaining away the explanatory gap. In particular, a proponent may suggest that I raised the bar unnecessarily high by stipulating that our epistemic situation includes the *truth-values* of our beliefs, and by including their status as *knowledge*. It may be suggested that there is a sense in which truth-value is external to our epistemic situation, and that the phenomenal concept strategy needs only to explain our epistemic situation more narrowly construed.

I can think of three main versions of this strategy. A proponent may suggest: (1) that a physically explicable account of phenomenal concepts can explain the *justification* of our phenomenal beliefs; or (2) that such an account can explain the *inferential disconnection* between our physical and phenomenal beliefs, including the fact that the latter are not deducible from the former, for example (this suggestion meshes especially well with Hill's account of phenomenal concepts in terms of dual conceptual roles); or (3) that such an account can explain the *existence* of our phenomenal beliefs and of associated beliefs, such as the belief in an explanatory gap. In each of these cases, proponents may claim that corresponding features will be present in zombies, so that there is no obstacle to a physical explanation.

I think that each of these strategies is interesting, but each suffers from the same problem. To restrict the ambition of the phenomenal concept strategy in this way undercuts its force in supporting type-B materialism. Recall that the strategy is intended to resist the antiphysicist's inference from an epistemic gap to an

ontological gap by showing how the relevant epistemic gap may exist even if physicalism is true. In the antiphysicalist's arguments, the relevant epistemic gap (from which an ontological gap is inferred) is characterized in such a way that truth and knowledge are essential. For example, it is crucial to the knowledge argument that Mary gains new factual *knowledge* or, at least, new true beliefs. It is crucial to the conceivability argument that one can conceive beings that lack phenomenal states that one actually has. And it is crucial to the explanatory gap that one has cognitively significant knowledge of the states that we cannot explain. If one characterized these gaps in a way that were neutral on the truth of phenomenal beliefs, the arguments would not get off the ground. So truth-value is essential to the relevant epistemic gaps. If so, then to undercut the inference from these gaps to an ontological gap, the phenomenal concept strategy needs to show how the relevant truth-involving epistemic gaps are consistent with physicalism. The strategies above do not do this, so they do nothing to undercut the inference from the epistemic gap to an ontological gap.

Perhaps proponents could augment their explanation of the narrow epistemic situation with an additional element that explains why the relevant beliefs are true and qualify as knowledge. For example, one might augment it with an explanation (perhaps via a causal theory of reference?) of why phenomenal beliefs refer to physical states and an explanation (perhaps via a reliabilist theory of knowledge?) of why such beliefs constitute knowledge. However, the augmented explanation is now subject to the original dilemma. If such an account applies equally to a zombie (as might be the case for simple causal and reliabilist theories, for example), then it cannot account for the crucial epistemic differences between conscious beings and zombies. And if it does not apply equally to a zombie (if it relies on a notion of acquaintance, for example), then crucial explanatory elements in the account will not be physically explainable.

So I think that none of these strategies gives any support to type-B materialism. Each of them deserves brief discussion in its own right, however. For example, it is worth noting that strategy (1), involving justification, has a further problem, in that it is plausible that Mary's introspective beliefs have a sort of justification that Zombie Mary's corresponding beliefs do not share. One could make this case by appealing to the widely accepted view that conscious experience makes a difference to the justification of our perceptual and introspective beliefs. Or one could make it by

considering the scenario directly: whereas Mary's belief that she is currently conscious and having a color experience is plausibly justified with something approaching Cartesian certainty, there is a strong intuition that Zombie Mary's corresponding belief is not justified to the same extent, if it is justified at all. If so, then a physically explicable account of phenomenal concepts cannot explain even the justificatory status of Mary's phenomenal knowledge.

The second strategy, involving inferential disconnection, does not have this sort of problem, as it is plausible that a zombie's physical and quasi-phenomenal beliefs are no more inferentially connected than a conscious being's beliefs. Here, the main problem is that given above. Whereas the inferential disconnection strategy may physically explain an inferential disconnection between physical and phenomenal *beliefs*, the antiphysicist's crucial epistemic gap involves a disconnection between physical and phenomenal *knowledge*. This strategy does not help to reconcile this crucial epistemic gap with physicalism, so it lends no support to type-B materialism. At best, it shows that zombie-style analogs of phenomenal beliefs (inferentially disconnected from physical beliefs) are compatible with physicalism, but this is something that we knew already.

The most interesting version of strategy (3) is the one that appeals to phenomenal concepts to explain our belief in an epistemic gap (including our belief that Mary gains new knowledge, that zombies are conceivable, and that there is an explanatory gap). For the reasons given above, this strategy cannot help the type-B materialist undermine the *inference* from an epistemic gap to an ontological gap. However, one might think that it helps undermine the premise of that inference by explaining why the belief in such a gap is to be predicted even if no such gap exists. This is an important strategy, but it is one more suited to a type-A materialist than to a type-B materialist. The type-B materialist agrees with the antiphysicist, against the type-A materialist, on the datum that there *is* an epistemic gap (e.g., that zombies are conceivable, that Mary gains new phenomenal knowledge, and that there is an explanatory gap). Given this datum, and given that the inference from an epistemic gap to an ontological gap is unchallenged by this strategy, then the strategy does nothing to support type-B materialism against the antiphysicist.

### **Option 3: Assert That Zombies Share Our Epistemic Situation**

The third reaction is to assert that zombies share our epistemic situation. Where we have beliefs about consciousness, zombies have corresponding beliefs with the same truth-values and the same epistemic status. And where Mary acquires new phenomenal knowledge on seeing red for the first time, Zombie Mary acquires new knowledge of a precisely analogous sort. If this is right, then the crucial features of phenomenal concepts might simultaneously be physically explicable and able to explain our epistemic situation.

Of course, a zombie's crucial beliefs will not be *phenomenal* beliefs, and Zombie Mary's crucial knowledge will not be *phenomenal* knowledge. Zombies have no phenomenal states, so they cannot have true beliefs that attribute phenomenal states to themselves, and they cannot have first-person phenomenal knowledge. Instead, the proponent of this strategy must conceive of zombies as attributing some *other* sort of state to itself. We might think of these states as 'schmenomenal states,' and the corresponding beliefs as 'schmenomenal beliefs.' Schmenomenal states stand to phenomenal states roughly as 'twater,' the superficially identical liquid on Twin Earth, stands to water: schmenomenal states are not phenomenal states, but they play a role in zombies' lives that is analogous to the role that phenomenal states play in ours. In particular, on this proposal, a zombie's schmenomenal beliefs have the same truth-value and epistemic status as a non-zombie's phenomenal beliefs.

One might worry that in a type-B materialist view, schmenomenal states must be the same as phenomenal states, since both are identical to the same underlying physical states. In reply, one can note that the discussion of zombies falls within the scope of a conceivability operator, and the type-B materialist allows that although physical states are identical to phenomenal states, it is at least conceivable that they are not so identical. The zombie scenario will presumably be understood in terms of conceiving that the same physical states are identical to nonphenomenal (schmenomenal) states instead. To avoid this complication, one might also conduct this discussion in terms of a functionally identical silicon zombie, rather than in terms of a physically identical zombie. Then the type-B materialist can simply say that ordinary humans have neural states that are identical to phenomenal states, whereas silicon zombies have silicon states that are identical to schmenomenal states. On the current view, silicon zombies will have schmenomenal knowledge that is epistemically analogous to humans' phenomenal knowledge.

This proposal might be developed in two different ways: either by deflating the phenomenal knowledge of conscious beings or by inflating the corresponding knowledge of zombies. That is, a proponent may argue either that Mary gains *less* new knowledge than I suggested earlier or that Zombie Mary gains *more* new knowledge than I suggested earlier. Earlier, I argued that Mary gains new cognitively significant non-indexical knowledge, whereas Zombie Mary does not. The deflationary strategy proposes that Mary gains no such knowledge; the inflationary strategy proposes that Zombie Mary gains such knowledge, too.

The deflationary strategy will presumably involve the claim that the only new factual knowledge that Mary gains upon seeing red for the first time is *indexical* knowledge. That is, Mary gains knowledge of the form ‘I am in *this* state now,’ where ‘*this* state’ picks out the state that she happens to be in: presumably some sort of neural state. According to this proposal, Zombie Mary gains analogous knowledge, also of the form ‘I am in *this* state now,’ where ‘*this* state’ picks out the state she happens to be in: presumably a neural state or a silicon state. There seems to be no problem in principle with the idea that Zombie Mary could gain indexical knowledge of this sort, at least if a zombie can have knowledge at all. This strategy meshes particularly well with the proposal that phenomenal concepts are a species of indexical concept.

In response, I think there is good reason to accept that Mary gains more than indexical knowledge. I have made this case elsewhere (Chalmers 2003a), so I will just recapitulate it briefly here. First, there is a sense in which indexical knowledge is perspective-dependent, and vanishes from an objective perspective. For me, full objective knowledge is incomplete unless I know that I am David Chalmers, but no one else with full objective knowledge can be ignorant of the fact that I am David Chalmers in this way. The same goes for indexical knowledge of my current time and location: no one with full objective knowledge can be ignorant of this in the way that I can be ignorant. Mary’s indexical knowledge that *this* brain state is such-and-such brain state is of the same sort: that is, no one else with full physical knowledge can be ignorant of this in the way that Mary can be ignorant. But Mary’s phenomenal knowledge of what it is like for her to see a red tomato is not like this. Other beings with full physical knowledge can be ignorant of what it is like for Mary to see a tomato, just as Mary was ignorant before she saw the tomato, regardless of their

perspective or the brain states they happen to be in. This strongly suggests that Mary's phenomenal knowledge is not indexical knowledge.

Second, just as Mary gains nontrivial knowledge that such-and-such is what it is like to see red, where 'such-and-such' corresponds to her deployment of a phenomenal concept, she also gains nontrivial indexical knowledge that *this* state is such-and-such, where '*this* state' corresponds to an indexical concept picking out whatever phenomenal state she happens to be in, and 'such-and-such' again corresponds to her deployment of a phenomenal concept. This knowledge is cognitively significant knowledge that Mary gains upon introspection. But this knowledge involves the deployment of an indexical concept on one side of an identity, and Mary's crucial phenomenal concept on the other side. Again, this strongly suggests that the phenomenal concept is distinct from the indexical concept, and that Mary's cognitively significant knowledge *I am in such-and-such state now* is distinct from her trivial indexical knowledge *I am in this state now*. If so, then Mary gains more than this indexical knowledge, and the deflationary strategy fails.

The inflationary strategy involves the proposal that just as Mary gains cognitively significant non-indexical knowledge involving phenomenal concepts, Zombie Mary gains analogous cognitively significant non-indexical knowledge involving schmenomenal concepts. So where Mary gains significant knowledge of the form *Tomatoes cause such-and-such phenomenal state, I am in such-and-such phenomenal state, and This state is such-and-such phenomenal state*, Zombie Mary gains significant knowledge of the form *Tomatoes cause such-and-such schmenomenal state, I am in such-and-such schmenomenal state, and This state is such-and-such schmenomenal state*. Zombie Mary's new beliefs have the same truth-value, the same epistemic status, and the same epistemic connections as Mary's corresponding beliefs.

Here, the natural response is that this scenario is simply not what we are conceiving when we conceive of a zombie. *Perhaps* it is possible to conceive of a being with another sort of state—call it 'schmonsciousness'—to which it stands in the same sort of epistemic relation that we stand in to consciousness. Schmonsciousness would not be consciousness, but it would be epistemically just as good. It is by no means obvious that a state such as schmonsciousness is conceivable, but it is also not obviously inconceivable. However, when we ordinarily conceive of zombies, we are not conceiving of beings with something analogous to consciousness that is

epistemically just as good. Rather, we are conceiving of beings with nothing epistemically analogous to consciousness at all.

Put differently: when we conceive of zombies, we are not conceiving of beings whose inner life is as rich as ours, but different in character. We are conceiving of beings whose inner life is dramatically poorer than our own. And this difference in inner lives makes for dramatic difference in the richness of our introspective knowledge. Where we have substantial knowledge of our phenomenal inner lives, zombies have no analogous introspective knowledge: there is nothing analogous for them to have introspective knowledge of.

Perhaps a zombie can have a sort of introspective knowledge of some of its states: its beliefs and desires, say, or its representations of external stimuli. But this sort of introspective knowledge is not analogous to our phenomenal introspective knowledge. Rather, it is analogous to our nonphenomenal introspective knowledge. Phenomenology is not all that is available to introspection, and it is not out of the question that zombies could have the sort of nonphenomenal introspective knowledge that we have. But none of this knowledge will have the character of our introspective knowledge of phenomenal states because there is nothing analogous for zombies to introspect.

At this point a proponent might appeal to certain naturalistic theories of the mind: perhaps a functionalist theory of belief, a causal theory of mental content, and/or a reliabilist theory of knowledge. Zombies have the same functional organization as conscious beings and the same reliable causal connections among their physical states, so a proponent could suggest that these theories entail that zombies will have corresponding beliefs with the same epistemic status as ours. It is not obvious that the theories will make this prediction: this depends on whether they are a priori theories that apply to all conceivable scenarios. If they do not, then they do not undermine the conception of zombies whose epistemic status differs from ours. But in any case, to appeal to these theories in this context is to beg the question. Consideration of the Mary situation and related matters gives us good reason to believe that consciousness is relevant to matters such as mental content and epistemic status. It follows that if consciousness is not itself explainable in physical/functional terms, then any entirely physical/functional theory of content or knowledge will be incomplete. If a theory predicts that a nonconscious zombie would have the same sort of introspective knowledge that we do, then this is reason to reject the theory.

The upshot of all this is that the inflationary strategy does not adequately reflect what we are conceiving when we conceive of a zombie. *Perhaps* it is conceivable that a nonconscious duplicate could have some analogous state, schmonsciousness, of which they have analogous introspective knowledge. But it is also conceivable that a nonconscious duplicate would have no such analogous introspective knowledge. And this latter conceivability claim is all that the argument against the phenomenal concept strategy needs.

#### **Option 4: Reject the Link between Conceivability and Explanation**

The fourth possible reaction for proponents of the phenomenal concept strategy is to deny the connection between conceivability and explanation. Such proponents might allow that  $P \& \sim C$  is conceivable, but hold that nevertheless,  $P$  explains  $C$ . Or they might allow that  $C \& \sim E$  is conceivable, but hold that nevertheless,  $C$  explains  $E$ .

Of course, everyone should allow that there are *some* sorts of explanation such that explaining  $B$  in terms of  $A$  is consistent with the conceivability of  $B$  without  $A$ . For causal explanation, for example, this is precisely what one expects. The crucial claim is that there is a *sort* of explanation that is tied to conceivability in this way, and that this sort of explanation is relevant to the explanatory gap. This is the sort of micro-macro explanation that I earlier called transparent explanation: explanation that makes transparent why relevant high-level truths obtain, given that low-level truths obtain. If it is conceivable that the low-level truths obtain without the high-level obtaining, the explanation will not be transparent in the relevant way. Instead, one will need to appeal to substantive further principles to bridge the divide between the low-level and high-level domain. It is just this sort of transparent explanation that is absent in the original explanatory gap.

An opponent may deny that this sort of transparent explanation is required for a good reductive explanation or that it is present in typical reductive explanations. Or he may at least deny this for a notion of transparent explanation that is strongly tied to conceivability. For example, Block and Stalnaker (2001), Levine (2001), and Yablo (2002) all argue that typical cases of micro-macro explanation—the explanation of water in terms of  $H_2O$ , for example—are not associated with an a priori entailment of macro truths by micro truths. If they are right about this, then insofar as the notion of transparent explanation is tied to a priori entailment, it is not required for ordinary micro-macro explanation. But I have argued elsewhere (Chalmers and Jackson 2001)

that they are not right about this: even in cases such as the relation between microphysics and water, there is a sort of associated a priori entailment, and this sort of entailment is crucial for a good reductive explanation.

It is also worth noting that even if these theorists are right, this will at best undermine a link between one sort of conceivability and explanation. As before, let us say that  $S$  is *negatively* conceivable when the truth of  $S$  cannot be ruled out a priori. Then the claim that  $A$  entails  $B$  a priori is equivalent to the claim that  $A \& \sim B$  is negatively conceivable. If these theorists are right, then even ‘zombie- $H_2O$ ’ (Levine’s [2001] term for a microphysically identical substance that is not water) will be negatively conceivable, so that ordinary micro-macro explanation of  $B$  by  $A$  cannot require that  $A \& \sim B$  is negatively conceivable. However, Levine himself notes that there is a different sort of ‘thick’ conceivability such that zombies are conceivable in this sense and zombie- $H_2O$  is not, and he notes that this sort of conceivability is tied to explanation:  $A \& \sim B$  is thickly conceivable if and only if there is an explanatory gap between  $A$  and  $B$ . If so, we can use this sort of thick conceivability in the previous arguments.

I think that Levine’s thick conceivability corresponds closely to what I earlier called *positive* conceivability, which requires a clear and distinct positive conception of a situation that one is imagining. Positive conceivability is arguably the central philosophical notion of conceivability. And it is highly plausible that in cases of ordinary reductive explanations of  $B$  by  $A$ ,  $A \& \sim B$  is not positively conceivable: we can form a positive conception of a zombie in a way that we cannot form a positive conception of zombie- $H_2O$ . Furthermore, this positive conceivability seems to be particularly strongly associated with the sense of apparent contingency that goes along with the explanatory gap. So it remains plausible that for the sort of explanation that is relevant here, positive conceivability of  $A \& \sim B$  entails an explanatory gap between  $A$  and  $B$ .

An opponent may insist more strongly that no sort of conceivability is tied in this way to micro-macro explanation. She may hold that this sort of explanation simply requires a relevant correlation or a relevant identity between the low-level and high-level domains, whose existence does not require any strong conceptual connection between low-level truths and high-level truths. I think that this gets the character of micro-macro explanation wrong, by failing to account for the sense of transparency in a good micro-macro explanation. But in any case, an opponent of this sort is unlikely

to be too worried by the explanatory gap in the first place. If this sort of move works to dissolve the explanatory gap between physical processes and phenomenal concepts, say, then it will work equally well to dissolve the original explanatory gap between physical processes and consciousness. If so, then once again the phenomenal concept strategy is rendered redundant in explaining the explanatory gap.

Of course, such a theorist may still appeal to the phenomenal concept strategy to explain the remaining epistemic gaps (such as the conceivability of zombies) in their own right, independent of any connection to the explanatory gap. Here the general idea will be as before: there is no valid inference from these epistemic gaps to an ontological gap because the existence of these epistemic gaps is compatible with physicalism. But as before, an opponent will question the strategy on the grounds that there is as much of an epistemic gap between physical processes and phenomenal concepts (as characterized by the proponent's account), or between phenomenal concepts and our epistemic situation, as there was between physical processes and consciousness. To respond, the opponent must either deny this epistemic gap (which will raise all the previous issues) or give independent reasons to think that the epistemic gap is compatible with physicalism (which will render the phenomenal concept strategy redundant). Either way, the theoretical landscape will be much as before.

(Strictly speaking, there may be one version of this strategy on which the theoretical landscape will differ. A proponent might appeal to phenomenal concepts solely to explain Mary's new knowledge, without using it to explain either the conceivability of zombies or the explanatory gap. If so, then the conceivability of zombies who do not satisfy this account of phenomenal concepts will not raise the usual regress worry. To avoid a residual epistemic gap with the same character as the original epistemic gap, the proponent would simply need to make the case that Mary could know all about the relevant structural features of phenomenal concepts from inside her black-and-white room. Of course, this proponent will then need some other means to deal with the conceivability argument, and with the explanatory gaps posed both by consciousness and by the account of phenomenal concepts.)

In any case, I think that many of the central points of this paper can also be made directly in terms of explanation, without proceeding first through conceivability. The analysis in terms of conceivability is useful in providing a tool for fine-grained analyses and arguments, and to get a sense of the options in the theoretical landscape.

But with these options laid out, one can also make the case directly that any given account of phenomenal concepts will generate either an explanatory gap between physical processes and phenomenal concepts, or between phenomenal concepts and our epistemic situation. I will make this sort of case in the next section.

## 5 Applications

I will now look at some specific accounts of phenomenal concepts in light of the preceding discussion. If what has gone before is correct, then any fully specific account of phenomenal concepts will fall into one of two classes. It will either be a ‘thick’ account, in which the relevant features of phenomenal concepts are not physically explainable (although they may explain our epistemic situation), or it will be a ‘thin’ account, in which the relevant features of phenomenal concepts do not explain our epistemic situation (although they may be physically explainable).

I have already discussed the indexical account of phenomenal concepts, (of Ismael, O’Dea, Perry, and others) under the third reaction above. For the reasons given there, I think that this account is clearly a thin account; for example, it does not adequately explain the character of Mary’s new cognitively significant knowledge. So there is reason to believe that phenomenal concepts are not indexical concepts.

I have also discussed the dual-conceptual-role account (of Nagel, Hill, McLaughlin, and others), under the second reaction above. If this account is understood in wholly functional terms, involving the distinctness in functional role of certain representations in the brain, then it is clearly a thin account. For reasons discussed earlier, this account may help to explain an inferential disconnection between physical and phenomenal beliefs, but it cannot explain the character of phenomenal *knowledge*. Perhaps the account could be supplemented by some further element to explain this character (for example, postulating a special faculty of sympathetic knowledge), but then the original dilemma will arise once again for the new account.

The ‘quotational’ account (of Block, Chalmers, Papineau, and others) might be understood either as a thin or a thick account, depending on how it is specified. One may understand this either in a ‘bottom-up’ way, in which we start with purely physical/functional materials and make no assumptions about consciousness, or in a

‘top-down’ way, in which we build consciousness into the account from the start. I will examine each of these versions in turn.

The bottom-up version of the quotational account is specified in purely physical/functional terms, without building any assumptions about consciousness. The basic idea will be that there are some neural states  $N$  (those that correspond to phenomenal states, though we will not assume that) that can come to be embedded in more complex neural representations by a sort of ‘quotation’ process, which allows the original state to be incorporated as a constituent. Perhaps this will go along with some sort of demonstrative reference to the original neural state, so that the complex state has the form ‘That state:  $N$ .’ Of course, it is not obvious that one can explain any sort of demonstrative reference in physical/functional terms, but I will leave that point aside.

At this point, we can think of the account as an engineer might. If we designed a system to meet the specifications, what sort of results would we expect? In particular, what sort of knowledge of state  $N$  would one expect? I think the answer is reasonably clear. One would expect a sort of indexical knowledge of the state, of the form ‘I am in this state now.’ But one would not expect any sort of cognitively significant knowledge of the state’s intrinsic character. To see this, note that one might design an identical system where state  $N$  is replaced by a different state  $M$  (perhaps another neural state, or a silicon state), with different intrinsic properties. From a bottom-up perspective, we would not expect this change to affect the epistemic situation of the subject in the slightest. States  $N$  and  $M$  may make a difference to the subject’s knowledge by virtue of their functional role, but from an engineering perspective there is no reason to think that the subject has access to their intrinsic character.

So the bottom-up version of the quotational account is best understood as a thin account of phenomenal concepts. It may ground a sort of indexical or demonstrative knowledge of neural states, but it cannot ground the sort of significant non-indexical knowledge of internal states that Mary gains on leaving her black-and-white room. In this respect, the bottom-up version of the quotational account seems to be no better off than the indexical account.

On the top-down version of the quotational account, we build consciousness into the account from the start. In particular, we assume that our initial state  $Q$  is a *phenomenal* state. (It does not matter to what follows whether we assume in addition that  $Q$  is or is not a neural state, or whether we stay silent on the matter.) We then

stipulate a sort of concept-forming process that incorporates phenomenal states as constituents. Perhaps this process will involve a sort of demonstrative reference to the original phenomenal state, so the resulting concept has the form ‘That state: *Q*.’ What sort of results will we then expect?

We *might* then reasonably expect the subject to have some sort of cognitively significant knowledge of the character of *Q*. In general, when we make demonstrative reference to phenomenal states, we can have cognitively significant knowledge of their character. We could also imagine a functionally identical subject who, in place of *Q*, has a different phenomenal state *R*. In this case, one might expect the substitution to affect the subject’s epistemic situation: the new subject will have cognitively significant non-indexical knowledge that it is in phenomenal state *R*, which is quite different from the first subject’s knowledge.

This top-down version of the quotational account is quite clearly a thick account of phenomenal concepts. By building phenomenal states into the account, it has the capacity to help explain features of our epistemic situation that the bottom-up account cannot. But precisely because the account builds in phenomenal states from the start, it cannot be transparently explained in physical terms. This version of the account presupposes the special epistemic features of phenomenal states rather than explaining them.

(Papineau’s version of the quotational account appears to be a thin version. Papineau discusses a silicon zombie (2002: 125–27) and suggests that it will have semantic and epistemic features analogous to those of a conscious being. His account seems to point in a direction in which the relevant phenomenal knowledge is all a kind of indexical or demonstrative knowledge, although he does not explicitly make this claim or address the objections to it. By contrast, my own version of this sort of account [Chalmers 2003a] is certainly intended as a thick account.)

The recognitional-concept account (of Loar, Carruthers, Tye, and others) can be handled in a similar way. If we understand the concept in a bottom-up way, involving recognitional processes triggered by neural states, what sort of knowledge will we expect? Here, I think we would once again expect a sort of indexical or demonstrative knowledge of the neural states in question, without any cognitively significant knowledge of their intrinsic character. Once again, we would expect that substituting one neural state for another would make no significant difference to a subject’s epistemic situation. So this version of the account can be understood as a thin account.

On the other hand, if we understand the account in a top-down way, as involving recognitional concepts triggered by phenomenal states, then one might well expect it to lead to significant knowledge of the character of these states. It is plausible that merely having a phenomenal state enables us to have a conception of its character, by which we can recognize it (or at least, recognize states reasonably similar to it) when it reoccurs, and such that substituting a different phenomenal state will make a difference to our epistemic situation. This top-down account might well capture something about the difference between a conscious being's epistemic situation and a zombie's situation. But again, this account presupposes the existence of consciousness, along with some of its special epistemic features, so the account is clearly a thick account.

(Loar's own account appears to be a thick account. His discussion of phenomenal concepts presupposes both the existence of consciousness and some of its special epistemic features. In particular, his account crucially relies on the thesis that phenomenal states are presented to us under noncontingent modes of presentation, thus enabling significant knowledge of their character. He defends this assumption by saying that the nonphysicalist accepts the thesis, so the physicalist is entitled to it as well. But of course, the thesis poses a special explanatory burden on the physicalist. How can a neural state of a physical system be presented to a subject under a noncontingent mode of presentation, thus enabling significant knowledge of its character? Loar does not say.)

What about Sturgeon's account of phenomenal concepts, according to which phenomenal states constitute their own canonical evidence? I think that this is probably best understood as a thick account. From a bottom-up perspective, would we expect neural states to constitute their own canonical evidence? When zombies deploy their analogs of phenomenal concepts, do they have analogous states that constitute their own canonical evidence? The answer is not entirely obvious, but on the face of it, the more plausible answer is no. If so, then Sturgeon's account can be seen as a thick account, one that rests on a special epistemic feature of phenomenal concepts.

What do the thick accounts of phenomenal concepts have in common? All of them implicitly or explicitly build in special epistemic features of phenomenal concepts: the idea that phenomenal states present themselves to subjects in especially direct ways, or the idea that simply having a phenomenal state enables a certain sort of knowledge of the state, or the idea that the state itself constitutes evidence for the

state. If we build in such features, then we may be able to explain many aspects of our distinctive epistemic situation with respect to consciousness. But the cost is that such features themselves pose an explanatory problem. If these features are powerful enough to distinguish our epistemic situation from that of a zombie, then they will themselves pose as much of an explanatory gap as does consciousness itself.

If one rejects physicalism, there is no obvious problem in accommodating these epistemic features of consciousness. Dualists sometimes postulate an epistemic relation of acquaintance that holds between subjects and their phenomenal states, and that affords knowledge of these states. If necessary, a dualist can simply take this relation as primitive: the dualist is already committed to positing primitive mental features, and this relation may reasonably be taken to be part of the primitive structure of consciousness. However, this move is not available to a physicalist. The physicalist must either explain the features or accept a further explanatory gap.

Our examination of a specific account of phenomenal concept reaches a conclusion very much compatible with that of Levine (2006). It appears that such accounts either build in strong epistemic relations such as acquaintance, which themselves pose problems for physical explanation, or they build in weak epistemic relations such as indexical or demonstrative reference, in which case they cannot explain our epistemic situation with regard to consciousness. The arguments earlier in the chapter suggest that this is not a mere accident of these specific accounts that a better account may evade. Any account of phenomenal concepts can be expected to have one problem or the other. For this reason, the phenomenal concept strategy cannot reconcile ontological physicalism with the explanatory gap.

## References

- Aydede, M., and Güzeldere, G. (2005). Cognitive Architecture, Concepts, and Introspection: An Information-Theoretic Solution to the Problem of Phenomenal Consciousness. *Nous* 39: 197–255.
- Balog, K. (1999). Conceivability, Possibility, and the Mind-Body Problem. *Philosophical Review* 108: 497–528.
- Block, N., and Stalnaker, R. (1999). Conceptual Analysis, Dualism, and the Explanatory Gap. *Philosophical Review* 108: 1–46.
- Block, N. (2006). Max Black's Objection to Mind-Body Identity. In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, ed. T. Alter and S. Walter. Oxford: Oxford University Press.

- Carruthers, P. (2004). Phenomenal Concepts and Higher-Order Experiences. *Philosophy and Phenomenological Research* 68: 316–36.
- Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- Chalmers, D. J. (1999). Materialism and the Metaphysics of Modality. *Philosophy and Phenomenological Research* 59: 473–96.
- Chalmers, D. J. (2002). Does Conceivability Entail Possibility? In *Conceivability and Possibility*, ed. T. Gendler and J. Hawthorne. Oxford: Oxford University Press: 145–200.
- Chalmers, D. J. (2003a). The Content and Epistemology of Phenomenal Belief. In *Consciousness: New Philosophical Perspectives*, ed. Q. Smith and A. Jokic: 220–72. Oxford: Oxford University Press.
- Chalmers, D. J. (2003b). Consciousness and Its Place in Nature. In *The Blackwell Guide to the Philosophy of Mind*, ed. P. Stich and T. Warfield. Oxford: Blackwell. Reprinted in *The Philosophy of Mind: Classical and Contemporary Readings*, ed. D. Chalmers: 247–72. New York: Oxford University Press, 2002.
- Chalmers, D. J., and Jackson, F. (2001). Conceptual Analysis and Reductive Explanation. *Philosophical Review* 110: 315–61.
- Dennett, D. (1991). *Consciousness Explained*. Boston: Little, Brown.
- Dennett, D. (2006). What RoboMary Knows. In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, ed. T. Alter and S. Walter. Oxford: Oxford University Press.
- Hawthorne, J. (2006) Dancing Qualia and Direct Reference. In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, ed. T. Alter and S. Walter. Oxford: Oxford University Press.
- Hill, C. S. (1997). Imaginability, Conceivability, Possibility, and the Mind-Body Problem. *Philosophical Studies* 87: 61–85.
- 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.
- Ismael, J. (1999). Science and the Phenomenal. *Philosophy of Science* 66: 351–69.
- Jackson, F. (1998). *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Clarendon Press.
- Jackson, F. (2003). Mind and Illusion. In *Minds and Persons: Royal Institute of Philosophy Supplement* 53, ed. A. O'Hear: 251–71. Cambridge: Cambridge University Press.
- Levine, J. (2006). What is a Phenomenal Concept? In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, ed. T. Alter and S. Walter. Oxford: Oxford University Press.
- Levine, J. (1983). Materialism and Qualia: The Explanatory Gap. *Pacific Philosophical Quarterly* 64: 354–61.
- Levine, J. (2001). *Purple Haze: The Puzzle of Consciousness*. Oxford: Oxford University Press.

- Levine, J. (2006) Phenomenal Concepts and the Materialist Constraint. In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, ed. T. Alter and S. Walter. Oxford: Oxford University Press.
- Loar, B. (1990/97). Phenomenal States. *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.
- Nagel, T. (1974). What Is It Like to Be a Bat? *Philosophical Review* 4: 435–50.
- O’Dea, J. (2002). The Indexical Nature of Sensory Concepts. *Philosophical Papers* 31: 169–81.
- Papineau, D. (2002). *Thinking about Consciousness*. Oxford: Oxford University Press.
- Papineau (2006). Phenomenal Concepts and the Materialist Constraint. In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, ed. T. Alter and S. Walter. Oxford: Oxford University Press
- Perry, J. (1979). The Problem of the Essential Indexical. *Nous* 13: 3–21.
- Perry, J. (2001). *Knowledge, Possibility, and Consciousness*. Cambridge: MIT Press.
- Stoljar, D. (2005). Physicalism and Phenomenal Concepts. *Mind and Language* 20: 469-94.
- Sturgeon, S. (1994). The Epistemic Basis of Subjectivity. *Journal of Philosophy* 91: 221–35.
- Tye, M. (2003). A Theory of Phenomenal Concepts. In *Minds and Persons*, ed. A. O’Hear: 91–105. Cambridge: Cambridge University Press.
- White, S. (1986). Curse of the Qualia. *Synthese* 68: 333–68.
- Yablo, S. (2002). Coulda, Woulda, Shoulda. In *Conceivability and Possibility*, ed. T. Gendler and J. Hawthorne. Oxford: Oxford University Press: 441–92.