David Papineau

Phenomenal and Perceptual Concepts

Contents

---

1 Introduction

Phenomenal concepts are common coin among nearly all contemporary philosophers working on consciousness. They are recognized both by ontological dualists, who take them to refer to distinctive non-material (phenomenal) properties, and by the majority of contemporary materialists, who respond that phenomenal concepts are distinctive only at a conceptual level, and refer to nothing except material properties that can also be referred to using non-phenomenal material concepts.

In speaking of the majority of contemporary materialists, I have in mind the school of thought that David Chalmers (2003a) has dubbed 'type-B physicalism'. In effect, type-B physicalism is a concession to the classic anti-materialist arguments of Frank Jackson (1986) and Saul Kripke (1980). Older (type-A) physicalists took all concepts of conscious states to be functional concepts—that is, concepts that referred by association with causal roles. Because of this, they denied the initial premises of Jackson's and Kripke's arguments. In response to Jackson's 'Mary' argument, they argued that any functional concepts of conscious states would have been be available to Mary before she left her room, and so that there was no sense in which she acquired any new knowledge of 'what it is like' to see something as red. Relatedly, in response to Kripke's argument, they argued that it was inconceivable, and so obviously impossible, that a being could be fully physically identical to us yet lack consciousness. However, these responses to Jackson and Kripke are now widely agreed to be unsatisfactory. It seems clear that the pre-emergence Mary does lack some concepts of colour experiences, and moreover that zombies are at least conceivable. By recognizing phenomenal concepts, type-B physicalists aim to concede this much to Jackson and Kripke. At the same time, they argue that, once we do recognize phenomenal concepts, then we can see that the

subsequent stages of Jackson's and Kripke's arguments do not provide a valid route to ontologically dualist conclusions. (Cf. Loar 1990, Papineau 2002 chs 2 and 3.)

What is the nature of phenomenal concepts? Here there is far less consensus. Among those who trade in phenomenal concepts, some take them to be sui generis (Tye, 2003, Chalmers, 2003b), while others have variously likened them to recognitional concepts (Loar, 1990), to demonstratives (Horgan 1984, Papineau 1993a, Perry 2001), or to quotational terms (Papineau 2002, Balog forthcoming).

In my Thinking about Consciousness (2002), I developed a 'quotational-indexical' of phenomenal concepts account on roughly the following lines. To have a phenomenal concept of some experience, you must be able introspectively to focus on it when you have it, and to recreate it imaginatively at other times; given these abilities, you can then form terms with the structure *the experience: —*, where the gap is filled either by a current experience, or by an imaginative recreation of an experience; these terms then comprise a distinctive way of referring to the experience at issue.

In the book I argued that this account of phenomenal concepts not only allows a satisfactory materialist response to Jackson's and Kripke's arguments, but also explains why dualism seems so compelling even to those unfamiliar with those arguments. According to my analysis, we all experience a basic 'intuition of mind-brain distinctness', which is prior to any philosophical investigation (and indeed which lends a spurious plausibility to the standard anti-materialist arguments, by independently adding credibility to their conclusions). However, once we understand the structure of phenomenal concepts, I argued, we can see how this intuition arises, and why it provides no real reason to doubt materialism.

In this paper, I want to return to the topic of phenomenal concepts. It now seems to me that the treatment in Thinking about Consciousness was inadequate in various respects. Here I want to try to improve on that account. In particular, I shall develop an extended comparison of phenomenal concepts with what I shall call 'perceptual concepts', hoping thereby to throw the nature of phenomenal concepts into clearer focus.

While the position I shall develop in this paper will involve some significant revisions of the claims made in my book, I think that the main arguments in the book are robust with respect to these revisions. In particular, the responses to Jackson and Kripke will stand pretty much as before, and an explanation of the persistent 'intuition of distinctness' will continue to be available.

The revised account will also enable me to deal with a common worry about phenomenal concepts[1[1]]. Suppose Mary has come out of her room, seen a red rose, and as a result acquired a phenomenal concept of the experience of seeing something red (though she mightn't yet know that this experience is conventionally so-called). On most account of phenomenal concepts, including the one developed in my book, any exercise of this phenomenal concept will demand the presence of the experience itself or an imaginatively recreated exemplar thereof. The trouble, however, is that it seems quite possible for Mary to think truly, using her new phenomenal concept, *I am not now having that experience (nor recreating it in my imagination)*—but this would be ruled out if any exercise of her phenomenal concept did indeed depend on the presence of the experience or its imaginative recreation. The revised account of phenomenal concepts to be developed here will not require this, and so will be able to explain Mary's problematic thought.

[1[1]] This worry has long been pressed on me by my London colleagues Tim Crane and Scott Sturgeon, and is developed in Crane's contribution to a forthcoming symposium on Thinking about Consciousness in Philosophy and Phenomenological Research. The same objection is attributed to Kirk Ludwig in a typescript by Ned Block entitled 'Max Black's Objection to Mind-Body Identity'.

The rest of this article contains four sections. The next two analyse perceptual and phenomenal concepts respectively. The penultimate section checks that my revised account of phenomenal concepts will still serve to block the standard arguments for dualism. The final section defends my position against a recent argument by David Chalmers against the whole Type-B strategy of defending physicalism by appeal to phenomenal concepts.

2 Perceptual Concepts

2.1 Perceptual Concepts are not Demonstrative

Let me turn away from phenomenal concepts for a while, and instead consider perceptual concepts. Getting clear about perceptual concepts will stand us in good stead when we turn to the closely related category of phenomenal concepts.

We can start with this kind of case. You see a bird at the bottom of your garden. You look at it closely, and at the same time think *I haven't seen THAT in here before*. Later on you can recall the bird in visual imagination, perhaps thinking *I wonder if THAT was a migrant*. In addition, on further perceptual encounters with birds, you sometimes take some bird to be the same bird again, and can again form further thoughts about it, such as *THAT bird has a pleasant song*. (Let me leave it open for the moment whether you are thinking of a particular bird or a type of bird; I shall return to this shortly.)

In examples like this, I shall say that subjects are exercising perceptual concepts. Perceptual concepts allow subjects to think about perceptible entities. Such concepts are formed when subjects initially perceive the relevant entities, and are re-activated by latter perceptual encounters. Subjects can also use these concepts to think imaginatively about those entities even when they are not present.

Now, it is tempting to view concepts of this kind as 'demonstrative'. For one thing, it is natural to express these concepts using demonstrative words, as the above examples show ('. . . *THAT* . . .'). Moreover, uses of perceptual concepts involve a kind of perceptual attention or imaginative focus, and this can seem analogous to the overt pointing or other indicative acts that accompany the use of verbal demonstratives.

However, I think it is quite wrong to classify perceptual concepts as demonstratives. If anything is definitive of demonstrative terms, it is surely that they display some species of characterlikeness. By this I mean that the referential value of the term is context-dependent—the selfsame term will refer to different items in different contexts. However, there seems nothing characterlike about the kind of perceptual concept illustrated in the above examples. Whenever it is exercised, your perceptual concept refers to the same bird. When you use the concept in question, you don't refer to one bird on the first encounter, yet some possibly different bird when later encountering or visually imagining it. Your concept picks out the same bird whenever it is exercised.

It is possible to be distracted from this basic point by failing to distinguish clearly between perceptual concepts and their linguistic expression. If I want to express some perceptual thought in language, then there may be no alternative to the use of demonstrative words. In order to convey my thought to you, I may well say 'That bird has a present song', while indicating some nearby bird. And I agree that the words here used—'that bird'—are demonstrative, in that they will refer to different birds in different contexts of use. But this does not mean that my concept itself is demonstrative. As I have just urged, my concept itself will refer to the same bird whenever it is exercised.

The reason we often resort to demonstrative words to convey thoughts involving non-demonstrative perceptual concepts is simply that there is often no publicly established linguistic term to express our concept. In such cases, we can nevertheless often get our ideas across by demonstratively indicating some instance of what we are thinking about. Of course, this possibility assumes that some such instance is available to be demonstrated—if there isn't, then we may simply find ourselves unable to express what we are thinking to an audience.

By insisting that perceptual concepts are not demonstrative, even if the words used to express them are, I do not necessarily want to exclude characterlikeness from every aspect of the mental realm. Millikan (1990) has argued that mental indexicality plays no ineliminable role in the explanation of action, against Perry (1979) and much current orthodoxy, and I find her case on this particular point persuasive. Even so, I am open to the possibility that primitive mental demonstratives may play some role in pre-conceptual attention (*what was THAT?)* and also to the possibility that there may be characterlike mental terms constructed with the help of predicates (*I'm frightened of THAT DOG-ie the dog in the corner of the room)*.[2[2]] In both these kinds of case I allow that the capitalised expressions may express genuinely characterlike mental terms—that is, repeatable mental terms that have different referents on different occasions of use. My claim in this section has only been that perceptual concepts in particular are not characterlike in this sense, but carry the same referent with them from one occasion of use to another.

2.2 Perceptual Concepts as Stored Templates

I take perceptual concepts to involve a phylogenetically old mode of thought that is common to both humans and animals. We can helpfully think of perceptual concepts as involving stored sensory templates. These templates will be set up on initial encounters with the relevant referents. They will then be reactivated on later perceptual encounters, via matches between incoming stimuli and stored template—perhaps the incoming stimuli can be thought of as 'resonating' with the stored pattern and thereby being amplified. Such stored templates can also be activated autonomously even in the absence of any such incoming stimuli—these will then constitute 'imaginative' exercises of perceptual concepts.[3[3]]

The function of the templates is to accumulate information about the relevant referents, and thereby guide the subject's future interactions with them. We can suppose that various items of information about the referent will become attached to the template as a result of the subject's experience. When the perceptual concept is activated, these items of information will be activated too. They may include features of the referent displayed in previous encounters. Or they may simply comprise behavioural information, in the form of practical knowledge that certain responses are appropriate to the presence of the referent. When the referent is re-encountered, the subject will thus not only perceive it as presently located at a certain position in egocentric space, but will also take it to possess certain features that were manifested in previous encounters, but may not yet be manifest in the re-encounter. Imaginative exercises of perceptual concepts may further allow subjects to process information about the referent even when it is not present.

Note how this function of carrying information from one use to another highlights the distinction between perceptual concepts and demonstratives. Demonstrative terms do not so carry a body of information with them, for the obvious reason that they refer to different entities on different occasions of use. Information about an entity referred to by a demonstrative on one occasion will not in general apply to whatever entity happens to be the referent the next time the demonstrative is used. By contrast, perceptual concepts are suited

---

[2[2]] I will take no stand on whether or not such 'mixed demonstratives' are equivalent to definite descriptions.

[3[3]] Cf. Prinz 2002 especially chs 6 and 7.

to serve as repositories of information precisely because they refer to the same thing whenever they are exercised.

2.3 Perceptual Semantics

I have said that perceptual concepts refer to perceptible entities. However, what exactly determines this relation between perceptual concepts, conceived as stored sensory templates, and their referents? In particular, what determines whether such a concept refers to a type or a token? I suggested earlier that you might look at a bird, form some stored sensory template, and then use it to think either about that particular bird or about its species. But what decides between these two referents? At first pass, it seems that just the same sensory template might be pressed into either service.

Some philosophers think of perceptual concepts as 'recognitional concepts' (Loar 1990). This terminology suggests that perceptual concepts should be viewed as referring to whichever entities their possessors would recognize as satisfying them. A stored sensory template will refer to just that entity which will activate it when encountered. If none but some particular bird will activate some template, then that particular bird is the referent. If any member of a bird species will activate a template, then the species is the referent.

This recognitional account would serve adequately for most of the further purposes of this paper. But in fact it is a highly unsatisfactory account of perceptual reference. Now I have raised this topic, I would like to digress briefly and explain how we can do better.

First, let me briefly point out the flaws in the recognitional account. For a start, it's not clear that recognitional abilities are fine-grained enough to make the referential distinctions we want. Could not two people have just the same sensory template, and so be disposed to recognize just the same instances, and yet one be thinking about a particular bird, and the other about the species? It is not obvious, to say the least, that my inability to discriminate perceptually between the bird in my garden and its conspecifics means that I must be thinking about the whole species rather than my particular bird; nor, conversely, is it obvious that I must be thinking of my bird rather than its species if I mistakenly take some idiosyncratic marking of my bird to be a characteristic of the species. In any case, the equation of referential value with recognitional range faces the familiar problem that it seems to exclude any possibility of misrecognition: if the referent of my perceptual concept is that entity which includes all the items I recognize as satisfying the concept, then there is no room left for me to misapply the concept perceptually. However, this isn't what we want—far from guaranteeing infallibility, perceptual concept possession seems consistent with very limited recognitional abilities.

I think we will do better to approach reference by focusing on the function of perceptual concepts rather than their actual use. As I explained in the last subsection, the point of perceptual concepts is to accumulate information about certain entities and make it available for future encounters. Given this, we can think of the referential value of a perceptual concept as that entity which it is its function to accumulate information about. Give or take a bit, this will depend on two factors: the origin of the perceptual concept, and the kind of information that gets attached to it.

Let me take the second factor first. Note that the kind of information that it is appropriate to carry from one encounter to another will vary, depending on what sort of entity is at issue.[4[4]] For example, if I see that some bird has a missing claw, then I should expect this to hold on other encounters with that particular bird, but not across other encounters with members of that species. By contrast, the information that the bird eats seeds is appropriately carried over

[4[4]] Here I am very much indebted to Ruth Millikan's On Clear and Confused Ideas (2000).

to other members of the species. The point is that different sorts of information are projectible across encounters with different types of entity. If you are thinking about some metal, you can project melting point from one sample to another, but not the shape of the samples. If you are thinking about some species of shellfish, you can project shape, but not size. If you are thinking about individual humans, you can project ability to speak French, but not shirt colour. And so on.

Given this, we can think of the referents of perceptual concepts as determined inter alia by what sort of information the subject is disposed to attach to that concept. If the subject is disposed to attach particular-bird-appropriate information, then the concept refers to a particular bird, while if the subject is disposed to attach bird-species-appropriate information, then reference is to a species. In general, we can suppose that the concept refers to an instance of that kind to which the sort of information accumulated is appropriate.

To make this suggestion more graphic, we might think of the templates corresponding to perceptual concepts as being manufactured with a range of 'slots' ready to be filled by certain items of information. Thus a particular-bird-concept will have slots for bodily injuries and other visible abnormalities; a particular-person-concept will have slots for languages spoken; a metal-concept will have a slot for melting point; and so on. Which slots are present will then determine which kind of entity is at issue.

The actual referent will then generally be whichever instance of that kind was responsible for originating the perceptual concept. As a rule, we can suppose that the purpose of any perceptual concept is to accumulate information about that item (of the relevant kind) that was responsible for its formation. This explains why there is a gap between referential value and recognitional range. I may not be particularly good at recognizing some entity. But if that entity is the source of my concept, then the concept's function is still to accumulate information about it.

Of course, if some perceptual concept comes to be regularly and systematically triggered by some entity other than its original source, and as a result information derived from this new entity comes to eclipse information about the original source, then no doubt the concept should come to be counted as referring to the new entity rather than the original source. But this special case does not undermine the point that a perceptual concept will normally refer to its origin, rather than to whichever entities we happen to recognize as fitting it.

Now that I have explained how it is possible for perceptual concepts to refer differentially to both particular tokens and general types, some readers might be wondering how things will work with subjects who have perceptual concepts both for some token and its type—for example, suppose that I have a perceptual concept both for some particular parrot and for its species. To deal with cases like this, we need to think of perceptual concepts as forming structured hierarchies. When someone has perceptual concepts both for a token and its type, the former will add perceptual detail to the latter, so to speak. The same will also apply when subjects have concepts of some determinate (*mallard*, say) of some determinable type (*duck*). In line with this, when some more detailed perceptual concept is activated, then so will any more general perceptual concepts which covers it, but not vice versa. Since any items of information that attach to such more general concepts will also apply to the more specific instances, this will work as it should, giving us any generic information about the case at hand along with any case-specific information.

Before proceeding, let me make it clear how I am thinking about the relationship between perceptual concepts and conscious perceptual experience. I want to equate conscious perceptual experiences with the activation of perceptual concepts, due either to exogenous stimulation or to endogenous imagination. This does not necessarily mean that any activations of perceptual concepts are conscious. There may be states that fit the

specifications of perceptual concepts given so far, but whose activations are too low-level to constitute conscious states—early stages of visual processing, say. My assumption will only be that there is some range of perceptual concepts whose activations constitute conscious perceptual experience.[5[5]] In line with this, I shall restrict the term 'perceptual experience' to these cases—that is, I shall use 'experience' is a way that implies consciousness. In addition, I shall also assume that the phenomenology of these states goes with the sensory templates involved, independently of what information the subject attaches to those templates or is or is disposed to attach to them. (So if you and I use the same sensory pattern to think about a particular bird and a bird species respectively, the what-it's-likeness of the resulting experiences will nevertheless be the same[6[6]].)

2.4 Perceptually Derived Concepts

The discussion so far has assumed that thoughts involving perceptual concepts will require the subject actually to be perceiving or imagining. In order for the perceptual concept to be deployed, the relevant stored template needs either to be activated by a match with incoming stimuli, or to be autonomously activated in imagination.[7[7]]

However, now consider this kind of case. You have previously visually encountered some entity—a particular bird, let us suppose—and have formed a perceptual concept of that bird. As before, you exercise this perceptual concept when you perceive further birds as the same bird again, or when you imagine the bird. However, now suppose that you think about the bird when it is not present, and without imaginatively recreating your earlier perception. You simply think 'That bird must nest near here', say, without any accompanying perceptual or imaginative act. I take it that such thoughts are possible.[8[8]] Having earlier established perceptual contact with some entity, you can subsequently refer to it without the active help of either perception or imagination. I shall say that such references are made via perceptually derived concepts.

Here is one way to think about this. Initially your information about some referent was attached to a sensory template. But now you have further created some non-perceptual 'file', in which your store of information about that entity is now also housed. This now enables you to think about the entity even when you are not perceiving or imagining it. When you later activate the file, you automatically refer to the same entity as was referred to when the file was originally created.[9[9]]

[5[5]] It should be noted that my assumption that phenomenology goes with categorization, rather than with basic physical object representation, is denied by Jackendorf 1987 and Prinz 2000. However, I find their arguments uncompelling.

[6[6]] It is consistent with this that there should be phenomenological differences when different patterns are involved within or across subjects when they identify an individual bird (*Jemima*), or its type (*mallard*), or indeed both (*Jemima and a mallard*).

[7[7]] Couldn't you have a standing thought 'that bird is a female', say, even when you aren't actively rehearsing the thought—and won't this imply a sense in which you can think about the bird even when not perceiving or imagining it? Maybe so. But I am interested here in the involvement of concepts in occurrent thoughts, not in standing ones.

[8[8]] Let me be more specific about the possibility at issue here. The idea is not that you might refer non-perceptually using familiar indexical constructions, as in 'the bird I saw in my garden yesterday morning'. Rather, the non-perceptual reference is supposed to derive more immediately from the prior perceptual contact, in such a way as to remain possible even if you have forgotten where and when you previously perceived the bird.

[9[9]] Those who think in terms of 'recognitional concepts' may be inclined to argue that perceptually derived concepts, along with the perceptual concepts they derive from, require an ability to recognize referents perceptually. Against this, I have already argued that even ordinary perceptual concepts do not require their possessors to be any good at recognizing referents. With perceptually derived concepts I would say that recognitional powers can atrophy still further, even to the extent where there

Perhaps the ability to create such non-perceptual files is peculiar to linguistic creatures. This is not to say that any such file must correspond to a term in a public language: you can think non-perceptually about things for which you have no name—for example, you may have no name for the bird that that you think nests nearby. Still, in evolutionary terms it seems likely that the ability to think non-perceptually depended on the emergence of language. In this connection, note that an ability to think about things that you have not perceived, and so cannot perceptually recognize or imagine, must play an essential part in mastery of a public language. For public languages are above all mechanisms that allow those who have first-hand acquaintance with certain items of information to share that information with others—which means that those who receive such information will often need to create non-perceptual 'files' for entities they have never perceived themselves. By contrast, languageless creatures will have no channels through which to acquire information about items beyond their perceptual ambit, and so no need to represent those items non-perceptually. This provides good reason to suppose that the ability to create non-perceptual 'files' arrived only with the emergence of language. If this is right, then only language-using human beings will be able to transcend perceptual concepts proper by constructing what I am calling 'perceptually derived concepts'. Of course, as noted at the beginning of this paragraph, humans will also sometimes use this ability to create non-perceptual 'files' that correspond to no word in a public language. But, still, when they do so, they may well be drawing on an ability that evolved only along with linguistic capacities.

Perhaps there is an issue about counting concepts here. I have distinguished between 'perceptual concepts' and 'perceptually derived concepts'. Do I therefore want to say that a thinker who has constructed a 'perceptually derived concept' from a prior 'perceptual concept' now has <u>two</u> concepts that refer to the same thing? From some perspectives, this might seem like double counting. In particular, it is not clear that the standard Fregean criterion of cognitive significance will tell us that there are two concepts here. After all, if the creation of a 'perceptually derived concept' is simply a matter of housing your store of information in a non-perceptual file, as I put it above, and if any subsequently acquired information about the relevant referent automatically gets attached to both sensory template and non-perceptual tag, then it seems that the subject will always make exactly the same judgements whether using the 'perceptual' or 'perceptually derived' concept, and so fail the Frege test for possession of distinct concepts. And this would suggest that we simply have <u>one</u> concept here, not two, albeit a concept which can be exercised in two ways—perceptually and non-perceptually.[10[10]]

There is no substantial issue here. To the extent that the flow of information between the two ways of thinking is smooth, the Frege test gives us reason to say that there is only one concept. On the other hand, to the extent that there are cognitive operations that distinguish a perceptually derived concept from its originating perceptual concept, there is a rationale for speaking of two concepts, and I shall do so when this is convenient.

<u>3 Phenomenal Concepts</u>

---

is no disposition perceptually to identify <u>anything</u> as the referent of your concept. Suppose that you previously referred to something perceptually. But your stored sensory template has faded, and you can no longer perceptually reidentify new instances when you come across them, and perhaps you can't even perceptually imagine them. It doesn't seem to me that this need stop you being able to rehearse your belief '<u>That</u> bird was female', say, where the underlined phrase still refers to the original referent. <u>Anaphora</u> is perhaps a useful model here. Consider someone who first thinks about some entity perceptually, and then keeps coming back to it in thought, even after the ability to reidentify the entity has faded. It seems natural to suppose that these thoughts will continue to refer to the same entity, even in the absence of continued recognitional abilities.

[10[10]] I would like to thank Dorothy Edgington for drawing my attention to this issue.

3.1 The Quotational-Indexical Model

Let me now turn to phenomenal concepts. My earlier 'quotational-indexical' model, recall, viewed phenomenal concepts as having the structure *the experience: —,* where the gap was filled either an actual perceptual experience or an imaginative recreation thereof. It now seems to me that this 'quotational-indexical' model ran together a good idea with a bad one. The good idea was to relate phenomenal concepts to perceptual concepts. The bad idea was to think that phenomenal concepts, along with perceptual ones, are some kind of 'demonstrative'.

Let me first explain the bad idea. Suppose that perceptual concepts were demonstrative, contrary to the argument of the last section. Then presumably they would be constructions that, on each occasion of use, referred to whichever item in the external environment was somehow salient to the subject. By analogy, if phenomenal concepts worked similarly, then they too would refer to salient items, but now in the 'internal' conscious environment. This thus led me to the idea that phenomenal concepts were somehow akin to the mixed demonstrative construction *that experience*. On this model, phenomenal concepts would employ the same general demonstrative construction (*that*) as is employed by ordinary mixed demonstratives, but the qualifier *experience* would function to direct reference inwards, so to speak, ensuring that some salient element is the conscious realm is picked out. The 'quotational' suggestion then depended on the fact that this demonstrated experience would itself be present in the realm of conscious thought, unlike the non-mental items referred to by most demonstratives. This made it seem natural to view phenomenal concepts as 'quoting' their referents, rather than simply referring to distal items. Linguistic quotation marks, after all, are a species of demonstrative construction: a use of quotation marks will refer to <u>that word</u>, whatever it is, that happens to be made salient by being placed within the quotation marks. Similarly, I thought, phenomenal concepts can usefully be thought of as referring to <u>that experience</u>, whatever it is, that is currently made salient in thought.

However, this now seems to me all wrong. Not only is it motivated by a mistaken view of perceptual concepts, but it runs into awkward objections about the nature of the notion of *experience* used to form the putative construction *that experience*.

There seem two possible models for the concept of *experience* employed here. It might be <u>abstracted</u> from more specific phenomenal concepts (*seeing something red*, *smelling roses*, and so on); alternatively, it could be some kind of <u>theoretical</u> concept, constituted by its role in some theory of experiences. However, neither option seems acceptable.

The obvious objection to the abstraction strategy is that it presupposes such specific phenomenal concepts as *seeing something red*, *smelling roses*, and so on, when it is supposed to explain them. If we are to acquire a generic concept of experience via first thinking phenomenally about more specific experiences, and then abstracting a concept of what they have in common, then it must be possible to think phenomenally about the more specific experiences prior to developing the generic concept. But if thinking phenomenally about the more specific experiences requires us already to have the generic concept, as on the demonstrative account of phenomenal concepts, then we are caught in a circle.

What if our notion of experience is constituted by its role in some theory of experiences (our folk psychological theory perhaps)? Given such a theoretically defined generic concept of experience, there would be no barrier to then combining it with a general-purpose 'that' to form demonstrative concepts of specific experiences. Since the generic concept wouldn't be derived by abstraction from prior phenomenal concepts of specific experiences, there would be no circle in using it to form such specific phenomenal concepts.

This picture may be cogent in principle, but it seems to be belied by the nature of our actual phenomenal concepts. If a generic concept of 'experience' were drawn from something like folk psychological theory, then we could expect it to involve some commitment to the assumption that experiences are internal causes of behaviour. Folk psychology surely conceives of experiences inter alia as internal states with characteristic causes and behavioural effects. But then it would seem to follow that anything demonstrated as *that experience*, where *experience* is the folk psychological concept, must analytically have some behavioural effects. It needn't be analytic which specific behavioural effects *that experience* has—you could know that all experiences have characteristic effects without knowing what specific effects *that experience* has—but still, it would be analytic that *that experience* had some behavioural effects. However, this doesn't seem the right thing to say about phenomenal concepts. There is surely nothing immediately contradictory in the idea that an experience picked out by some phenomenal concept has no subsequent effects on behaviour or anything else. Epiphenomenalism about phenomenal states doesn't seem to be a priori contradictory.[11][11] Yet it would be, if our ways of referring to phenomenal states analytically implied that they had behavioural effects.

3.2 Phenomenal Concepts as Perceptual Concepts

I said above that my old model of phenomenal concepts ran together the good idea that phenomenal concepts are related to perceptual concepts with the bad idea that both kinds of concepts are 'demonstratives'. Let me now try to develop the good idea unencumbered by the bad one.

My current view is that phenomenal concepts are simply special cases of perceptual concepts. Consider once more the example where I perceptually identify some bird and make some judgement about it (*THAT is a migrant*). I earlier explained how the perceptual concept employed here could either be a concept of an individual bird or the concept of a species. I want now to suggest that we think of phenomenal concepts as simply a further deployment of the same sensory templates, but now being used to think about perceptual experiences themselves, rather than about the objects of those experiences. I see a bird, or visually imagine a bird, but now I think, not about that bird or a species, but about the experience, the conscious awareness of a bird.[12][12]

The obvious question is—what makes it the case that I am here thinking about an experience, rather than an individual bird or a species? However, we can give the same answer here as before. I earlier explained how the subject's dispositions to carry information from one encounter to another can decide whether a given sensory template is referring to an individual rather than a species, or vice versa—if the subject projects species-appropriate information, reference is to a species, while if the subject projects individual-appropriate information, reference is to an individual. So let us apply the same idea once more—if the subject is disposed to project experience-appropriate information from one encounter to another, then the sensory template in question is being used to think about an experience. For example,

---

[11][11] This leaves it open, of course, that there may be other good arguments against epiphenomenalism, apart from a priori arguments. Cf. Papineau 2002 section 1.4.

[12][12] Does this mean that perceptual experiences are the only items that can be thought about phenomenally? This seems doubtful. To consider just a couple of further cases, what about emotions, and pains? At first pass, it certainly seems that these states too can be picked out by phenomenal concepts—yet they are not obviously examples of perceptual experiences. There are two ways to go here. One would be to understand perception in a broad enough way to include such states. After all, emotions and pains are arguably representational states, and so could on these grounds be held to be a species of perception. Alternatively, we might distinguish these states from perceptions, but nevertheless allow that they are similar enough for us to think about them in ways that parallels phenomenal thinking about obviously perceptual states. I have no strong views on this choice, but in what follows I shall simplify the exposition by sticking to perceptions.

suppose I am disposed to project, from one encounter to another, such facts as that what I am encountering ceases when I close my eyes, goes fuzzy when I am tired, will be more detailed if I go closer, and so on. If this is how I am using the template as a repository of information, then I will be referring to the visual experience of seeing the bird, rather than the bird itself. More generally, if they are used in this kind of way—to gather experience-appropriate information, so to speak—the same sensory templates that are normally used to think about perceptible things will refer to experiences themselves.

Can phenomenal concepts pick out experiential particulars as well as types? In the perceptual case, as I have just explained, there is room for such differential reference to both particular objects and to types, due to the possibility of differing dispositions to carry information from one encounter to another. In principle it may seem that the same sort of thing could work in the phenomenal case. The trouble, however, is that particular experiences, by contrast with ordinary spatio-temporal particulars, do not seem to persist over time in the way required for re-encounters to be possible. Can the same particular pain, or particular visual sensation, or particular feeling of lassitude, re-occur after ceasing to be phenomenally present? It is true that we often say things like 'Oh dear, there's that pain again—I thought I was rid of it'. But nothing demands that we read such remarks as about quantitative rather than qualitative identity: nothing forces us to understand them as saying that the same particular experience has re-emerged, as it were, rather than that the same experiential type has been re-instantiated (note in particular that experiences do not seem to allow anything analogous to the spatio-temporal tracking of ordinary physical objects). In line with this, note that information about experiences, as opposed to information about spatio-temporal particulars, does not seem to divide into items that are projectible across encounters with a particular and items that are projectible across encounters with a type.

Given all this, I am inclined to say that phenomenal concepts cannot refer differentially both to particulars and to types. Rather they always refer to types—that is, to the kind of mental item that can clearly re-occur. As I am conceiving of perceptual and phenomenal concepts, the function of a concept is to carry information about its referent from one encounter to another—and it seems that only phenomenal types and not particulars can be re-countered.

The corollary is that, when we do refer to particular experiences, we cannot be using our basic apparatus of phenomenal concepts, given that these are only capable of referring to phenomenal types. Rather, we must be invoking more sophisticated conceptual powers, such as the ability to refer by description (thus *the particular pain I am having now*, or *the particular experience of crimson I enjoyed at last night's sunset*).

3.3 Phenomenal Use and Mention

This model of phenomenal concepts as a species of perceptual concept retains one crucial feature from my earlier quotational-indexical model, namely, that phenomenal references to an experience will involve an instance of that experience, and in this sense will use that experience in order to mention it.

To see why, think about what happens when a phenomenal concept is exercised. Some sensory template is activated, and is used to think about an experience. This sensory activation will either be due to externally generated sensory stimuli or to autonomous imaginative activity. That is, you will either be perceiving the environment, or employing perceptual imagination. For example, either you will be perceiving a bird, or you will be perceptually imagining one. Except, when phenomenal thought is involved, this template is also used to think about perceptual experience, rather than just about the objects of perceptions. You look at a bird, or visually imagine that bird, but now use the sensory state to think about the visual experience of seeing the bird, and not only about the bird itself.

This means that that any exercise of a phenomenal concept to think about a perceptual experience will inevitably either involve that experience itself or an imaginary recreation of that experience. If we count imaginary recreations as 'versions' of the experience being imagined, then we can say that phenomenal thinking about a given experience will always use a version of that experience in order to mention that experience.

Note how this model accounts for the oft-remarked 'transparency of experience' (Harman 1990). If we try to focus our minds on the nature of our conscious experiences, all that happens is that we focus harder on the objects of those experiences. I try to concentrate on my visual experience of the bird, but all that happens is that I look harder at the bird itself. Now, there is much debate about exactly what this implies for the nature of conscious experience (cf. Stoljar, forthcoming). But we can by-pass this debate here, and simply attend to the basic phenomenon, which I take to be the phenomenological equivalence of (a) thinking phenomenally about an experience and (b) thinking perceptually with that experience. What it's like to focus phenomenally on your visual experience of the bird is no different from what it's like to see the bird.

On my model of phenomenal thinking, this is just what we should expect. I said at the end of the last section that the phenomenology of perceptual experiences is determined by which sensory template they involve, and not by what information they carry with them. I have now argued that just the same sensory templates underly both perceptual experiences and phenomenal thoughts about those experiences. It follows that perceptual experiences and phenomenal thoughts about them will have just the same phenomenology. This explains why thinking phenomenally about your visual experience of a bird feels no different from thinking perceptually about the bird itself.

3.4 A Surprising Implication

The story I have told so far has an implication that some might find surprising. On my account, the semantic powers of phenomenal concepts would seem to depend on their cognitive function, rather than their phenomenal nature. I have argued that phenomenal concepts refer to conscious experiences because it is their purpose to accumulate information about those experience. As it happens, exercises of such concepts will in part be constituted by versions of the conscious experiences they refer to, and so will share the 'what-it's-likeness' of those experiences. But this latter, phenomenal fact seems to play no essential role in the semantic workings of phenomenal concepts. To see this, suppose that we had evolved to attach information about conscious experiences to states other than sensory templates—to words in some language of thought, perhaps. Wouldn't these states refer equally to experiences, and for just the same reason, even though their activation did not share the phenomenology of their referents? However, this might seem in tension with the idea that phenomenal concepts involve some distinctive mode of phenomenal self-reference to experiences. If the phenomenality of phenomenal concepts is incidental to their referential powers, then in what sense are they distinctively phenomenal? (Cf. Block, forthcoming.)

Note that my earlier 'quotational-indexical' account of phenomenal concepts is not open to this kind of worry. On that account, phenomenal concepts used experiences as exemplars, rather than as ways of implementing a cognitive role. Given this, it is essential to the phenomenal concept of *seeing something red*, say, that 'quotes' some version of that experience, just as it is essential to the quotational referring expression ' "zymurgy" ' that it contain the last word in the English dictionary within the inner quotation marks. On the quotational-indexical account then, there is no question of some state referring to an experience in the same way as a phenomenal concept does, yet its exercise not involve the experience.

Note also that the worrisome implication is not peculiar to the particular theory of the semantics of phenomenal concepts I have defended in this paper. It will arise on any theory that makes the semantic powers of phenomenal concepts a matter of their conceptual role, or their informational links to the external world, or any other facet of their causal-historical workings. For any theory of this kind will make it incidental to the referential powers of phenomenal concepts that they have the same phenomenology as their referents. Any such theory leaves it open that some other states, with different or no phenomenology, could have the same causal-historical features, and so refer to experiences for the same reason that phenomenal concepts do.

My response to this worry is that there is no real problem here. On my account, it is indeed true that phenomenal concepts refer because of their cognitive function, not because of their phenomenology, and therefore that other states with different or no phenomenology, but with the same cognitive function, would refer to the same experiences for the same reasons. I see nothing wrong with this. Of course, it is a further question whether we would wish to include any such non-phenomenological states within the category of 'phenomenal' concepts, given their lack of what-it's-likeness (cf. Tye 2003). But this is no grounds for denying that they would refer to experiences for just the same reason as phenomenal concepts do.

I shall come back to the issue of what counts as a 'phenomenal' concept in the next section. But first let me ask a somewhat different question. Given that other items could in principle play the cognitive role that determines reference to experiences, why do we use experiences themselves for this purpose? What is it about conscious experiences that makes them such a good vehicle for referring to themselves?

One possible answer is that this use of experiences is somehow well-suited to answering certain questions. To adapt an example of Michael Tye's (2003, p. 102), suppose that we are wondering whether the England one-day cricket strip is visually darker than the Indian one. By thinking phenomenally about these colours, we will generate versions of the relevant experiences, and so be in a position to compare them directly.

This makes some sense, but I think a simpler answer may be possible. Consider the analogous question: why do we use perceptual experience to represent perceptible items such as people, physical objects, animals, plants, shapes, colours, and so on? After all, in this case too the referential powers of these states are presumably determined by some type of cognitive role, which could in principle have been played by something other than perceptual experiences themselves. Here the obvious answer seems to be that the perceptions are especially good for thinking about perceptible entities simply because they are characteristically activated by those entities, and so are well-suited to feature in judgements that those entities are present. It would unnecessarily duplicate cognitive mechanisms to use the perceptual system to identify perceptible entities, yet something other than perceptual experiences as the vehicle for occurrent thoughts that imply that those entities are present.

This thought applies all the more in the phenomenal case. Conscious experiences are excellent vehicles for thinking about those selfsame experiences, simply because they are automatically present whenever their referents are. The fact that we use experiences to think about themselves means that we don't have to find other cognitive resources to frame occurrent thoughts about the presence of experiences.

4. Phenomenal Concepts and Anti-Materialist Arguments

4.1 The Knowledge Argument

In the last subsection I raised the issue of what exactly qualifies a concept as 'phenomenal'. I have no definite answer to this definitional question. Far more important, from my point of

view, is whether phenomenal concepts as introduced so far provide effective answers to the standard anti-physicalist arguments. I shall now aim to show that they do this. In the course of doing so, however, I shall highlight those features of phenomenal concepts that are important to their serving this philosophical function. We can leave it open which features of phenomenal concepts are essential to their counting as 'phenomenal'. But it is worth being clear about which features matter to the philosophical arguments.

Let me begin with Frank Jackson's knowledge argument. Here the Type-B physicalist response is that there is indeed a sense in which Mary doesn't 'know what seeing red is like' before she comes out of her room, despite her voluminous material knowledge. But this is not because there is any objective feature of reality that her material knowledge fails to capture, but simply because there is a way of thinking about the experience of seeing red that is unavailable to her while still in the room. Before she comes out of the room, she lacks a phenomenal concept of the experience of seeing red. She could always think about the experience using her old material concepts all right, but not with any phenomenal concept. This is why she did not know that *seeing red =THAT experience* (where this is to be understood as using a material concept on the left-hand side and a phenomenal concept on the right-hand side).

The crucial feature of phenomenal concepts, for the purposes of this argument, is that they are experience-dependent: the concept's acquisition depends on its possessor having previously undergone the experience it refers to. This is why she doesn't 'know what seeing red is like' before she comes out of the room. She needs to see red in order to acquire the conceptual wherewithal to think *seeing red = THAT experience*.

The reason that Mary's new concept is experience-dependent is that it requires a sensory template, and her acquisition of this template depends on her visual system having previously been activated by some red surface. This is of course a contingent feature of human beings. We can imagine beings who are born with the sensory templates that we acquire from colour experiences (cf. Papineau 2002 section 2.8). Still, as it happens, humans are not like this. They are born with few, if any, sensory templates, but must rather acquire them from previous experiences. (If humans were born with the sensory templates activated by red surfaces, then physicalists could not answer the knowledge argument by saying that Mary needs a red experience in order to acquire a phenomenal concept of red. But if humans were like that, then physicalists wouldn't need to answer the knowledge argument in the first place, since Mary would already have a phenomenal concept of red before she left her room, and so would already be in a position to know that *seeing red = THAT experience*, by courtesy of an imaginative exercise of her phenomenal concept on the right-hand side.)

4.2 *I am not now having or imagining THAT experience*

It is worth distinguishing the experience-dependence of phenomenal concepts from the use-mention feature discussed in subsection 3.3 above. Even though normal examples of phenomenal concepts, like the one Mary acquires on leaving her room, have both the experience-dependence and use-mention features, there is space in principle for concepts which are phenomenal in the sense of being experience-dependent but which don't use experiences to mention themselves. Indeed, I would argue that this is not just an abstract possibility—there are actual concepts which display experience-dependence, but not the use-mention feature.

To see why, recall the earlier discussion of perceptually derived concepts. These derived concepts involved the creation of some non-sensory file to house the information associated with some perceptual concept, and they made it possible to think about perceptible entities even when those entities were not being perceived or perceptually imagined. Analogously, we can posit a species of 'phenomenally derived concept'. Suppose someone starts off, like

Mary, by thinking phenomenally using a sensory template instilled by previous experiences. But then she creates a non-sensory file in which to house the information that has become attached to that template, and which will henceforth allow her to think about the experience without any sensory activation. I say she now has a phenomenally derived concept. Exercises of this concept won't activate the experience it mentions, and so this concept will fail to satisfy the use-mention requirement. But this phenomenally derived concept will still satisfy the experience-dependence requirement, in that its creation depends on a prior phenomenal concept which in turn depends on previous experiences.

The possibility of phenomenally derived concepts offers an answer to an objection raised in the Introduction. This was that standard accounts of phenomenal concepts seem to imply that any exercise of a phenomenal concept demands the presence of the experience it refers to or an imaginatively recreated exemplar thereof. However, this seems too demanding. Surely someone like Mary can use her new concept to think truly that *I am not now having THAT experience (nor recreating it in my imagination)*. Yet this should be impossible, if any exercise of her phenomenal concept does indeed require the relevant experience or its imaginative recreation.

My response to this objection is that Mary thinks the problematic thought with the help of a phenomenally derived concept.[13[13]] She starts with a phenomenal concept based on some sensory template, and then creates a non-sensory file to carry the information associated with the template. This allows her to think about the relevant experience without activating the associated sensory template—that is, without either having or imaginatively recreating the experience in question. She thinks *I am not now having or imagining THAT experience*—and since she is using a phenomenally derived concept, what she thinks can well be true.

Some readers might wonder whether it is really appropriate to say that Mary is here exercising a phenomenal concept. After all, if this concept is realized non-sensorily, then why is it any more 'phenomenal' that the general run of ordinary concepts? In particular, would we want to say that someone knows 'what it is like' to see something red, merely in virtue of thinking *seeing red = THAT experience*, where the right hand side deploys a phenomenally derived concept, and so does not require the thinker actually to have the relevant experience or its imaginative recreation?

Well, I have no principled objection if someone wants to withhold the description 'phenomenal' on these grounds. But note that this move is not available to someone who wants to press the objection at hand, which after all is precisely that there seems room for a thinker to exercise a 'phenomenal' concept while not having any version of the experience referred to. For this objection to make any sense, 'phenomenal' cannot be understood as requiring sensory realization per se. Rather, it has to be understood simply as standing for those concepts whose acquisition depends on undergoing the relevant experience. And in this sense of 'phenomenal'—experience-dependence—phenomenally derived concepts do explain how someone can think phenomenally without having any version of the corresponding

---

[13[13]] It might occur to some readers that another answer to the challenge would be to insist that Mary must really be using some old pre-exposure material concept of red experience when she thinks the supposedly problematic thought. However, this will not serve. To see clearly why, it will help to vary the Mary thought-experiment slightly. Suppose that, on her exposure, Mary was shown a coloured piece of paper, rather than a rose, and that she wasn't told what colour it was. The objection would seem still to stand. The conceptual powers she acquires from her exposure would still seem to enable her later truly to think, *I am not now having or imagining THAT experience*. But now she can't be using any of her old pre-exposure concepts to refer to the experience. For, if she doesn't know what colour the paper was, she won't know which of her old concepts to use.

experience. For phenomenally derived concepts are certainly experience-dependent, given that they derive from phenomenal concepts that derive from prior experiences.

4.3 Semantic Stability and A Posteriori Necessity

Let me now turn to anti-materialist arguments which turn on modal considerations. The best know of these is Kripke's argument against the identity theory in Naming and Necessity (1980). But before addressing this, I would like to consider a different modal argument, which I shall call 'the argument from semantic stability'. As it turns out, both these arguments can be blocked by appealing to the use-mention feature of phenomenal concepts. But the way this works is rather different in the two cases.

The argument from semantic stability hinges on the fact that Type-B physicalists take identity claims like *nociceptive-specific neuronal activity = pain* (where the right hand side uses a phenomenal concept) to be a posteriori necessities. The distinctness of the concepts on either side of the identity claim means that there is no question of knowing such claims a priori. Even after she had acquired both concepts, Mary still needs empirical information to find out that *pain* was the same experience as *nociceptive-specific neuronal activity*. In this respect, Type-B physicalists take mind-brain identity claims to be akin to such familiar a posteriori necessities as *water = H_2O*, *lightning = electric discharge*, or *Hesperus = Phosphorous*.

The objection to Type-B physicalism is then that phenomenal mind-brain identities cannot possibly be akin to these familiar a posteriori necessities, because a posteriori necessity is characteristically due to 'semantic instability', but phenomenal concepts are semantically stable.[14[14]]

Let me unpack this. Note first that, in all the examples of familiar a posteriori necessities listed above, the referential value of at least one of the concepts involved—*water, lightning, Hesperus* (and indeed *Phosphorus*)—depends, so to speak, on how things actually are. If it had turned out that XYZ and not H_2O is the colourless liquid in rivers, etc, then *water* would have referred to XYZ. If it had turned out that some heavenly body other than Venus is seen in the early morning sky, then *Hesperus* would have referred to that other heavenly body. And so on.

This observation suggests the hypothesis that a priority and necessity only come apart in the presence of semantically unstable concepts. On this hypothesis, claims formulated using semantically *stable* concepts will be necessary if and only if they are a priori. Certainly there are plenty of concepts that seem to be stable in the relevant sense, that is, whose referents seem not to be actual-fact-dependent. Physical concepts like *electron* or *H_2O* seem to be like this, as do such everyday concepts like *garden* or *baseball*. And if we stick to claims involving only such stable concepts (*electrons are negatively charged*, say, or *baseball is a game*) then it does seem plausible that these claims will be necessary if and only if they are a priori.

The general idea here is that necessities will only be a posteriori when you are ignorant of the essential nature of some entity you are thinking about. If your concepts are transparent to you, if their real essence coincides with their nominal essence, so to speak, then you will be able to tell a priori whether claims involving them are necessary or not. But with semantically unstable concepts we need empirical information to know what they refer to, and so to ascertain whether a necessary proposition is expressed. To take just the first example above,

---

[14[14]] For versions of this argument see Chalmers 1996, Jackson 1998, Bealer 2002. The notion of semantic stability is due to Bealer. I would like to thank Philip Goff for helping me to understand these issues.

it takes empirical work to discover that H_2O is the referent of *water*, and so that *water = H_2O* is necessarily true.

The claim is thus that a posteriori necessity always turns on the presence of concepts whose reference is actual-fact-dependent; correlatively, if we keep away from such concepts, then necessity and a priority will always go hand in hand.

If this claim is accepted, then it is indeed hard to see how phenomenal mind-brain identity claims like *pain = nociceptive-specific neuronal activity* could be a posteriori necessities. For the phenomenal concepts like *pain* do not seem to be semantically unstable. There seems little sense to the idea that it could have turned out, given different empirical discoveries, that *pain* referred to something other than its actual referent.

But then, given the general thesis that a posteriori necessity requires semantic instability, it follows that *nociceptive-specific neuronal activity = pain* cannot be an a posteriori necessity. Since we don't need any empirical information to know what *pain* refers to, we must already know what proposition the claim *nociceptive-specific neuronal activity = pain* expresses, just in virtue of our grasp of the concept *pain*, and so ought to be able to tell a priori that this claim is true, if it is. But we can't, so it can't be true.

In the face of this argument, Type-B physicalists need to deny that a posteriori necessities require semantically unstable concepts. There seems no doubt that phenomenal concepts are semantically stable. And it is constitutive of Type-B physicalism that phenomenal mind-brain identities are a posteriori. So the only option left is to insist that these identities are a posteriori necessities which involve no semantic instability.

Opponents will ask whether phenomenal mind-brain identities are the only such cases. If they are, then the Type-B physicalist move can be charged with unacceptable ad hocness. Type-B physicalists would seem to be guilty of special pleading, if the connection between a posteriori necessity and semantic instability holds good across the board, excepting only those cases where phenomenal concepts are involved.

One obvious way for Type-B physicalists to respond to this charge of ad hocness is to seek other examples of a posteriori necessities that do not involve semantic instability. Obvious possibilities are identities involving proper name concepts that (unlike *Hesperus* or *Phosphorous*) do not have their references fixed by salient descriptions (*Cicero = Tully* , say) or again identities involving perceptual concepts (such as *reflectance profile Φ = red*, where the right hand side uses a perceptual colour concept).

However, opponents of Type-B physicalism will deny that these a posteriori necessities are free of semantic instability. Maybe the concepts involved don't have their references fixed by salient descriptions. But they will insist that this is not the only way for concepts to be semantically unstable, and that more careful analyses of semantic stability will show that proper name and perceptual concepts are indeed unstable, while phenomenal concept are not, and are thus still anomalous among concepts that enter into a posteriori necessities.[15[15]]

[15[15]] Thus we might this of semantic instability are requiring, not the descriptive fixing of reference, but only that thinkers who possesses the relevant concepts will pick out different entities as their referents given different scenarios presented in fundamental terms (Chalmers and Jackson 2001, Chalmers 2002). (Since physicalists will equate fundamental terms with physical terms, this characterization of semantic instability then leads to the thesis that physicalism implies that all the facts must follows a priori from the physical facts.) I myself am very dubious that proper name and perceptual concepts are unstable in this sense—this seems to follow only if we illegitimately presuppose that thinkers need a potential appreciation of the semantic workings of their proper name and perceptual concepts in order to possess them. Another possible understanding of semantic

I remain to be persuaded about this charge of anomalousness. However, I shall not dig my heels in at this point. Rather let me concede, for the sake of the argument, that phenomenal mind-brain identities are indeed anomalous in not involving any semantically unstable concepts. I don't accept that this means that these identities cannot be true. Rather, I say that phenomenal mind-brain identities are anomalous because phenomenal concepts are very peculiar. More specifically, phenomenal concepts have the very peculiar feature of using the experiences they refer to. When we reflect on this, we will see that it is unsurprising that identities involving phenomenal concepts should be unusual in combining semantic stability with a posteriori necessity.

The underlying anti-physicalist thought, recall, was that semantic stability goes hand in hand with knowledge of real essences; conversely, if thinkers are ignorant of real essences, they must be using unstable concepts. The complaint about Type-B physicalism, then, is that it requires the possessors of phenomenal concepts like *pain* to be ignorant of the real physical essence of pain, even though the concept *pain* is manifestly stable. The anti-physicalists thence conclude that *pain* must refer to something non-physical, something with which the possessors of the concept are indeed directly acquainted.

But Type-B physicalists can respond that, however it is with other concepts, this combination of semantic stability with ignorance of essence is just what we should expect given the use-mention feature characteristic of phenomenal concepts. Even if phenomenal concepts don't involve direct knowledge of real essences, they will still come out semantically stable, for the simple reason that the use-mention feature lead us to think of the referent as 'built into' the concept itself. Since the concept uses the phenomenal property it mentions, this alone seems to eliminate any conceptual or metaphysical space wherein *that* concept might have referred to something different.

Above I said that I remained to be persuaded that phenomenal concepts are distinguished from proper name and perceptual concepts in uniquely displaying this combination of semantic stability and ignorance of essence. Still, at an intuitive level we can see why phenomenal concepts should appear special in this way. When we think of proper name concepts like *Cicero* or perceptual concepts like *red*, we seem able to make intuitive sense of scenarios where the reference-fixing facts are different, where the concept *Cicero* names some other person than Cicero, and scenarios where the perceptual observational concept *red* refers to other reflectance profiles. But we don't seem able to do this with phenomenal concepts—it doesn't seem that there are any scenarios whose actuality would make *pain* refer to something other than pain. This is because we think of phenomenal concepts as essentially using the very phenomenal properties that are being referred to. This seems to leave no room for the idea that that a given phenomenal concept could have referred to some other property than it does refer to, if the facts had turned out differently. As long as it remains the same phenomenal concept, then its exercises will involve the same phenomenal property—and then, since it mentions whichever phenomenal property it uses, it will refer to that property, however the actual facts turn out.[16[16]]

---

instability is as requiring only that the facts which relate concepts to their referents might have been different. (Thus Chalmers 1996, p. 373 'if the subject cannot know that R is P a priori, then reference to R and P is fixed in different ways and the reference-fixing intensions can come apart in certain conceivable situations.') Note that this latter understanding will not require physicalists to hold that all the facts must follows a priori from the physical facts, given that it does not seem an a priori matter which facts relate concepts to their referents. I have an open mind on whether this last understanding of semantic stability leaves phenomenal concepts stable while implying that proper name and perceptual concepts are not. Cf. the next footnote.

[16[16]] Does this line of thought really distinguish proper name and perceptual concepts from phenomenal concepts? There are worries on both sides. First, it is not clear how we are to conceive of *Cicero* and *red* as referring to different things while remaining the same concept—note in particular

In a sense, phenomenal concepts are too close to their referents for it to seem possible that those same concepts could refer to something else. With other concepts that enter into a posteriori identities, including proper name and perceptual concepts, we can imagine the 'outside world' turning out in such a way that they referred to something other than their actual referents. Some other person might have turned out be the historical source of my *Cicero* concept, some other physical property might have turned out to answer to my *red* concept, and so on. But in the case of phenomenal concepts, the referent seems to be part of the concept itself, leaving no room for any such possibility.[17][17]

If this is right, then the semantic stability of phenomenal concepts provides no reason to think that they must refer to non-physical properties with which their possessors are directly acquainted. For the use-mention feature of phenomenal concepts yields an independent explanation of why they should be semantically stable, even while their possessors remain ignorant of the real physical essences of their referents.

4.5 Kripke's Original Argument

Let me now turn to Saul Kripke's original argument against the mind-brain identity theory. There are significant differences between this and the argument from semantic stability. Kripke doesn't seem to be committed to the thesis that necessity only comes apart from a priority in the presence of semantic instability. Kripke's paradigm cases of a posteriori necessities involve names whose reference is determined in line with his causal theory of reference (*Cicero = Tully*), and there is nothing in Kripke to suggest that this a posteriority demands that the referential values of these names must be actual-fact-dependent. From a Kripkean point of view, there is nothing special here that needs explaining—a posteriori necessities are simply the natural consequence of the non-descriptive way reference is determined for normal names.

Kripke's argument hinges not on the a posteriority of the physicalist's identity claims, but on their apparent contingency. Kripke has no complaint about the a posteriority of claims like *pain = nociceptive-specific neuronal activity*—a posteriori necessities are par for the course, from a Kripkean point of view. What Kripke takes to be problematic about these mind-brain identities, rather, is that it seems that they might have been false: intuitively we feel that there are possible worlds—zombie worlds, say—where nociceptive-specific neuronal activity is not identical to pain. Now, there is nothing per se incoherent in the idea of a necessary identity that appears contingent. For example, we can make sense of the idea that *Hesperus = Venus* might have been false, by construing this identity claim as saying that the heavenly body that appears in the morning is Venus—something which might well have been otherwise. But now—and this is Kripke's point—this way of explaining the appearance of contingency does require you to construe the relevant referring term as semantically unstable, for it demands that you read the term in a way that makes it come out referring to something different if the actual facts are different. Yet *pain* cannot be read in such a way, which means, say Kripke, that the physicalist has no satisfactory way of explaining the apparent contingency of phenomenal mind-brain identities.

---

that it is by no means enough that the words 'Cicero' or 'red' might have referred to different things. Second, it is not obvious exactly how to individuate phenomenal concepts like *pain* so as to make it impossible for them to refer to different things—the issues discussed in subsection 3.4 are relevant here.

[17][17] Thus consider Chalmers 2003b p. 233: 'Something very unusual is going on here. . . . In the pure phenomenal case . . . the quality of the experiences plays a role in constituting the epistemic content of the concept . . . One might say very loosely that in this case, the referent of the concept is somehow present inside the concept's sense, in a way much stronger than in the usual cases of "direct reference".'

So Kripke gets to the same place as the argument from semantic stability, but from a different starting point. However, the differences between the two arguments mean that Kripke's argument demands a rather different response from the physicalist. It is no good to reply to Kripke that a posteriori necessity is consistent with semantic stability, for he agrees about this, and indeed will allow that there are non-phenomenal examples of this combination (*Cicero = Tully*). What he insists on, however, and this is a different point, in that apparently contingency is inconsistent with semantic stability, and that physicalists therefore have no way of explaining the apparent contingency of mind-brain identities.

In response to the argument for semantic stability, I denied that a posteriori necessity required semantic instability, at least where mind-brain identities are involved. However, I don't think that there is any corresponding room to deny that apparent contingency must involve semantic instability. If a claim can be understood as actually true yet possibly false, then some of the concepts involved must shift reference. Given that the concepts in phenomenal mind-brain identity claims are all semantically stable, this leaves the physicalist with one option—deny that phenomenal mind-brain identities are apparently contingent.

This might seem all wrong—isn't it agreed on all sides that we can cogently conceive of zombies (even if they aren't really possible), and therewith that mind-brain identities at least seem possibly false? So what room is there for the physicalist to deny that mind-brain identities are apparently contingent?

Well, I agree that physicalists are compelled to allow that phenomenal mind-brain identities seem possibly false. But this isn't yet to allow that they seem contingent. Contingency requires not only falsity in some possible worlds, but also truth in the actual world. And it is specifically this combination that generates the need for semantic instability, to give a non-actual referent that falsifies the claim in some possible world. But there is another way for an identity claim to seem possibly false—namely, for it simply to seem false. And in that case there is nothing to require semantic instability.

Let me go more slowly. Consider people who think that Cicero is actually different from Tully. To them the claim that *Cicero = Tully* will of course seem possibly false, because it will seem necessarily false. But nothing here demands that we understand them as thinking of either of these names in a semantically unstable way, as possibly referring to something other than their actual referents. Such a construal is only called for when we have the combination of both apparent possible falsity and actual truth. If somebody thought that *Cicero = Tully* is true, but might have been false, then we must construe them as thinking *the greatest Roman orator = Tully*, or some such, to explain how they have room, so to speak, for the thought that a necessary truth might have turned out to be false . But there's no need to read them this way if they simply think that Cicero and Tully are actually different.

So my suggestion is that physicalists should say that mind-brain identities strike us just like *Cicero = Tully* strikes people who think Cicero and Tully are different people. They seem non-necessary simply because they seem false. Zombies seem possible simply because pain seems actually distinct from nociceptive-specific neuronal activity. From this point of view, Kripke has misdescribed the crucial zombie intuition from the start. It's not an intuition of apparent contingency—some confused intuition that pain could come apart from nociceptive-specific neuronal activity in some other possible world—but simply a direct intuition of falsity—pain is different from nociceptive-specific neuronal activity in the actual world.[18[18]] [19[19]]

[18[18]] It is consistent with this diagnosis that thought-experiments about possible zombies might nevertheless play an epistemologically significant role in clarifying the content of our intuitions. (I owe this point to George Bealer.) Suppose it is agreed on all sides that pains and nociceptive-specific

4.5 The Intuition of Distinctness

Some readers might be wondering why the last subsection hasn't conceded the anti-physicalist case to Kripke. My suggestion is that physicalists should explain our attitude to the possibility of zombies by allowing that mind-brain identity claims strike us as false. But isn't this tantamount to denying physicalism?

Not necessarily. I say physicalists should allow that physicalism seems false, not allow that it is false. That is, physicalists should maintain that we have an intuition of mind-brain distinctness, but that this intuition is mistaken.

This is by no means ad hoc. It seems undeniable that most people have a strong intuition of mind-brain distinctness—an intuition that pains are something extra to brain states, say. This intuition is prior to any philosophical analyses of the mind-brain relation, and indeed persists even among those (like myself) who are persuaded by those analyses that dualism must be false. Given this, it is a requirement on any satisfactory physicalist position that it offer some explanation of why we should all have such a persistent intuition of mind-brain distinctness, even though it is false. Physicalists need to recognize and accommodate the intuition of distinctness, quite apart from requiring it to deal with Kripke's argument.

There are a number of possible ways of explaining away the intuition of distinctness, especially for physicalists who recognize phenomenal concepts. I myself favour an explanation that hinges on the use-mention feature of phenomenal concepts, and which elsewhere I have called 'the antipathetic fallacy'.[20[20]]

Suppose you entertain a standard phenomenal mind-brain identity claim like *pain = nociceptive-specific neuronal activity*, deploying a phenomenal concept on the left-hand side and a material concept on the right. Given that the phenomenal concept uses the experience it mentions, your exercise of this concept will depend on your actually having a pain, or an imagined recreation thereof. Because of this, exercising a phenomenal concept will feel like having the experience itself. The activity of thinking phenomenally about pain will introspectively strike you as involving a version of the experience itself.

---

neuronal activity are perfectly correlated in the actual world. We might then wonder whether this is a matter of property identity or whether it is a correlation between two distinct properties. One way of resolving this would be to ask whether 'these' properties could come apart—for example, are zombies possible? In this way, the intuitive possibility of zombies could serve to make it clear to us that, even though we take pain and nociceptive-specific neuronal activity to be perfectly correlated in the actual world, we still view them as distinct properties. However, viewing the intuition that zombies are possible as having this epistemological significance is quite different from saying that pain = nociceptive-specific neuronal activity seems contingent. The story I have just told has no place for the thought that pain is identical to nociceptive-specific neuronal activity in the actual world but distinct in some other possible world. On the contrary, the epistemological significance just ascribed to the zombie thought-experiment hinges on the fact that actual identity stand or falls with necessary identity. I owe this point to George Bealer.

[19[19]] Physicalists are often too ready to see things in Kripkean terms, and to seek some reading of mind-brain claims on which true mind-brain identities might have been false. Thus Perry (2001) assimilates phenomenal concepts to indexicals, and then says that the possibility that 'pain'/'this brain state' might have picked out some different state than C-fibres firing 'underlies the sense of contingency'. I don't recognize the intuition that this story is supposed to explain. When I think that 'pains might not have been C-fibres firing', my thought isn't that 'this cognitive process (whatever it is) might (have turned out) not (to) be C-fibres firing', but simply that 'pains aren't C-fibres firing'.

[20[20]] See Papineau 1993b. For an alternative explanation of the intuition of distinctness, see Melnyk 2003.

Things are different with the exercise of the material concept on the right-hand side. There is no analogous phenomenology. Thinking of nociceptive-specific neuronal activity doesn't require any pain-like feeling. So there is a intuitive sense in which the exercise of this material concept 'leaves out' the experience at issue. It 'leaves out' the pain in the sense that it doesn't activate any version of it.

Now, it is all too easy to slide from this to the conclusion that, in exercising such a material concept, we are not thinking about the experiences themselves. After all, doesn't this material mode of thought 'leave out' the experiences, in a way that the phenomenal concept does not? And doesn't this show that the material concept simply doesn't refer to the experience denoted by our phenomenal concept of pain?

This line of thought is terribly natural. (Consider the standard rhetorical ploy: 'How could pain arise from mere neuronal activity?') But of course it is a fallacy. There is a sense in which material concepts do 'leave out' the feelings. Uses of them do not in any way activate the experiences in question, by contrast with uses of phenomenal concepts. But it simply does not follow that these material concepts "leave out" the feelings in the sense of failing to refer to them. They can still refer to the feelings, even though they don't activate them.

After all, most concepts don't use or involve the things they refer to. When I think of being rich, say, or having measles, this doesn't in any sense make me rich or give me measles. In using the states they refer to, pheneomenal concepts are very much the exception. So we shouldn't conclude on this account that materialal concepts, which work in the normal way of most concepts, in not using the states they refer to, fail to refer to those states.

Still, fallacious as it is, this line of thought still seems to me to offer a natural account of the intuitive resistance to physicalism about conscious experiences. This resistance arises because we have a special way of thinking about our conscious experiences, namely, by using phenomenal concepts. We can think about our conscious experience using concepts to which they which bear a phenomenal resemblance. And this then creates the fallacious impression that other non-phenomenal ways of thinking about those experiences fail to refer to the felt experiences themselves.[21[21]]

5 Chalmers on Type-B Physicalism

5.1 Chalmers' Dilemma

David Chalmers has recently mounted an attack on the whole Type-B physicalist strategy of invoking phenomenal concepts in order explain the mind-brain relation (Chalmers 2006—his contribution to this volume). He aims to present Type-B physicalists with a dilemma. Let C be the thesis that humans possess phenomenal concepts. As Chalmers sees it, Type-B physicalists require both (a) that C explains our epistemic situation with respect to consciousness and (b) that C is explicable in physical terms. However, Chalmers argues that there is no version of C that satisfies both these desiderata—either C can be understood in a way that makes it physically explicable, or in a way that allows as to explain our epistemic situation, but not both.

In order to develop the horns of this dilemma, Chalmers asks the physicalist whether or not C is conceptually guaranteed by the complete physical truth about the universe, P. That is, is P and not-C conceivable?

[21[21]] Note how my explanation implies that the intuition of distinctness will only arise when we are thinking with phenomenal concepts, which use the very states they mention, and not when we are thinking with phenomenally derived concepts. This seems to me quite in accord with the facts.

Suppose the physicalist says that this combination is conceivable. This makes the existence of phenomenal concepts conceptually independent of all physical claims. But then, argues Chalmers, all the original puzzles about the relation between the brain and phenomenal states will simply reappear as puzzles about the relation between the brain and phenomenal concepts themselves. So Chalmers holds that on this option C fails to be physically explicable.

The other option is for the physicalist to say that P and not-C is not conceivable. On this horn, claims about phenomenal concepts will not be conceptually independent of P (suppose that phenomenal concepts are conceived physically or functionally, say). However, this would mean that zombies (conceivable beings who are physically identical to us but lack consciousness) would be conceived as having phenomenal concepts. However, argues Chalmers, we don't conceive of zombies as epistemically related to consciousness as we are (after all, they are conceived as not having any consciousness). This argues that something more than C is needed to explain our peculiar relation to consciousness. So Chalmers hold that on this option C fails to explain our epistemic situation.

So—either P and not-C is conceivable, or it isn't. And on neither option, argues Chalmers, is C both physically explicable and explanatory of our epistemic situation.

5.2 The Dilemma Embraced

Far from viewing Chalmers as offering a nasty choice, I am happy to embrace both horns of his dilemma. I say that we can conceive of phenomenal concepts in a way that makes them conceptually independent of the physical facts, and also conceive them in a way that doesn't make them so independent. Moreover, I think that both these ways of thinking about phenomenal concepts allows phenomenal concepts to be simultaneously physically explicable and explanatory of our epistemic situation.

It is the use-mention feature of phenomenal concepts that allows them to be thought of in two different ways. Exercises of phenomenal concepts involve versions of the phenomenal states they refer to. Given this, thinking about phenomenal concepts requires us to think of the phenomenal states that they use. But according to Type-B physicalism, these used phenomenal states, like phenomenal states in general, can be thought of in two different ways—phenomenally and non-phenomenally. So we can think about (first-order) phenomenal concepts phenomenally, using (second-order) phenomenal concepts to think about the phenomenal states involved, or we can think about them non-phenomenally, thinking about the involved phenomenal states in physical or functional terms, say. Since the (second-order) phenomenal concepts used on the first option, like all phenomenal concepts, will be a priori distinct from any physical or functional concepts, the first way of thinking about (first-order) phenomenal concepts will make P and not-C conceivable, while the second way of thinking about (first-order) phenomenal concepts, as physical or functional, will make P and not-C inconceivable.[22[22]]

However, for a Type-B physicalist, these two ways of thinking still refer to the same entities—(first-order) phenomenal concepts. And these entities will have the same nature and cognitive role, however they are referred to. So the way they are referred to ought to make no difference to whether they are physically explicable and explanatory of our epistemic

[22[22]] This 'first-order' and 'second-order' talk should not be taken to imply some hierarchy of differently structured concepts. When I say that we can think about a 'first-order' phenomenal concept using 'second-order' one, all I mean is that the phenomenal property used by the 'first-order' phenomenal concept can itself be thought about using a phenomenal concept. There is no reason why the latter phenomenal concept should be differently structured from any other—indeed it can be the same (first-order) phenomenal concept that uses the phenomenal property in question. Still, the terminology of 'first-' and 'second-order' will help clarify my line of argument.

situation. They should satisfy these two desiderata however they are referred to. Let me now show that they do.

5.3 The First Horn

On the first horn, we think about (first-order) phenomenal concepts by using further (second-order) phenomenal concepts. That is, we note that exercises of (first-order) phenomenal concepts involve uses of phenomenal states, and when we think of the phenomenal states thus involved, we do so using further (second-order) phenomenal concepts.

The problem on this horn, according to Chalmers, relates to the epistemic and explanatory gap between physical and phenomenal claims. Chalmers views this gap as making a strong case for dualism, a case which Type-B physicalists seek to block by showing that the existence of this 'distinctive gap can be explained in terms of certain distinctive features of phenomenal concepts' (p. 00). However, even if invoking phenomenal concepts can so succeed in explaining the original gap between physical and phenomenal claims, Chalmers argues that physicalists will now need to explain away a new gap between physical claims and claims about the possession of phenomenal concepts—for, after all, on this horn of the dilemma they agree that P and not-C is conceivable, that is, that the physical facts do not conceptually necessitate claims about (first-order) phenomenal concepts.

The natural physicalist response is to argue that they can explain this new gap in just the way that they explained the original one. If we are conceiving of (first-order) phenomenal concepts using further (second-order) phenomenal concepts, then of course there will be a conceptual gap between physical claims and claims about phenomenal concepts. Still, if the original gap could be 'explained in terms of certain distinctive features of [first-order] phenomenal concepts', as Chalmers is allowing for the sake of the argument, why can't the new gap be explained in terms of the same features of (second-order) phenomenal concepts?

Chalmers objects that this explanation-repeating move will be either regressive or circular (p. 00). But it is not obvious to me why this should be so. In particular, there doesn't seem to be anything regressive or circular in repeating the explanatory use that I myself have made of phenomenal concepts, as I shall show in a moment.

I suspect that Chalmers' charge of regression or circularity reflects the very high demands he is placing on Type-B explanations of the conceptual gap. For the most part he leaves it open how such explanations might go, being happy to conduct his argument on an abstract level. But just before his charge of regression or circularity, he does propose one possible explanation of the original gap, suggesting that Type-B physicalists might say that phenomenal concepts give their possessors a distinctive kind of direct acquaintance with their referents, of a kind that 'one would not predict from just the physical/functional structure of the brain' (p. 00). Now, I agree that this kind of explanation is going to get Type-B physicalists into trouble—though not especially because it becomes regressive or circular when it is repeated at the higher level, but simply because it is unacceptable to start with for a physicalist to posit distinctive semantic powers of direct reference that correspond to nothing identifiable in physical or functional terms.

Still, this doesn't mean that any Type-B physicalist explanation of the conceptual gap is going to run into trouble. There is no question here of cataloguing all the different ways in which different Type-B physicalists have appealed to phenomenal concepts in order to account for various 'gaps'. Let me simply remind readers of some of the things I said earlier, and show that there is nothing regressive or circular about saying the same things about the relation between physical claims and phenomenally conceived claims about phenomenal concepts.

Like all Type-B physicalists, I take the existence of (first-order) phenomenal concept to imply that (first-order) phenomenal mind-brain identity claims are a posteriori. In response to the challenge that these claims are unique among a posteriori necessities in not hinging on semantically unstable concepts, I argued that such uniqueness is adequately explained by the use-mention feature of phenomenal concepts: this feature explains why we take it that phenomenal concepts will refer to the same referent whatever the facts, even though phenomenal concepts are arguably unlike other semantically stable concepts in not requiring transparent knowledge of the essential nature of their referents. As to the feeling that, even after this has been said, there remains something disturbingly unexplanatory about phenomenal mind-brain identities, my view is that this feeling doesn't stem from any semantic or epistemic peculiarity of these identities, but simply from the prior 'intuition of distinctness' that militates against our believing these identities to start with. (To the extent we embrace this intuition, then of course we will feel a real 'explanatory gap', for we will then want some causal explanation of why the physical brain should 'give rise to' the supposedly separate phenomenal mind.)[23[23]] As to the source of the intuition of distinctness, I explained this by once more appealing to the use-mention feature of phenomenal concepts, and the way it makes us think that non-phenomenal modes of thought 'leave out' the phenomenal feelings.

Now I don't see why I can't simply say all these things again, if Chalmers challenges me to account for the extra gap between P and C which arises when we are thinking of (first-order) phenomenal concept in (second-order) phenomenal terms, as on this first horn of his dilemma. (Second-order) phenomenal claims identifying the possession of (first-order) phenomenal concepts with physical states will be a posteriori necessities. If these are held to be unusual among a posteriori necessities in not involving semantic instability, I can point out that the use-mention feature of (second-order) phenomenal concepts explains why this should be so. If it is felt that, even after this has been said, there seems to be something disturbingly unexplanatory about claims identifying the possession of (first-order) phenomenal concept with physical states, I attribute to a higher-level 'intuition of distinctness', which arises because physical/functional ways of thinking about (first-order) phenomenal concepts seems to 'leave out' the feelings which are present when we think about (first-order) phenomenal concepts using (second-order) phenomenal concepts.

In short, just as the use-mention feature of (first-order) phenomenal concepts accounts for any peculiarities of the conceptual gap between physical/functional claims and (first-order) phenomenal claims about phenomenal properties, so does the use-mention feature of (second-order) phenomenal concepts account for any similar peculiarities in the gap between physical/functional claims P and (second-order) phenomenal claims C about the possession of phenomenal concepts.

Of course, Chalmers may now wish to ask about the relationship between physical/functional claims P and (third-order) phenomenal claims C about the possession of (second-order) phenomenal concepts. But I am happy to go on as long as he is.

5.4 The Second Horn

Let me now turn to the other horn of Chalmers' dilemma. Here P and not-C is not conceivable. We conceive of phenomenal concepts in physical/functional terms, and so conceive of zombies as sharing our phenomenal concepts, in virtue of conceiving them as sharing our physical/functional properties.

[23[23]] For my reasons for thinking that there is nothing semantically or epistemically peculiar about mind-brain identities, and that the impression of a distinctive 'explanatory gap' derives solely from the associated intuition of distinctness, see Papineau 2002 ch 5.

Chalmers' worry on this horn is that phenomenal concepts so conceived will fail to explain our epistemic relationship to consciousness. For we don't conceive of zombies as epistemically related to consciousness as we are, even though on this horn we are conceiving them as sharing our phenomenal concepts. So something more than phenomenal concepts seems to be needed to explain our peculiar epistemic relation to consciousness.

In order to rebut this argument, and show that a physical/function conception of phenomenal concepts allows a perfectly adequate account of our epistemic relation to consciousness, I need to proceed in stages. Observe first that none of the points I have made about our relationship to consciousness in this paper demands anything more than a purely physical/functional conception of phenomenal concepts. To confirm this, we can check that all these points would apply equally to zombies, conceived of as having physical/functional phenomenal concepts, but no inner phenomenology. Note first that the zombies' 'phenomenal' concepts (the scare quotes are to signal that we are not now conceiving of these concepts as involving any phenomenology) will be just as experience-dependent as our own. Zombie Mary will need to come out of her room to acquire a 'phenomenal' concept of red experience, and when she does she will acquire some new non-indexical knowledge: she will come to know that *seeing red =THAT experience* (where this is to be understood as using a material concept on the left-hand side and her new experience-dependent 'phenomenal' concept on the right-hand side). Moreover, this kind of knowledge is arguably unusual, in that it lays claim to an a posteriori necessity, yet it doesn't display the semantic instability characteristic of such claims. Still, zombie Type-B physicalists can invoke the use-mention feature of zombie Mary's new 'phenomenal' concept to explain why that concept should come out as semantically stable even though its possessors can be ignorant of the essential nature of its referent. Not that the zombie Type-B physicalists are likely to have things all their own way, for they will also have to contend with the zombie 'intuition of distictness'—zombies who reflect on the nature of their 'phenomenal' brain-mind claims might well note (using second-order 'phenomenal' concepts) that the left-hand sides 'leave out' a mental property that is used on the right-hand sides, and conclude on this basis that non-'phenomenal' concepts don't mention the same mental properties as are mentioned by 'phenomenal' concepts. Still, zombie Type-B physicalists can point out that this is a confusion, engendered by the peculiar use-mention feature of zombie 'phenomenal' concepts.

All in all, then, everything I have said about our own epistemic relation to our conscious states will be mirrored by the zombies' relation to their corresponding states. I take this symmetry with zombies to show that our own relationship to consciousness can be perfectly adequately explained using a physical/functional conception of phenomenal properties. But Chalmers urges that the comparison cuts the other way. Maybe, he allows, we can suppose that zombies have states to which they stand in the same sort of epistemic relation that we have to consciousness. But we mustn't forget, he insists, that we are also conceiving zombies as beings who lack our inner life, who have no phenomenology. Given this, Chalmers argues that an explanation of mental life that works for zombies can't possibly explain our relation to our own conscious phenomenology.

At this point it will help to recall the basic Type-B physicalist attitude to zombies. Since Type-B physicalists hold that human consciousness is in fact physical, they hold that zombies are metaphysically impossible; any being who shares our physical properties will therewith share our conscious properties; not even God could make a zombie. At the same time, Type-B physicalists recognize that we have two way of thinking about phenomenal properties. This is why zombies are conceivable even though impossible. We can apply one way of thinking about phenomenal matters, but withhold the other—that is, we can think of zombies as sharing our physical/functional properties, but as lacking our phenomenal properties phenomenally conceived.

Given this, there is no obvious reason why Type-B physicalists should be worried that a physical/functional explanation of our epistemic relationship to consciousness will apply equally to zombies. Physical/functional duplicates of us will necessarily be conscious, just like us. True, our ability to think in phenomenal terms makes it possible for us also to conceive of these duplicates as lacking phenomenal properties, and so not related to consciousness as we are. But the fact that we can so conceive of zombies needn't worry the physicalist, who after all thinks that we are here conceiving an impossibility which misrepresents our actual relationship to consciousness. We can imagine beings who are physically/functionally just like us but who lack our inner life—but that doesn't mean that the physical/functional story is leaving something out, given that in reality our inner life isn't anything over and above the physical/functional facts.

Let me conclude by turning to 'silicon zombies'. Here things come out rather differently, since Type-B physicalism leaves it open that silicon zombies are metaphysically as well as conceptually possible. Silicon zombies are possible beings who share our functional properties, if necessary down to a fine level of detail, but who are made of silicon-based materials rather than our carbon-based ones, and on that account lack our conscious properties. As it happens, I am inclined to the view that conscious properties are identical with functional properties rather than strictly physical properties, and that silicon zombies are therefore metaphysically impossible, even though conceivable, just like full-on zombies. However, nothing in Type-B physicalism as I have presented it (nor indeed anything I have written about consciousness) requires this identification of conscious properties with functional rather than physical ones, so I am prepared to concede for the sake of the argument that silicon zombies would lack conscious properties.

Now suppose further that the 'physical/functional' conception of phenomenal concepts which defines the second horn of Chalmers' dilemma is in fact a functional conception. (This seems reasonable—all the non-phenomenal claims I have made about phenomenal concepts have hinged on their functional workings, not their physical nature.) Since silicon zombies are our functional duplicates, they will therefore have 'phenomenal' concepts, functionally conceived, and these will mimic the operations of our own phenomenal concepts: silicon Mary will need to come out of her room to acquire a 'phenomenal' concept of red 'experience', silicon subjects will suffer a 'dualist intuition of distinctness', and so on.

So my putative explanation of our own epistemic relationship to consciousness is mirrored by the matching relationship of the silicon zombies to its corresponding states, even though the silicon zombies are missing the crucial thing that we have—consciousness. And now, since we are dealing with silicon zombies rather than full-on physical duplicates, I can't say that this asymmetry is an illusion generated by our conceiving an impossibility, since by hypothesis the silicon zombies really wouldn't have the conscious properties that we humans possess. Unlike full-on duplicates, silicon zombies really do lack something that we have.

At this point, I think that Type-B physicalists should bit the bullet and say that the thing that differentiates us from the silicon zombies doesn't make any difference to the explanatory significance of phenomenal concepts. We might be related to something different, but this doesn't mean that we enjoy some special mode of epistemological access our states which is not shared by the silicon zombies. After all, the silicon zombies' 'phenomenal' concepts do successfully refer a certain range of silicon mental properties—'schmoncious' properties—and Type-B physicalists can say that the silicon zombies' 'phenomenal' concepts relate them to these schmonscious properties in just the way that our own phenomenal concepts relate us to our conscious properties. True, these schmonscious properties are not conscious ones, since by hypothesis we are supposing that consciousness requires carbon-based physical make-up. But this does not mean that there is any substantial explanatory asymmetry between the way our phenomenal concepts relate us to our conscious states and the way the zombies' 'phenomenal' concepts relate them to their schmonscious states. (After all, note

that silicon zombie philosophers can point out that we lack something that they have, given that we lack the silicon-based make-up required for schomsciousness.)

Of course, if you are a dualist, like David Chalmers, or indeed like anybody who is still in the grip of the intuition of distinctness, then you will hold that there is some very special extra property generated by carbon-based brains, and nothing corresponding in the silicon zombies. And you will think our introspective awareness relates us to this special extra property, and so must involve some special capacity that we do not share with the silicon zombies. But physicalists reject any such special properties, extra to physical/functional ones, and so have no reason to think that our relation to our conscious properties is any different in kind from the silicon zombies' relation to their schmonscious properties. Phenomenal concepts, functionally conceived, provide a perfectly good explanation of both relationships.[24[24]]

Bibliography

Balog, K. Forthcoming. 'The "Quotational" Account of Phenomenal Concepts'.

Bealer, G. 2002. 'Modal Epistemology and the Rationalist Renaissance' in J. Hawthorne and T. Gendler (eds) Conceivability and Possibility. Oxford: Oxford University Press.

Block, N. Forthcoming. 'Max Black's Objection to Mind-Body Identity'.

Chalmers, D, 1996. The Conscious Mind. Oxford: Oxford: Oxford University Press.

Chalmers, D. 2002. 'Does Conceivability Entail Possibility?' in J. Hawthorne and T. Gendler (eds) Conceivability and Possibility. Oxford: Oxford University Press

Chalmers, D. 2003a. 'Consciousness and its Place in Nature', in S. Stich and F. Warfield (eds) The Blackwell Guide to the Philosophy of Mind, Oxford: Blackwell.

Chalmers, D. 2003b. 'The Content and Epistemology of Phenomenal Belief', in Q. Smith and A. Jokic (eds) Consciousness: New Philosophical Perspectives. Oxford: Oxford University Press.

Chalmers, D. 2006. 'Phenomenal Concepts and the Explanatory Gap', in this volume.

Chalmers, D. and Jackson, F. 2001. 'Conceptual Analysis and Reductive Explanation', Philosophical Review 110: 315-61.

Crane, T. Forthcoming. 'Papineau on Phenomenal Concepts', Philosophy and Phenomenological Research.

Harman, G. 1990. 'The Intrinsic Quality of Experience', in J. Tomberlin (ed.) Philosophical Perspectives 4.

Horgan, T. 1984. 'Jackson on Physical Information and Qualia', Philosophical Quarterly 34.

Jackendorf, R. 1987. Consciousness and the Computational Mind. Cambridge, Mass: MIT Press.

Jackson, F. 1986. 'What Mary Didn't Know', Journal of Philosophy 83.

[24[24]] Earlier versions of this paper were read at the Tilburg workshop on Mind and Rationality in August 2003, at the Jowett Society in Oxford in October 2003, and to the King's College Departmental Seminar in January 2004. I would like to thank all those who commented on those occasions.

Jackson, F. 1998. From Metaphysics to Ethics. Oxford: Oxford University Press

Kripke, S. 1980. Naming and Necessity. Oxford: Blackwell.

Loar, B. 1990. 'Phenomenal States', in J. Tomberlin (ed.) Philosophical Perspectives 4.

Melnyk, A. 2003. Contribution to symposium on Thinking about Consciousness [website address to follow]

Millikan, R. 1990. 'The Myth of the Essential Indexical', Nous 24 723-34

Millikan, R. 2000. On Clear and Confused Ideas. Cambridge: Cambridge University Press.

Papineau, D. 1993a. Philosophical Naturalism. Oxford: Basil Blackwell.

Papineau, D. 1993b. 'Physicalism, Consciousness, and the Antipathetic Fallacy', Australasian Journal of Philosophy 71.

Papineau, D. 2002. Thinking about Consciousness. Oxford: Oxford University Press.

Perry, J. 1979. 'The Problem of the Essential Indexical', Nous 13:3-12.

Perry, J. 2001. Knowledge, Possibility and Consciousness. Cambridge, Mass: MIT Press.

Prinz, J. 2000. 'A Neurofunctional Theory of Visual Consciousness', Consciousness and Cognition 9: 243-59.

Prinz, J. 2002. Furnishing the Mind. Cambridge, Mass: MIT Press.

Stoljar, D. Forthcoming. 'The Argument from Diaphanousness', in R. Stainton and and M. Escurdia (eds), Language, Mind and World. Oxford: Oxford University Press.

Tye, M. 1995. Ten Problems of Consciousness. Cambridge, Mass: MIT Press.

Tye, M. 2003. 'A Theory of Phenomenal Concepts', in A. O'Hear (ed.) Minds and Persons. Cambridge: Cambridge University Press.