SWIF Philosophy of Mind, 09 May 2003
http://www.swif.uniba.it/lei/mind/forums/papineau.htm

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 H_2O 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 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 Phenomenonological 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 H_2 O). 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 'H_2 O' 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 ≠ 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'.