

SWIF Philosophy of Mind, Volume 4, 2003.

Forum on David, Papineau, *Thinking about Consciousness*

The Book

[Details from the publisher's site.](#)

[Table of Contents, Introduction, and Index](#) (PDF)

Andrew Melnyk

[Papineau on the Intuition of Distinctness](#)

Robert Kirk

[Thinking about Papineau's Thinking about consciousness](#)

David Papineau

[Reply to Kirk and Melnyk](#)

Notes on contributors

David Papineau ([Homepage](#)) is Professor of Philosophy of Science at King's College, University of London. His books include *For Science in the Social Sciences* (1978), *Theory and Meaning* (1979), *Reality and Representation* (1987) and *Philosophical Naturalism* (1993).

Robert Kirk ([Homepage](#)) is Emeritus Professor of Philosophy the Department at the University of Cottingham. His publications include *Translation Determined* (1986), *Raw Feeling* (1984), *Relativism and Reality* (1999).

Andrew Melnyk ([Homepage](#)) is Associate Professor of Philosophy and Director of Undergraduate Studies at the University of Missouri-Columbia. He is the author of *A Physicalist Manifesto: Thoroughly Modern Materialism* (2003) a numerous papers in philosophy of mind.

This is a online publication for the Sito Web Italiano per la Filosofia - SWIF - Italian Web Site for Philosophy Edizioni Digitali di Filosofia - Registration ISSN n. 1126-4780. Unless otherwise specified, the Copyright © of the articles belongs to the authors.

Papineau on the Intuition of Distinctness

Andrew Melnyk

Department of Philosophy
University of Missouri

[Homepage](http://melnyka@missouri.edu) - melnyka@missouri.edu

In his new book, amid other good things, David Papineau argues with exemplary clarity and force (i) that no extant philosophical argument succeeds in showing that phenomenal states and physical states are distinct, and (ii) that at least one empirical consideration suggests that they are not (Papineau 2002, Chs. 1, 2, 3, and 5). But he insists nonetheless that all of us – physicalists included – have what he calls the 'intuition of distinctness' to the effect that they are distinct. Moreover, this intuition does not merely generate some sort of nagging doubt about materialism. According to Papineau, the intuition of distinctness "stops us *really* believing the materialist identification of mind with brain, even those of us who profess materialism" (p. 94; italics original). So Papineau holds that (1) none of us *really* believes materialism about phenomenal states because (2) we feel the intuition of distinctness.

Claim (1) is a very strong one, and Papineau does little to argue for it, but it is plausible and deserves to be taken seriously, as it will be below, even though it is not clear what "really believing" something amounts to, or how it is related to regular believing or to one's subjective probabilities. And (1) cannot be refuted simply by noting that materialists profess materialism. For materialists may profess materialism because (i) they follow the policy of professing what they believe they *ought* to believe and (ii) they believe, having weighed up the evidence, that they ought to believe materialism; but neither (i) nor (ii) requires that they actually believe materialism. Claim (2) is even more plausible, for what materialist will deny that phenomenal properties *seem* utterly distinct from physical properties of any kind? And explaining claim (1) by appeal to claim (2) is an attractively economical next step.

But what explains claim (2)? In chapter 6, Papineau undertakes the important task of explaining why we feel the intuition of distinctness. The task is important because so long as the intuition is unexplained in some fashion compatible with materialism we will be tempted to explain it by supposing that it reflects a more or less dim recognition of what might some day be articulated as a sound argument against materialism. And, of course, one possible materialist approach to explaining the intuition of distinctness is to diagnose it as arising from a dim recognition of some already articulated but unsound argument against materialism. But Papineau does not adopt this approach, arguing against it in the earlier sections of his chapter.

His preferred account appeals to what he calls the *antipathetic fallacy*. According to Papineau, even though phenomenal states are physical states, and can therefore be thought of via the exercise of third-personal concepts drawn from the neurosciences or functionalist psychology, we also have a special first-personal way of thinking about them that is available only to those who have actually undergone them. These first-personal concepts – phenomenal concepts – are special in that, when they are exercised to think about an experience, "the experience itself is in a sense being *used* in our thinking, and so is present in us" (p. 170; italics original). Third-personal concepts – material concepts – are, of course, not like that. Now the antipathetic fallacy allegedly arises when we notice that our material concepts do not *use* the experiences they supposedly refer to, and then infer that material concepts do not therefore *mention* those experiences; the antipathetic fallacy is thus "a species of use-mention fallacy" (p. 171).

But I am not yet convinced that it is our commission of the antipathetic fallacy -- at least as I have stated it so far -- that explains the intuition of distinctness. An immediate worry is that since, in order to explain the intuition, one must attribute fallacious reasoning to those who have it, it follows, since Papineau himself confesses to having the intuition, that one must charge Papineau himself with committing the antipathetic fallacy. But it is rather hard to believe that this charge could be true of someone as unmuddled and logically sophisticated as Papineau, especially since he is the one who has pointed out the fallacy and exposed its fallacious character so clearly!

This worry is not decisive, however, since it is possible that humans should be constitutionally prone to engage in fallacious reasoning of a certain sort, that they should be capable of fully understanding that the reasoning is fallacious, and yet that they should be quite incapable of making appropriate modifications to their more primitive dispositions to reason. An analogy for this possibility is provided by the notorious recalcitrance of perceptual illusions: I may be as certain as I am of anything that the two lines in the Müller-Lyer illusion are of equal lengths (perhaps because I have measured them with a ruler) while remaining quite unable to stop myself, at some relatively primitive level of perceptual representation, from representing them as unequal. The question, therefore, is whether a proneness to commit the antipathetic fallacy is an actual example of this sort of possibility.

Ironically, Papineau's discussion provides evidence that it is not. For, in showing that the antipathetic fallacy is indeed fallacious, he points out that, for most concepts, one's failing to *be* F does not entail that one is failing to *think* about being F, since one's thinking about being F (e.g., sick) does not require one's being F. But, I wish to add, nobody claims that it does; we simply feel no *general* temptation at all to commit the antipathetic fallacy. It looks like a poor candidate, then, for a cognitive tendency so powerful that we cannot eradicate or override it. Indeed, although Papineau calls the antipathetic fallacy "terribly natural" (p. 170), he offers no reason, beyond noting that it can be seen as a sort of use-mention fallacy, to think that people are much inclined to commit it at all. And how could they be, if they only ever commit it in connection with a tiny minority of the concepts they possess?

Now my account of Papineau's appeal to the antipathetic fallacy to explain the intuition of distinctness, while faithful to its letter, has nonetheless been unfaithful to its spirit. For my account has represented the fallacy as being committed by those who reflect *only* upon their deployment of *material* concepts (for phenomenal states): they notice that in deploying a certain material concept they are not thereby *in* pain, and they conclude that therefore their concept is not *of* pain. But it is clear that, for Papineau, commission of the fallacy somehow involves reflection *also* upon one's deployment of *phenomenal* concepts (for phenomenal states), and indeed upon an invidious *comparison* between one's deployment of phenomenal concepts and one's deployment of material concepts. But how, exactly, might we bring reflection upon one's deployment of phenomenal concepts into the explanation of the intuition of distinctness?

Papineau does not say. But one idea might be that reflection upon one's deployment of phenomenal concepts, and in particular one's noticing that such deployments involve being in (something like) a phenomenal state, generates a certain expectation, namely, that *all* ways of thinking about phenomenal states must involve being in (something like) a phenomenal state; but then someone with such an expectation, who subsequently noticed that his or her deployment of a material concept did *not* involve being in (anything like) a phenomenal state, would be quite right to infer that the material concept was not of a phenomenal state.

But I have two concerns with this idea. The minor concern is that it requires that people have the capacity to reflect upon their thoughts about phenomenal states while they are thinking those very thoughts: people must be able to think phenomenally about pain (say), and hence, on Papineau's account, simultaneously be in pain (or a pain-like state), while they are also thinking *about* their thinking (phenomenally) about pain, and while they are thinking, indeed, *that* their thinking (phenomenally) about pain involves being in pain (or a pain-like state). I have no particular reason to doubt that people do possess this mental capacity, but it is surely a very sophisticated one. The major concern with the idea of the last paragraph is that it requires that people jump to a universal conclusion about *all* ways of thinking about phenomenal states on the strength of what is observed to be true about *one* way of thinking about phenomenal states. But why would they do that? Inductions from a single case *can* be good, given the right background knowledge, but the single-case induction being postulated here does not appear to be an example. So the question remains why people would engage in it, and in particular why here, and it is too ad hoc to answer that they just do. And there is a further question: even if we are indeed constitutionally prone to employ the single-case induction in this sort of circumstance, how come we cannot reject its conclusion when we consider it in light of our *total* evidence? It still remains to explain our inability to *believe* the conclusion.

So it is not clear how to bring reflection upon one's deployment of phenomenal concepts into an adequate explanation of the intuition of distinctness. And yet we surely do all feel the intuition, so it does need an explanation, and, for any materialist, an explanation consistent with a materialist account of phenomenal states. What, then, is to be done?

My hunch is that there is still life in a fascinating suggestion that Papineau makes only to dismiss. The suggestion is that we literally cannot believe identity claims framed using a phenomenal concept and a material concept, and that we cannot do so because believing identity claims in general is a matter of something like *mental file-merging*, and, in the special case of identity claims framed using a phenomenal concept and a material concept, our cognitive architecture prevents such file-merging from taking place (p. 165). And, though Papineau does not say this, our hypothesized inability to believe that pain (say) is a material state would be quite consistent with our believing that we ought to believe that pain is a material state.

But Papineau, as I say, dismisses this suggestion. Let me conclude by explaining why I think he does so prematurely. His reason, in a nutshell, is that (premiss 1) if there were some cognitive-architectural obstacle to file-merging across the phenomenal-material conceptual divide, then, especially given the close connection alleged in chapter 4 between phenomenal concepts and perceptual concepts, there ought equally to be a cognitive-architectural obstacle to file-merging across the perceptual-theoretical conceptual divide; but (premiss 2) there is no such obstacle in the perceptual-theoretical case – for example, we have no trouble believing that visually-conceived kestrels are identical with theoretically-conceived kestrels (pp. 165-167).

I think there may be room to doubt premiss 2 here, since it might be suggested that how much trouble one has in believing that visually-conceived kestrels are identical with theoretically-conceived kestrels depends upon how one thinks of the *secondary qualities* of the visually-conceived kestrels. For if it is insisted that the colors (say) of a kestrel be treated as entirely intrinsic to the bird, then ordinary people might well bridle at the identification of visually-conceived kestrels with theoretically-conceived ones. But if, by contrast, the kestrel's colors are permitted to reside, at least partially, in the minds of observers, then, though ordinary people may now willingly accept the identification of visually-conceived with theoretically-conceived kestrels, the suspicion will be raised that identifications across the perceptual-theoretical divide are unproblematic only if identifications across the phenomenal-material divide are unproblematic – the very claim at issue.

Premiss 1 may be independently objectionable. For there is a possible reason why file-merging across the phenomenal-material divide should be unachievable even if file-merging across the perceptual-theoretical divide is not. The reason I have in mind is that *one* kind of phenomenal concept seems to be usable only to refer to a phenomenal state as one undergoes it ("*That* is going on in me now"), and *not* to be usable to *re-identify* a phenomenal state, not even to re-identify it as *one of those again*. Now if phenomenal concepts of this kind exist, and if concepts in general can be viewed as analogous to files, then a phenomenal concept of this kind will constitute a file that is only temporary, a file that persists only as long as one is undergoing the experience it picks out. But any file corresponding to a material concept will presumably be permanent; at the very least it will permit the re-identification of whatever it picks out. And, on the not too implausible assumption that no temporary file can be merged with a permanent file, it follows that no phenomenal concept of the kind in question can be merged with a material concept, and hence, if believing identity claims is a matter of mental file-merging, that no identity claim framed using a phenomenal concept of the kind in question and a material concept can be believed.

So Papineau's fascinating suggestion seems to me eminently worthy of further exploration. It promises at least to explain why it is during intense episodes of introspection – as we think to ourselves, "*That* couldn't be a brain state!" -- that we are most strongly inclined to doubt materialism about phenomenal states.

REFERENCES

Papineau, David. 2002. *Thinking About Consciousness*. New York, NY: Oxford University Press.

Thinking about Papineau's *Thinking about consciousness*

Robert Kirk

Department of Philosophy
University of Nottingham
Robert.Kirk@nottingham.ac.uk

'We don't need any fancy new concepts to understand consciousness', David Papineau says on the first page of this challenging book, 'For there isn't anything really mysterious about it in the first place'. He concedes there *seems* to be a mystery, but holds that the puzzles 'would simply dissolve if we fully accepted that conscious states are one and the same as brain states' (2). However, 'something stops us embracing such identities', and his book 'is an attempt to understand this dualist compulsion, and free us from its grip' (3). This 'something' is an 'intuition of distinctness': the powerful idea that states of phenomenal consciousness just cannot be understood as physical states. The key to understanding this compulsion is to be found in a correct grasp of the difference between material concepts and the concepts in terms of which we introspect and think about our conscious experiences: 'phenomenal concepts'. 'By carefully analysing the workings of ... phenomenal concepts', he says, 'I am able to explain why it should seem so obvious that conscious states are distinct from material states, even though in reality they are not'. So his book is not just about consciousness, but about 'the special ways in which we *think* about consciousness' (5).

Papineau defends some very special, sometimes surprising, approaches to philosophical problems in this area. Although I agree with a great deal, I also find some of his central claims puzzling. One problematic cluster of ideas for me is his distinctive account of the character of phenomenal concepts and the 'divide' between them and other concepts. Another is what he calls the 'explanatory asymmetry' between psycho-physical and other scientific reductions. A third centres on his diagnosis of 'the intuition of distinctness'. Finally there is his rejection of what he calls 'the a priori characterisation' of materialism. All these topics are intertwined, but as far as possible I will take them in turn.

1. *The divide between psychological and phenomenal concepts*

Papineau holds that our conscious properties are *identical* with certain 'material' properties. Material properties are picked out by 'material concepts', typically by association with their causal roles. But material concepts are not just the ones we use to characterise material things; they include mental or psychological concepts. The ones that concern us 'pick out conscious properties as items in the third-personal, causal world. Most commonly, these will be role concepts, ... which refer by describing some causal or other role, such as pain's role in mediating between bodily damage and avoidance behaviour' (48). Following David Chalmers, Papineau calls this special class of material concepts 'psychological concepts', and it is in this sense that I will use the word 'psychological' from now on (see Chalmers 1996, 11-32).

So whenever and however we think about our conscious experiences, we are in fact thinking about material properties. But Papineau thinks there is also a special class of concepts, 'phenomenal concepts', which enable us to think about exactly the same material properties in a very different way. 'When we use phenomenal concepts, we think of mental properties, ... in terms of *what they are like*' (48). This way of thinking of them involves re-creating past conscious experiences in imagination, or classifying the experiences we are actually having. He argues that Jackson's famous 'knowledge argument' forces us to accept that there are indeed these two kinds of concepts, a view which Papineau calls 'conceptual dualism'. (It seems safe to assume readers are familiar with the knowledge argument: Jackson 1982.) Jackson himself held at the time that the argument showed not just that there are two kinds of *concepts*, but that there are also two kinds of *properties*. Papineau, as a materialist, maintains that the argument does no such thing - but that it does force us to accept that Mary, when she comes out of her grey prison, acquires new concepts: phenomenal concepts. She becomes able to use these concepts in sentences such as 'So *this* is what people experience when they look at ripe tomatoes'. (These views make Papineau what he calls, following Block, an 'inflationist' about phenomenal concepts.)

The special phenomenal concepts are not only taken to be concepts of states that *feel* a certain way. In addition they are taken to refer 'directly' to their referents; that is, they do not refer via some description or other, or by means of causal roles, as 'material' and 'psychological' concepts do. Papineau's account of these special phenomenal concepts plays a central role in the book. Notably, the point that 'uses of phenomenal concepts will standardly be *accompanied* by versions of the experiences referred to ... will provide the crucial ingredient' for his explanation of the 'intuition of distinctness' (105).

So far his points about phenomenal concepts may seem unremarkable, and consistent with enlightened materialism. But he goes on to make claims that I find both surprising and puzzling. It turns out that his point is not that *the concept of pain*, for example, is one of these special phenomenal concepts. That would misrepresent his position. He thinks we have *two* concepts of pain. One of these is indeed a phenomenal concept, but the other is 'psychological' (in the sense explained above). So on his account, expressions like 'pain' should be viewed 'as simultaneously expressing both sorts of concepts' (97); 'the term "pain" does indeed express two conceptually independent notions, phenomenal and psychological' (99). As things are, these two concepts refer to the same (material) property; but this is a contingent fact. And according to Papineau 'it is an entirely a posteriori matter whether the phenomenal and psychological concepts associated with everyday phrases like "pain" ... refer to the same property' (100).

Are there really these two concepts of pain? Supposing there are, is it really 'an entirely a posteriori matter' that they both refer to the same property? I was not persuaded.

Papineau is emphatic that materialism cannot be established a priori; and I for one would agree. It is indeed a posteriori whether or not such and such a physical process underlies, realises, or is identical with, such and such a phenomenal state: empirical investigations are needed to find out what, physically, is going on. But that is one thing; it is something else to claim, as he does, that there is 'no a priori route to the identification of their referents' (49). However, this assertion is not as clear as it might be. For one thing, an 'a priori route' might go in either of two directions: from or to the physical.

I don't think anyone would claim we could get a priori *from* conscious state *to* physical underpinnings. However, as he is well aware, plenty of philosophers hold that we could in principle go a priori in the other direction: from physical facts to conscious facts. He makes a point of denying this claim, which I will discuss later (section 4). At this stage I want to draw attention to some of the assumptions he seems to think support it. One is:

(A) 'If phenomenal and material concepts are quite distinct at the level of sense, there will be no a priori route to the identification of their referents' (49).

Taken in isolation that may seem reasonable. The concepts *water* and *H₂O*, for example, are distinct at the level of sense; it took empirical work to establish that nevertheless they do have the same referents. However, there is more to be said.

To start with, difference in sense cannot by itself be enough to entail there is no a priori route to identifying referents. '13' and 'the cube root of 2,197' have different senses; but we can discover a priori that they have the same referent. Papineau goes on to say that 'examination of the concepts themselves will not tell us that they refer to the same properties. Such knowledge can only be arrived at a posteriori, on the basis of empirical evidence about their actual referents' (49). Here we need to respect two distinctions.

It is one thing to say:

(B) Merely knowing all the relevant physical truths would not by itself equip even a sufficiently intelligent creature with phenomenal concepts, or enable it to work out a priori when, where, and which phenomenal concepts applied to the organisms whose existence those truths entailed.

It is something else to say:

(C) If a sufficiently intelligent creature *already possessed* the necessary phenomenal and related concepts, then knowing all the relevant physical truths would not enable it to work out a priori when, where, and which phenomenal concepts applied to the organisms whose existence those truths entailed.

I think Papineau's discussion makes it too easy to overlook the distinction exemplified by (B) and (C). If (A) were to be understood in the sense indicated by (B) I would have no objection to it. To deny it would seem to imply that necessary and sufficient conditions for phenomenal concepts are formulable in purely physical terms; which few today would be willing to maintain. In any case I don't see how that interpretation could fit in with his aims. If on the other hand (A) is to be understood in the sense indicated by (C), I don't think he has justified it. (More on this shortly.) The other distinction I think his discussion tends to slide over is between what we might call a superficial mastery of the concepts in question and a deep understanding of them and their interconnections. Someone with a mastery of the concept represented by the numeral '13' would not thereby be automatically equipped with mastery of the concept represented by 'the cube root of 2,197'. Further, someone with a mastery of both concepts would not necessarily come to realise that they both picked out the same number. Yet of course there is, in this case, 'an a priori route to the identification of their referents'.

If Papineau's claim in (A) were to be understood simply on the basis that a superficial mastery of the concepts involved is not enough to ensure that their users know that they have the same referent, then we would have to accept (A). However, he clearly cannot mean it in that sense, since he uses (A) and similar claims to support his rejection of the 'a priori characterisation of materialism' - a characterisation which obviously requires the more demanding sense of 'a priori route'.

Papineau asserts that 'the term "pain" does indeed express two conceptually independent notions, phenomenal and psychological' (99). But for the reasons sketched above, I am not persuaded that it is 'an entirely a posteriori matter' that these concepts both refer to the same property. If I am right, then even if there are two concepts of pain, they are not as independent as he claims. But is there now any good reason to say we have two concepts of pain?

We surely do have two very different ways of *using* our concept of pain. Using it to ascribe pain to others on the basis of their behaviour is one; using it to ascribe pain to ourselves on the basis of actually feeling or imagining pain is another. But that does not seem enough to justify the claim that there are two concepts. After all (to use an old example of David Armstrong's) the concept of *gene* was originally used on the basis of theoretical explanations, and only much later became usable on the basis of microscopic observation; yet that seems no basis for saying there are two concepts of gene. Papineau compares his claim about there being two concepts of pain with 'multi-criterial concepts' in science (98-100). But that comparison depends on his assumption that there are no a priori links from the physical to the phenomenal; so I don't think he could appeal to it in support of his claim. In any case the claim that there are two concepts rather than one starts to look merely terminological; much more important is the question of a priori connections.

Although he thinks there is no a priori route to identifying the referents of the alleged two distinct concepts of pain, he thinks there *is* an a priori route from causal roles and related 'material' facts to psychological truths. This is because he holds that psychological concepts are 'associated with causal roles mediating between canonical perceptual inputs and behavioural displays' (179). For that reason he claims 'it is an a priori matter which causal roles are associated with which mental terms' (179f.). He is therefore happy to concede that *if* a zombie world were possible (which he rightly thinks all materialists must deny), the zombies would nevertheless be describable in terms of the 'psychological' concepts. It is only the phenomenal concepts that would not apply to them. I will return to that point in section 4 below. Meanwhile let me note that I found nothing in Papineau's book which persuaded me to abandon a broadly functionalist approach to phenomenal consciousness. In this connection it is pertinent to draw attention to something that opponents of functionalism seem to overlook.

They often object that the performance of functions could not possibly constitute phenomenal consciousness because functions are a matter of *relations* between, for example, inputs, outputs, and the system's internal states, whereas phenomenal consciousness is a matter of events which: occur independently of their relations with the rest of the world; have their characters 'intrinsically' (see for example Levine 2001); and, as Papineau says, can be referred to 'directly' in the sense that the referring does not have to go via descriptions of causal roles. Functionalists can reply by stressing the following three considerations (which I think hold good regardless of the details of the brand of functionalism that is in question). First, while it is of course a relational matter whether a given function is being performed, its performance is itself an actual event. Second, the fact

that a given event is describable in relational terms does not prevent its also being describable in non-relational terms (that is my *cousin*, but also an *organism*). Third, and most important, functionalists may legitimately point out that while organisms and their internal states can be described in relational terms, they may acquire their own special concepts, different from those which observers might employ - concepts which may be suitable for thinking about the character of their experiences from their points of view. Suppose, as Papineau seems to concede, that the physical events constituting or identical to the experience of pain can be picked out in broadly functional terms. Functionalists can maintain that these same physical events can also be picked out *directly*, by means of the subject's special 'first-person' concepts. These concepts will no doubt *seem* to be insulated from the psychological facts and relations which nevertheless, if the functionalists are right, constitute the events in question and ensure that the organism is phenomenally conscious. On that basis functionalists can claim there is a clear enough sense in which the events in question have their own 'intrinsic' character, and can be referred to directly. There is no necessity for the subject to consider the relations between those events and the rest of the world in order to be in a position to judge either that they are happening, or what their subjective character is.

I don't think Papineau's reasoning undermines such approaches. If that is right, parallel reasons would seem to justify the view that even if there are phenomenal concepts, and even if they are not linked a priori in any obvious or direct way with causal or other functional concepts, they may still be linked in indirect, unobvious ways. (For further suggestions see Kirk 1996,1999.)

2. *The 'explanatory asymmetry' between psychophysical and other scientific reductions*

The situation which Joseph Levine calls the 'explanatory gap' is summarised in his own words as follows: 'On the one hand, we have excellent reason for thinking that conscious experience must be reducible, in the requisite sense, to a physical phenomenon, and, on the other hand, we don't see how it could be' (2001, 175). We can understand how it is that heat, for example, is molecular motion, or how it is that water is H₂O. But when it comes to understanding how the phenomenal is physical, we are at a loss. Discussing this alleged gap, Papineau concedes that 'there is indeed a kind of explanation which is not delivered by materialist reductions of conscious properties', and that this 'marks a contrast with materialist reductions in other areas of science' (142). But he argues that the problem is not that we can't find an explanation which is there to be found. It is that we can't resist seeking an explanation where none is either needed or possible. In his view there is no gap of the sort Levine discerns.

One thing we cannot do, Papineau points out (with what seemed to me excessive emphasis), is to explain why a thing is identical with itself. 'Identities need no explaining' (150). And certainly, if that had been what we were looking for in the psycho-physical case, we would have been wasting our time. But he recognises that that is not what we are looking for. What we want to explain is how, given that such and such purely physical descriptions hold, the phenomenal descriptions hold as well. What is it *about* the purely physical facts which guarantees that these phenomenal descriptions also apply? To many philosophers that seems to be a perfectly proper question, which we keenly want answered.

But Papineau, if I understand him rightly, thinks that it too is a mistaken question. He attempts to explain 'the basic point at issue' by comparing the case of proper names. Cicero is Tully; but if we ask what it is about this individual which ensured that Cicero is identical with Tully, there is no answer. Similarly for the identity of phenomenal properties with physical ones, he maintains. They are 'brute': we just have to accept that we are talking about *the same thing* in two different ways, and there is no more to be said. The reason is that 'phenomenal concepts are not associated with roles' (159).

You may now be starting to suspect that Papineau, having claimed 'there isn't anything really mysterious' about phenomenal consciousness, has actually introduced a big mystery with these 'brute' identities. Is the psycho-physical case relevantly like that of Tully and Cicero? True, if the materialist is right, there are in each case two different expressions referring to a single item. But the conditions for giving a man a name are easily understood and easily satisfied, while that is not so for the case of applying phenomenal concepts to physical properties. It certainly seems reasonable to wonder what it is about some physical properties which ensures that the phenomenal descriptions apply to them, while other physical properties do not qualify for such descriptions. And, of course, if the suggestions at the end of the last section are anything like correct, the question is in place after all. For if that

suggestion is right, phenomenal concepts *are* 'associated with roles', even if actually acquiring and using them does not require the subject to know what those roles are.

However, Papineau does maintain that it's a mistake to look for an explanation of the kind Levine and others say is needed and unavailable, just as much as it is a mistake to look for an explanation of why Tully is Cicero. What makes us hanker after such an explanation - of a kind we can't possibly have if he is right - is the 'intuition of distinctness' (145). He and other materialists 'will surely admit that they sometimes hanker for some further understanding of *why* brain activities should yield conscious feelings. ... [but this] is simply because mind-brain materialism is so hard to accept in the first place. The real fly in the ointment is the "intuition of distinctness" ...'. 'This arises quite independently of any questions of what materialism might or might not explain' (145). We had better examine his diagnosis of this intuition.

3. *Diagnosis of the 'intuition of distinctness'*

Papineau's diagnosis exploits his account of the special nature of phenomenal concepts. The core of that account is the idea that using phenomenal concepts to refer to conscious experiences essentially involves either imagining past experiences, or introspecting present ones, when both types of activity include actually having conscious experiences. Material concepts, in contrast, do not essentially involve having conscious experiences. He suggests that this contrast explains the power of the intuition that what phenomenal concepts refer to is essentially different from what any material concepts refer to. Using phenomenal concepts to refer to conscious experiences involves actually having experiences - 'exercising a phenomenal concept will *feel* like having the experience itself' - while using material concepts does not. So 'it is all too easy to slide ... to the conclusion that, in exercising material concepts, we are not thinking *about* the experiences themselves. ... This line of thought ... is what lies behind the inescapable conviction that the mind must be extra to the brain' (170).

I have already indicated why I am not persuaded that the two allegedly different kinds of concepts are as unconnected as Papineau makes out. For that reason I am sceptical about his explanation of the intuition of distinctness. But even supposing for argument's sake that he is right about phenomenal concepts, I don't see how that feature of phenomenal concepts could be the main explanation of our reluctance to accept that consciousness involves nothing beyond the physical.

Let us return briefly to the topic of the last section: the claim that the intuition of distinctness explains the feeling that there has to be some *explanation* of why phenomenal concepts apply to some of the items picked out by material concepts. Does that intuition really have a tendency to reinforce that feeling? Perhaps. But I have to say that, while of course I am not immune to the power of that intuition, I myself have long been persuaded that nothing beyond the physical is involved in our being phenomenally conscious - yet I *still* think there is an explanation of the kind Papineau claims to be in principle unavailable. For that reason I don't see how that intuition can be the main thing preventing us from 'realising' that no such explanation is possible.

4. *Papineau's rejection of 'the a priori characterisation' of materialism*

Papineau emphasises the importance of explaining exactly what is to be understood by 'materialism': is it 'type identity, token identity, realization, or supervenience' (8)? His preferred formulation is in terms of identity. He holds that conscious properties are identical either with first-order physical properties or with higher-order properties that are realised by first-order physical properties. He is wary of formulations in terms of supervenience, a notion he thinks is 'more trouble than it's worth ... any benefits offered by the notion of supervenience are more easily gained simply by identifying mental properties directly with higher-order properties or disjunctions of physical properties' (36). As he points out, supervenience will only serve to formulate a genuine materialism if it holds 'across all metaphysically possible worlds' (37). (As it happens, I have defended something close to that very view. However, I agree that current notions of supervenience are confusing, and have argued that strict implication does a better job: Kirk 1994, 1996, 2001.)

He also rejects what he calls the 'a priori characterisation of materialism'. This is that 'materialism is equivalent to the view that all truths - including all truths about the mind - follow a priori from the physical facts' (152). Now, in common with many materialists and also with dualists like David Chalmers, I think any materialist is committed to something like

that view whether they like it or not; so I am particularly interested in Papineau's rejection of it (Chalmers 1996, 1999; Jackson 1998). He accepts that it is at least plausible that such a view holds in the case of other scientific reductions: 'If there is a material reduction of *water*, then a full physical description of the world, plus the conceptual knowledge that water is the stuff that plays a certain role, will enable us a priori to identify which material kind reduces water'. And we have seen that he also accepts that a full physical description of the world will enable us a priori to discover the 'psychological' facts. But he rejects the a priori characterisation of materialism on the ground that 'phenomenal facts *cannot* be inferred a priori from a full physical description of the world' (153). This, he says, follows from the 'explanatory asymmetry' discussed in the last section: the view that while the references of material concepts are typically fixed by causal roles, no such roles are available to fix the references of phenomenal concepts. I explained in section 1 why I am not persuaded by his reasoning, which is encapsulated in the following remarks:

(D) 'The lack of any associated role to fix the reference of "phenomenal pain" stops us from inferring facts about phenomenal pains a priori from physical facts about brains. Suppose you know everything there is to know about brain activities, and about the typical physical causes and effects of those activities. This won't enable you to figure out a priori that certain brain states *feel* a certain way' (153).

To the contrary, I think a broadly functionalist approach, exploiting the way in which conscious organisms can refer directly to the events which constitute their conscious experiences, enables us to see that facts about 'phenomenal pains' are not isolated from facts about 'psychological pains' in the way he assumes.

At first you might infer from (D) that Papineau thinks a 'zombie world' is possible: a world sharing all its purely physical features (physical laws included) with the actual world, but totally lacking in phenomenal consciousness. Indeed some of his remarks might lead unwary readers to suppose he really does think that. For example he remarks that '... there doesn't seem anything metaphysically incoherent about creatures who are physically just like us, down to their nociceptive-specific neurons, but who have no feelings of pain' (77). However, he has a whole chapter entitled 'The Impossibility of Zombies', and argues, on the basis of Kripkean views about the necessity of identity, that '... materialists must deny that such things really are possible. So they need to say that zombies and ghosts are a kind of modal illusion' (79).

Still, there does seem to be a tension between (D) and the denial that such things as zombies are possible. If (D) holds, how can there fail to be a possible zombie world? Evidently Papineau leans on Kripke's views about a posteriori necessary identities. However, a line of attack on that approach has been developed by Chalmers and Jackson. It is deployed in terms of a semantic framework which distinguishes two types of intensions or meanings for key expressions (Chalmers 1996, 1999; Jackson 1998). It would have been interesting to know Papineau's response to such objections. But in any case it is possible to press the challenge without appealing to that 'two-dimensional' semantic framework. For unless there is some contradiction or incoherence in the description of a zombie world, in particular where there are episodes of 'psychological pain', it seems there must be more in the *actual* world than is provided for by the purely physical facts, in which case some kind of dualism is true (see Kirk 1994, 1996, 2001). Even if such contradictions or incoherences are hard to discover, I don't think Papineau has shown that there are none - for the reasons noted at the end of section 1 above.

I have been critical because that is the commentator's job. But I am strongly in favour of many features of Papineau's approach, and have simply ignored large areas where I find his treatments persuasive, instructive, and impressive - notably on the causal argument for materialism; on the relations between perceptual imagination and phenomenal concepts; on the argument from knowledge; on 'deflationist' attempts to meet that argument; on the methodology of empirical research into consciousness; and on the history of the principle of the completeness of physics. Let me end by saying how much I have enjoyed reading and being provoked by this stimulating book, and how much I have appreciated Papineau's refreshingly no-nonsense style.

REFERENCES

Chalmers, David J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York and Oxford: Oxford University Press.

Chalmers, David J. 1999 "Materialism and the Metaphysics of Modality." *Philosophy and Phenomenological Research* 59: 475-496.

Jackson, Frank. 1982. "Epiphenomenal Qualia." *Philosophical Quarterly* 32: 127-36.

Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Clarendon Press.

Kirk, Robert. 1994. *Raw Feeling*. Oxford: Clarendon Press.

Kirk, Robert. 1996. "Strict Implication, Supervenience, and Physicalism." *Australasian Journal of Philosophy* 74: 244-256.

Kirk, Robert. 1999 "Why There Couldn't Be Zombies." *Proceedings of the Aristotelian Society*, supplementary vol. 73: 1-16.

Kirk, Robert. 2001. "Nonreductive Physicalism and Strict Implication." *Australasian Journal of Philosophy* 79: 545-553.

Levine, Joseph. 2001. *Purple Haze: The Puzzle of Consciousness*. Oxford and New York: Oxford University Press.

Reply to Kirk and Melnyk

David Papineau

Department of Philosophy
King's College London

I am lucky to have two such penetrating commentators as Robert Kirk and Andrew Melnyk. It is also fortunate that they come at me from different directions, and so cover different aspects of my book. Robert Kirk has doubts about the overall structure of my enterprise, and in particular about my central commitment to a distinctive species of phenomenal concepts. Andrew Melnyk, by contrast, offers no objections to my general brand of materialism. Instead he focuses specifically on my discussion of the anti-materialist 'intuition of distinctness', raising questions about my attempt to explain this intuition away, and offering alternative suggestions of his own.

Let me first discuss Robert Kirk's comments. This will enable me to clarify some of the central themes of the book. After that, I shall turn to Andrew Melnyk's helpful comments on the intuition of distinctness.

Kirk's Comments

Kirk is unhappy with the 'inflationist materialism' that underpins the overall argument of my book. An 'inflationist' thinks that there are special ways of thinking about conscious states—using phenomenal concepts—which are a priori distinct from all other ways of thinking about conscious states, and in particular from functionalist ways of thinking about conscious states. Because of this, inflationist materialists will deny that materialism requires all true claims to follow a priori from the physical truths (that is, they will reject the 'a priori characterization of materialism'). Of course, inflationist materialists will allow that some kinds of non-physical claims follow a priori from the physical truths. Insofar as we refer to *water* using an everyday concept that invokes the causal role of water, then we can arguably deduce all true claims about water from a complete inventory of physical truths, for these truths will tell us that H₂O is the physical stuff that plays the watery role. But if phenomenal concepts don't invoke any causal roles, then there will be no such analogous a priori deduction of phenomenal claims from physical truths. ('Inflationist materialism' is another name for David Chalmers' 'Type-B materialism' (Chalmers, 2002). I would say that this position is now the standard view among materialist philosophers of mind. (Cf. Horgan, 1984, Peacocke, 1989, Loar 1990, Papineau, 1993, Sturgeon, 1994, Hill, 1997, Hill and McLaughlin, 1998, Tye, 1999.))

In the last section of his comments Kirk defends the 'a priori characterization of materialism' against inflationist materialism. As a materialist, he then has to deny any a priori divide between our concepts of conscious states and our grasp of their functional roles. I shall say something about Kirk's views about our concepts of conscious states in a moment. But first I would like to point out that the 'a priori characterization of materialism' makes extremely strong demands in general, even outside the realm of the mind-brain relation.

Take claims made using proper names—*Tully had brown hair*, say. Let us suppose that the totality of physical truths does indeed tell us about the hair colour of all the humans who ever existed. Still, will it tell us which of those human beings is *Tully*? In order for this to be deducible a priori from the physical truths, our grasp of the name 'Tully' will have to carry with it enough a priori information for us to single out Tully from all other humans. I see no reason to suppose that our competence with the name 'Tully' ensures this. After all, it is a commonplace of modern philosophy of language that our ability to use 'Tully' referentially depends inter alia on our causal-historical links to Tully, rather than on any uniquely identifying a priori description. (And the same goes for many other referring terms, apart from proper names of people.)

I take this point to discredit the a priori characterization of materialism. It is clearly no argument against materialism that *Tully had brown hair* cannot be deduced a priori from physical truths. The barrier is not that *Tully had brown hair* commits us to something non-

physical. It is just that we can understand 'Tully' without knowing how to pick out Tully from all the purely physically specified people.

Similarly, I say, with phenomenal terms. We cannot deduce that *phenomenal pain is nociceptive-specific neuronal activity*, say, a priori from the physical truths. But this isn't because 'phenomenal pain' refers to something non-physical. It's just that we can have the concept of phenomenal pain without knowing how to pick out its referent from among all the purely physically specified states.

This is why I think zombie worlds are conceivable but not possible. Since there is no a priori route from the physical truths to phenomenal pain, we will not violate any conceptual constraints if we posit a being that shares all our physical properties but not our phenomenal pains. Yet, for all that, if phenomenal pains *are* material states, then such a being will not be possible, by the necessity of identity. [\[1\]](#)

Let me now turn to Kirk's suggestions about our concepts of conscious states. Since Kirk is a materialist who upholds the a priori characterization of materialism, he must maintain that these concepts will allow us to deduce the phenomenal facts a priori from the physical facts. Given this, he faces a prima facie difficulty with Frank Jackson's famous 'Mary' thought-experiment, since at first sight the post-exposure Mary would seem to acquire phenomenal concepts that can't be linked a priori with the physical facts. So somehow Kirk must resist taking the Mary story at face value. (Here Kirk is in the same boat as the contemporary Frank Jackson. Since formulating the 'knowledge argument', Jackson has become persuaded that materialism must be true (Jackson, forthcoming). However, his original knowledge argument successfully demonstrates that materialism is inconsistent with the two claims (a) that Mary shows that our phenomenal concepts are a priori detached from the physical facts, and (b) that materialism requires phenomenal claims to follow a priori from the physical truths. So Jackson has been forced to give up (a) or (b). Somewhat surprisingly, he has stuck to (b), and given up (a). That is, he has kept faith with the a priori characterization of materialism, but now denies that the Mary story shows that phenomenal concepts are a priori detached the physical facts. If you ask me, he has kept the bad bit of the knowledge argument, and thrown away the good bit. The Mary story is a terrific demonstration of distinct phenomenal concepts. By contrast, the a priori characterization of materialism strikes me as quite misguided.)

It is not entirely clear from Kirk's comments how he himself would deal with the Mary thought-experiment. He is happy to allow that there is a sense in which Mary acquires new phenomenal concepts of her new experiences when she comes out of her grey prison. But he denies that these new concepts are entirely distinct from the old functional concepts that she could previously have used to refer to those experiences. Even if phenomenal concepts 'are not be linked in any obvious or direct way with . . . functional concepts, they may still be linked in indirect, unobvious ways'. Again, Kirk allows that we may 'have two very different ways of *using* our concept of pain. . . But that does not seem enough to justify the claim that there are two concepts.'

However, I do not see how these claims can be reconciled with a variant of the Mary thought-experiment that I discuss in my book. This is the case where Mary is shown a coloured piece of paper when she emerges from her grey prison, but isn't told what colour it is. Now she has a new concept (*that* kind of experience'), but this concept surely has no links, however indirect and unobvious, with any functional concepts. Mary can think about the type of experience in question, but no amount of a priori reflection is going to enable her to figure out its characteristic causes or effects. By the same coin, she will surely have no way of inferring the satisfaction of her new phenomenal concept a priori from the totality of physical truths.

Let me respond to one final aspect of Kirk's comments. He feels that, if phenomenal-material identities were brute—not deducible a priori from the physical facts—then they would be mysterious. In the book, I claim that brute phenomenal-material identities are no more mysterious than brute proper names identities like *Cicero = Tully*. Kirk objects that 'the conditions for giving a man a name are easily understood and easily satisfied, while that is not so for the case of applying phenomenal concepts to physical properties'. Now, I concede phenomenal concepts may not be so 'easily understood' as proper names. But *Thinking about Consciousness* is a sustained attempt to remedy this, by elaborating a detailed account of how phenomenal concepts work, and in particular of how they can have physical referents, even though they are not associated a priori with any functional roles.

This account is no doubt flawed in particular respects, but I see no reason in principle to rule out some such explanation of how phenomenal concepts apply to physical properties.

Melnyk's Comments

In the book I do agree that there is something intuitively mysterious about mind-brain identities. But I deny that this feeling of mystery derives from the non-role nature of phenomenal concepts and our consequent inability to deduce the phenomenal facts a priori from the physical facts. Many contemporary philosophers refer to the absence of such a priori brain-mind deductions as 'the explanatory gap' (Levine, 1983). As an inflationist materialist, I of course accept that there is such an 'explanatory gap'—there are indeed no a priori brain-mind deductions. However I don't think that this 'explanatory gap' is why we find the relationship of mind to brain so puzzling. 'Explanatory gaps' of this kind are two-a-penny, arising with the many other referring terms, such as ordinary proper names, which don't pick out their referents via roles. The feeling of mystery we feel in the mind-brain case is something else again—it is a real 'intuitive gap' rather than the commonplace and unpuzzling 'explanatory gap'. To understand the source of this intuitive gap we need to look elsewhere.

Andrew Melnyk's comments focus on my analysis of this intuitive gap. He starts by noting that my central claim about a widespread '*intuition* of (mind-brain) distinctness' is allied with the stronger thesis that this intuition of distinctness isn't just a nagging doubt, but actually stops any of us 'really *believing*' materialism. Moreover, as he notes, I do little to defend, or indeed clarify, this stronger claim in the book. Let me try to do a bit better here.

To get a hold on the issue, let me explain why I want to claim that none of us—including dyed-in-the-wool materialists like myself—really believes materialism. I need this claim in order to explain why even paid-up materialists continue to react to mind-brain identities in ways which according to my analysis commit them to dualism. To be specific, I need the claim to explain why even paid-up materialists continue to feel that zombies are *prima facie possible*; and I need it to explain why paid-up materialists continue to share the widespread feeling that there is something *mysterious* about mind-brain identities. In the book I argue in detail that neither of these reactions is explicable by the commonplace 'explanatory gap'—that is, by reference to the a priori separation of phenomenal concepts from functional and other material concepts. For we have the same a priori separation in other cases, such as proper name identities, yet people who come to accept such identities don't continue to regard them as mysterious, or their falsity as apparently possible (rather, they come to think: Cicero *couldn't* fail to be Tully—after all, they're the *same* person). So I offer an alternative explanation for these persistent reactions in the mind-brain case: we have these persistent reactions because we don't really believe that phenomenal states are brain states to start with—and then, of course, we do find their relation mysterious ('Why ever should brain states be accompanied by phenomenal states?'), and do think that brain states without phenomenal states are possible (simply because we think they are distinct properties, even if correlated in the actual world).

This now puts the question of disbelieving mind-brain identities into better focus. I need to attribute such disbelief to people to just the extent that they manifest the dualist reactions to mind-brain identities. In the book I simply assumed that even professed materialists will continue to have these dualist reactions, and inferred from this that they must really disbelieve materialism, despite any avowals to the contrary. But, now Andrew Melnyk has raised the issue, I see that there is room for a more nuanced treatment.

Perhaps different professed materialists continue to have the dualist reactions to different degrees. While some might continue to feel them fully, others might only feel them to a lesser degree ('Zombies don't strike me as so obviously possible any more'), and yet others might lose the reactions almost entirely. Correspondingly, alongside those professed materialists who don't actually believe materialism, there may also be those who give materialism some non-trivial degree of belief, and also those who give materialism pretty much full credence (for whom dualism is indeed just a 'nagging doubt').

Again, there could be complexities in the mode in which dualism is believed, rather than the degree. It is not always straightforward whether someone believes some proposition. You can fully believe something at a theoretical level, yet disbelieve it at some more primitive level. Consider people who cross their fingers when the aeroplane is taking off, or people who are 'in denial' about something for which they have overwhelming evidence, or indeed people who undergo the Müller-Lyer illusion. In all these cases, there is a sense in

which they both believe and disbelieve something. Maybe this is how it is with many professed materialists. They believe materialism at a theoretical level, but at some more primitive level they remain in the grip of dualism. Their primitive disbelief will then offer an explanation of their continued dualist reactions. To the extent that their thinking is influenced by their primitive disbelief in materialism, zombies will continue to strike them as possible, and the mind-brain relation will continue to seem mysterious.

Let me now turn from the issue of how far all of us believe dualism to the question of why we do so. In the book I offer 'the antipathetic fallacy' as my explanation. Melnyk raises some doubts about this explanation, and offers some alternative suggestions of his own. But before considering his points, it will help to make a methodological observation. When I first aired my 'antipathetic' diagnosis to colleagues in London in the early 1990s, my friend Scott Sturgeon said 'That's an interesting sociological hypothesis'. I was somewhat taken aback at this apparently belittling reaction to some years of hard philosophical work, but I quickly realized Scott was quite right. Claims about the source of dualist thoughts are clearly empirical claims, answering to facts about the cognitive processes of the individuals covered by the claims. This means that we need not regard such claims as a yes-or-no matter. One explanation for dualist thoughts may apply to some individuals, another to different individuals. I shall not dwell on this point in what follows, but readers will do well to bear it in mind. (This fits with the point made a moment ago, that the whether everybody believes dualism isn't a yes-or-no matter either, even before we start asking why. Just as different people may believe dualism to different degrees, and in different modes, so also may they believe if for different reasons.)

Melnyk wonders whether my 'antipathetic fallacy' is the right explanation for our intuitive inclinations towards dualism (the 'intuition of distinctness' henceforth). On my hypothesis, the fact that material concepts do not *use* phenomenal properties confuses people into thinking that material concepts do not *mention* them either. Of course, most referring concepts don't use the items they mention, but my idea, as Melnyk explains, is that it is specifically the *comparison* with phenomenal concepts, which *do* use the phenomenal states they refer to, that confuses people here.

Melnyk observes that my story requires some pretty sophisticated mental capacities. I need to suppose that, when people refer to some phenomenal state with some phenomenal concept, they can simultaneously think about their deployment of that phenomenal concept, and note that it involves that same phenomenal state. I agree that this is a pretty sophisticated ability, but not necessarily one that is beyond ordinary people (as Melnyk himself allows). Think what happens when people are invited to reflect on whether 'This technicolour phenomenology be produced by soggy grey matter'. They *introspect* or *imagine* seeing colours on the left hand side, and then note that the phenomenology of these acts is absent when they *think of* soggy grey matter on the right hand side. Moreover, it is worth remembering that my story doesn't require ordinary people to keep a very clear grip on what is going on in such cases—on the contrary, I suppose that, once they have vaguely noted that material concepts 'leave out' the feelings associated with phenomenal concepts, they then proceed to get caught up in a fallacious use-mention muddle.

Melnyk has a more definite worry about the antipathetic fallacy. Suppose ordinary people do note that their deployment of phenomenal concepts involves being in the phenomenal state referred to. Why ever should they conclude on this basis, via some sort of one-shot induction, that *all* concepts that refer to phenomenal states must so involve being in those states? Well, I agree that it is implausible that ordinary people should make such an induction. But that is not my hypothesis. To grasp clearly that phenomenal concepts use the selfsame phenomenal states that they mention, and to infer from this that all concepts that refer to phenomenal states must do the same, would be a rash induction, but at least it would be cogent. The reasoning I attribute to ordinary people is not rash, but muddled. They somehow note that non-phenomenal concepts 'leave out' the phenomenal states that phenomenal concepts 'involve', and fallaciously infer from this that non-phenomenal concepts don't refer to phenomenal states. If they could see clearly that the 'involvement' of phenomenal states in phenomenal concepts is a matter of the concepts simultaneously both using and mentioning the states, as Melnyk's inductive reconstruction of the antipathetic fallacy has it, then they would already be articulating things in a way that would enable them to avoid the confusion I attribute to them.

Perhaps Melnyk would want to pursue this line of objection. Let us agree that the antipathetic fallacy involves a kind of use-mention confusion, rather than a rash induction.

Still, why should this confusion arise only with concepts that refer to phenomenal states, and not with other kinds of concepts? But here there is a ready answer. We can think phenomenally about the deployment of any concept. But only in the case of phenomenal concepts will this phenomenal introspection inevitably mean we are also thinking about something phenomenally similar to the referent of the concept. For only phenomenal concepts refer by simultaneously activating some phenomenal state that is like their referent. So phenomenal concepts are indeed peculiar, in introspectively appearing to 'involve' their referents in a way that makes other ways of referring to those referents seem pale by comparison.

I want now to take up Melnyk's alternative positive suggestion about the source of the intuition of distinctness. This is that phenomenal and material concepts may be so cognitively differently that it is impossible for us to 'merge files' in the way we generally do when embracing an identity claim. In the book I briefly consider this suggestion, only to dismiss it on the grounds that phenomenal concepts are closely related to perceptual concepts, yet no such cognitive barrier seems to block file-merging across the perceptual-theoretical divide. Melnyk raises two doubts about this line of argument. First, he suggests I may be wrong to hold that there is no cognitive barrier to file-merging across the perceptual-theoretical divide. Second, and independently, he suggests that differences between the phenomenal and perceptual cases might explain why perceptual-theoretical file-merging is possible even when phenomenal-material file-merging is not. I am more persuaded by the first suggestion than the second. Let me consider them in reverse order.

Melnyk's second suggestion is that there may be a barrier to file-merging in the phenomenal-material case that is absent in the perceptual-theoretical case. His suggestion relates specifically to phenomenal concepts that are only usable when you are actually having the states they refer to ('That is going on in me now') and which don't even involve the ability to re-identify those states as the same again. In such cases, Melnyk suggests, any temporary file associated with the phenomenal concept would simply be too transient to be merged with any permanent material concept file. I find this unpersuasive for three reasons. First, I find it doubtful that any genuine referring term should be so transient as to be unavailable for merging with others; what's the point of being able to acquire facts involving some entity if you can't slot them informatively into your overall picture of the world? Second, I doubt that any phenomenal concepts fit Melnyk's very simple model; to pick out some phenomenal states as 'that' requires at least that you be able to attend to it, and it seems empirically likely that you can reidentify any experiences you can attend to. Third, I don't see why the kind of construction Melnyk has in mind should yield an asymmetry between phenomenal and perceptual concepts; any demonstrative analysis of phenomenal concepts would seem to have a natural parallel for perceptual concepts (thus, along with 'that (experience)', we would have 'that (observable property)').

Melnyk's first suggestion does not try to drive a wedge between the phenomenal-material and perceptual-theoretical cases; rather, he goes along with my assumption that the two cases stand or fall together, but argues that perceptual-theoretical examples support the conclusion that file-merging is blocked in both cases. In the book I argued the other way, urging that file-merging is possible in both cases: thus I maintained that there is no barrier to merging a visual concept of a kestrel (such as might be derived from first-hand observation) with a theoretical concept of a kestrel (as might be derived from a textbook of evolutionary biology). Melnyk wonders whether the impression that such merging is possible might not derive from our tendency to slice off the secondary qualities from the visually-conceived kestrels, so to speak, thus making it easier to conflate them with the theoretically-conceived kestrels. But if this is what is going on, he points out, it provides no argument for the possibility of phenomenal-material mergers. For we have made the perceptual-theoretical merger possible only by moving the hard parts—the secondary qualities—into the mind; so mergers which do involve these hard parts may well still be cognitively unviable.

As I said, I find this line of argument relatively persuasive. There is a lot more to say about it, and in particular about the relationship between phenomenal and perceptual concepts. But rather than pursue this complex issue here, let me finish on an irenic note, by recalling the methodological point made earlier. Explanations of the intuition of distinctness need not be a yes-or-no matter. We do not need to choose between the antipathetic fallacy and the no-file-merging explanations. Perhaps one explanation works in some cases, and the other works in other cases. Or perhaps the two explanations sometimes complement each other: there may be people who wouldn't succumb to the antipathetic fallacy on its own, and who

wouldn't be stopped from merging files solely by the cognitive divergence of phenomenal and material concepts, but who capitulate to the two influences acting in concert.

References

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

Chalmers, D. 2002. 'Consciousness and its Place in Nature', in Stich, S. and Warfield, F. (eds) *The Blackwell Guide to Philosophy of Mind*. Oxford: Blackwell.

Hill, C. 1997. 'Imaginability, Conceivability, Possibility, and the Mind-Body Problem.' *Philosophical Studies*, 87.

Hill, C. and McLaughlin, B. 1998. 'There are Fewer Things in Reality that are Dreamt of in Chalmers' Philosophy.' *Philosophy and Phenomenological Research*, 59.

Horgan, T. 1984. 'Jackson on Physical Information and Qualia.' *Philosophical Quarterly*, 32.

Jackson, F. Forthcoming. 'Mind and Illusion.'

Levine, J. 1983. 'Materialism and Qualia: The Explanatory Gap.' *Pacific Philosophical Quarterly*, 64.

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

Papineau, D. 1993. *Philosophical Naturalism*. Oxford: Blackwell.

Peacocke, C. 1989. 'No Resting Place: A Critical Notice of *The View from Nowhere*.' *Philosophical Review*, 98.

Sturgeon, S. 1994. 'The Epistemic View of Subjectivity.' *Journal of Philosophy*, 91

Tye, M. 1999. 'Phenomenal Consciousness: The Explanatory Gap.' *Mind*, 108.

Notes

[\[Note 1\]](#) Kirk wonders how I would respond to David Chalmers' appeal to 'two-dimensional semantics' to cast doubt on such materialist a posteriori necessary identities (Chalmers, 1996). I don't discuss 'two-dimensional semantics' explicitly in the book, but I think a clear enough answer is implicit there. Chalmers supposes that all terms have a 'primary intension', in addition to their referents as normally conceived. This 'primary intension' consists of those entities that the term would pick out in other possible worlds 'considered as actual' (for example, 'water' would pick out XYZ if the actual world's watery stuff were XYZ rather than H₂O). Chalmers then assumes that, if the claim that $a \neq b$ is so much as conceivable (for example, $water \neq H_2O$), this must be because 'a 's and 'b 's 'primary intensions diverge' (there must be worlds in which the terms 'water' and 'H₂O' would pick out different items), from which it follows that there is a genuinely possible world corresponding to the thought $a \neq b$. Applying this to the mind-brain case, we then get the Kripkean thesis that, if it is so much as conceivable that $pain \neq M$, where 'M' is some material concept, then there must be genuine possibilities where 'pain' and 'M' pick out different items. Moreover, if 'pain' is a priori distinct from *all* material concepts, as the inflationist materialist assumes, then this must mean that 'pain' must refer by invoking some distinctively non-material entity. As an inflationist materialist, I respond to all this simply by denying Chalmers' crucial premise. I don't accept that, whenever some $a \neq b$ is conceivable, then 'a ' and 'b ' must have 'primary intensions' which diverge. The terms 'a ' and 'b ' may simply refer directly, which means they won't have any 'primary intensions' different from their normal referents (different from their 'secondary intensions'). We can still conceive $a \neq b$ without conceptual contradiction, simply because 'a ' and 'b ' are different terms which are not interchangeable in our cognitive economy. But it does not follow from this that 'a ' and 'b ' must have different 'primary intensions', that they must pick out their referents in ways that would give them different referents in other possible worlds 'considered as actual'.