Synthese, 163, 2008, 133-43. (An earlier version of this paper was published in *Hommage á Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz*, ed. T. Rønnow-Rasmussen, B. Pettersson, J. Josefsson, and D. Egonsson, 2007.)

Is the mystery an illusion? Papineau on the problem of consciousness

Pär Sundström, Umeå University

ABSTRACT: A number of philosophers have recently argued that (i) consciousness properties are identical with some set of physical or functional properties and that (ii) we can explain away the frequently felt puzzlement about this claim as a delusion or confusion generated by our different ways of apprehending or thinking about consciousness. This paper examines David Papineau's influential version of this view. According to Papineau, the difference between our "phenomenal" and "material" concepts of consciousness produces an instinctive but erroneous intuition that these concepts can't co-refer. I claim that this account fails. To begin with, it is arguable that we are mystified about physicalism even when the account predicts that we shouldn't be. Further, and worse, the account predicts that an "intuition of distinctness" will arise in cases where it clearly does not. In conclusion, I make some remarks on the prospects for, constraints on, and (physicalist) alternatives to, a successful defence of the claim (ii).

1. Introduction

There is *something it is like* for me to, e.g., feel pain in my finger. We may think of what this is like as a *property* that is instantiated either by *me* at a time, or by a *state* that I'm in, or by an *event* in which I'm involved. This property can also be instantiated (if it's a property of me at a time) by some other subject or by me at

another time, or (if it's a property of a state of mine) by a state of some other subject or by another state of mine, or (if it's a property of an event in which I'm involved) by an event involving some other subject or another event in which I'm involved.

It's commonly held that such "consciousness properties" present us with a dilemma. On the one hand, we have reason to believe that they are identical with physical or functional properties (of subjects, states of subjects, or events involving subjects). But on the other hand, it seems mysterious how this could be so. Consciousness properties seem to be somehow "subjective" or "qualitative" or "feely", and it is hard to understand how – or believe that – such a property could be one and the same as some physical or functional property.

A number of recent philosophers have argued that we should embrace the first horn of this dilemma and, in a certain sense, "explain away" the second. On this view, consciousness properties *are* identical with physical or functional properties; we have *reason to believe* they are; and we also have the intellectual tools to *understand* how this could be so; but we remain puzzled about this identity because there are features of our ways of apprehending or thinking about consciousness that we get deluded or confused by (or about). The way to achieve satisfaction with physicalism is to understand how these delusions or confusions arise.

This paper examines David Papineau's influential version of this view. According to Papineau, we have convincing evidence that consciousness properties are physical or functional properties, but despite this evidence, we can't really make ourselves believe this identity: "something stops us *really* believing the materialist

identification of mind with brain, even those of us who officially profess materialism" (2002, 94). This compulsive "intuition of distinctness" is generated by the different character of two kinds of concept – "phenomenal concepts" and "material concepts" – that we use or can use to think about consciousness. Since *thinking* about consciousness properties under phenomenal concepts is so different from *thinking* about these same properties under material concepts, we have a hard time believing that these concepts can co-refer.

I shall argue that, while this account has some attractive features, it is unsuccessful. It is at least arguable that we are mystified about physicalism even when the account predicts that we shouldn't be. And, worse, the account predicts that an intuition of distinctness will arise in cases where it patently does not.

Section 2 contains a few preliminaries. It spells out Papineau's reasons for embracing physicalism, makes some observations about the puzzlement generated by this view, and distinguishes two general physicalist diagnoses of this puzzlement. In section 3, I present and criticise Papineau's account of the intuition of mind-brain distinctness. Section 4 considers, somewhat briefly, what general lessons can, and can't, be drawn about the mystery of consciousness from the preceding discussion.

A terminological note: 'Physicalism' will be used throughout for the thesis that consciousness properties are *identical* with some set of *physical or functional* properties (disjunctive properties included). I shall not attempt to define 'physical property'. But I will take physicalism to be, at least, a *non-trivial, reductionist* thesis.

(So, for example, you don't count as a physicalist merely by using 'physical property' in so broad a sense that consciousness properties are *trivially* physical.)

2. Preliminaries

Like many other philosophers, Papineau is convinced that physicalism is true by a version of what is often called "the causal argument". The argument goes roughly as follows.

To begin with, it is natural to suppose that instantiations of consciousness properties often have physical effects; for example, that what it's like for me to feel pain on a given occasion is part of what causes me, on that occasion, to put a band-aid on my finger. However, we also have reason to believe that all physical effects are caused by instantiations of material properties. So, we seem to want to say that both (i) that *some* physical effects are caused by instantiations of consciousness properties and (ii) that *all* physical effects are caused by instantiations of material properties. One possible account of how (i) and (ii) could both be true is that physical effects of consciousness properties have two *distinct* causes – like the death of the person who is shot by two bullets that enter his heart at the same time. However, it seems unappealing to suppose that conscious causation should generally be a matter of such overdetermination. The physicalist identity thesis offers a more appealing account: if consciousness properties are identical with material properties, then it's very easy to see how a given physical effect could have both a conscious and a material cause.[1]

[1] For Papineau's version of this argument, see (2002, chap. 1 and appendix). For doubts about the argument, see, e.g., Sturgeon (1998). One should note that the present argument is distinct from the more traditional consideration that a non-material mind would be *too different* from matter to have any

Papineau takes this argument to offer "definitive" support for physicalism (2002, 15). At the same time, he recognises, and shares, the common sentiment that "it certainly doesn't *seem* as if conscious properties are identical with brain properties … there is something very counter-intuitive about the phenomenal-material identity claims advocated by materialists" (ibid., 74). Indeed, he grants that "it seems *absurd* to identify conscious states with material states" (ibid., 1, emphasis added).

It's important to note that we normally don't feel this way about identity claims – at least not *after* we have sufficient evidence that they are true. Consider, for example, the claim that water = H_2O. *Until* we have sufficient reason to believe this claim, we may well find it counter-intuitive or even absurd: how could the transparent, cohesive liquid in my glass be an ensemble of tiny molecules? But once we have reason to believe that there are tiny molecules in the glass, and that their properties can explain the observable features of the liquid, it no longer seems absurd to identify the two.[2] The same is true in the case of identifications of individuals. One may initially be puzzled about the suggestion that Superman, the flying hero, should be identical with Clark Kent, the mild-mannered reporter. But once one is provided with enough evidence that this is so, no sense of puzzlement remains. However, in the case of consciousness and the brain, the situation seems different. Many philosophers are by now convinced – often by some version of the causal argument – that consciousness

---

causal relations with it. That idea plays no role in the present argument. The idea of the present argument is, rather, that the physical world is "causally complete" in the sense that physical effects already have complete physical causes, and that there is therefore no room for anything non-physical to make a causal contribution to the physical world.

[2] There may of course be room for philosophical debate about whether the relation here is really one of *identity*, or *constitution*, or something else (see, e.g., Johnston 1997). The point here is only that it doesn't seem *intuitively absurd* to suppose that the relation is identity.

properties must be identical with some physical or functional properties. But many, even among those who accept physicalism, *continue* to find this claim puzzling.[3]

A physicalist ought to say something about this residual puzzlement. And in general terms, it seems clear what the response should be: a physicalist should put the blame for our puzzlement about mind-brain identity squarely on *us*. By the physicalist hypothesis, there is, in reality, just one phenomenon – a set of consciousness/material properties. And there can hardly be anything "objectively mysterious" about that. If *we* are puzzled about it, this must be due to how *we* apprehend or think about this one phenomenon.

One can distinguish two different ways of developing this general suggestion. One proposal is that we are puzzled about the supposed consciousness-brain identity because our current concepts or theories don't adequately capture consciousness and whatever other phenomena we may need to grasp to fully understand consciousness. Thomas Nagel holds something like this view. According to Nagel, the present status of physicalism is "similar to that which the hypothesis that matter is energy would have had if uttered by a pre-Socratic philosopher. We do not have the beginnings of a conception of how it might be true" (1974, 447). This kind of diagnosis is also made by Colin McGinn (1989), and I believe it is also the view of Levine (2001). Call this the *Lack of Understanding Hypothesis*.[4]

---

[3] It may have been observed that the causal argument only (or at best) gives us reason to think that every consciousness property is identical with *some* material property; it doesn't suggest *which* material property a given consciousness property is identical with. But this can hardly be the source of our puzzlement about physicalism, because in other cases we are not puzzled about such claims. For example, any puzzlement I may initially feel about the suggestion that Superman is identical with *some* person in a football arena tends to evaporate once I'm provided with sufficient evidence for this.

[4] I take the Lack of Understanding Hypothesis to be compatible with different assessments about *how far* we are from attaining a proper understanding of consciousness. On one extreme of the spectrum,

An alternative proposal is that we already have whatever concepts and theories we need to understand physicalism, but we are (sometimes, often or always) puzzled about this view anyway because there are features of our ways of apprehending or thinking about consciousness that we get deluded or confused by (or about). It is as though we have all the tools we need for the task at hand, but we (sometimes, often or always) fail to use them in a clear-headed way. I will call this the *Illusion Hypothesis*.[5]

As we shall see in more detail shortly, Papineau's view is a version of the Illusion Hypothesis. He assures us that, "We don't need any fancy new concepts to understand consciousness" (2002, 1). As it happens, Papineau believes there is room for scientific progress regarding consciousness. But our dissatisfaction with materialism does not stem from the underdeveloped state of the relevant sciences and won't go away with progress in them. Rather, we find materialism puzzling because we "get confused by superficial features of our thinking". The recipe for getting clear about consciousness and its physical nature is to "unravel" our confusions (ibid., 4).

McGinn believes that we are "cut off by our very cognitive constitution" from understanding the true theory about the mind and the brain (1989, 350). (Indeed, McGinn is open to the possibility that *no* mind can understand its relation to its own brain; see ibid., 360-1). On a more optimistic note, Nagel thinks we can understand this, though only through a conceptual development of a "radical and scientifically unprecedented kind" (2000, 446). However, it's also possible to combine the Lack of Understanding Hypothesis with the view that normal scientific development will one day provide us with the correct account.

[5] A couple of observations about the two hypotheses:

First: There would be a claim for characterising both hypotheses as "illusion hypotheses": according to both hypotheses, we are deluded if we take our puzzlement about physicalism to reflect an *ontological* gap between consciousness and the physical world. But what I here call the Illusion Hypothesis has a stronger claim to this title, because on this hypothesis we are also deluded if we take our puzzlement to reflect that our present concepts or theories are inadequate to the phenomena.

Second: A defender of the Illusion Hypothesis need not hold that we know *all* there is to know about consciousness and its physical nature. Presumably, there are details left to learn on anybody's account. But the hypothesis entails that our puzzlement about physicalism isn't *due* to any failure in our concepts and theories to capture the phenomena, and so won't disappear with further improvements

Other versions of the Illusion Hypothesis are developed by Lycan (1987, 1996), Loar (1990/1997), Tye (1999), Perry (2001), Melnyk (2002), and Balog (manuscript).[6] I will make a few remarks about some of these other versions below, but the focus will be on Papineau's view.

3. Papineau's account of the "intuition of mind-brain distinctness"

The fundamental delusion that we suffer from in the case of consciousness is, according to Papineau, an "intuition of mind-brain distinctness", which flourishes even among those who accept the arguments for physicalism.[7] It's not clear whether Papineau thinks we can fully rid ourselves of this intuition. But we can understand its source and realise, at least at some level, that it is without merit (2002, 1).

The intuition of mind-brain distinctness arises, Papineau claims, because of a peculiar "use-mention feature" of our phenomenal concepts of consciousness properties. The details of this story have evolved over time. In the most recent version, phenomenal concepts are explained in terms of certain exercises of "sensory templates" that are also involved in uses of "perceptual concepts".[8] In brief outline, the account goes as follows.

---

in this regard. If we can make the puzzlement go away at all, it is by finding a way of thinking with (or about) our present concepts or theories in an unconfused way.

[6] A similar hypothesis, though agnostic about physicalism, is proposed by Sturgeon (1994).

[7] This is the *fundamental* illusion. According to Papineau, we may suffer some further illusions as a *consequence* of this one; see footnote 12 below.

[8] The most recent version is in Papineau (2007). References will be to this text unless otherwise noted. For earlier versions of the story, see Papineau (1993, chap. 4), and Papineau (2002, chap. 4 and 6).

Sensory templates, which we might think about as "patterns" or "moulds" in the brain, are originally acquired in perceptual encounters with the environment. Once acquired, a sensory template can be reactivated both in *perception*, "via matches between incoming stimuli and stored template", and *imaginatively*, in the absence of external stimuli (114-5). A given sensory template can be used to think about a variety of things. For example, you can use one and the same template to think about an *individual* bird, or about a *type* of bird, or about a type of *experience of* a bird.[9] When a template is used to think about features of the external environment, it functions as a "perceptual concept". When it is used to think about an experience, it functions as a "phenomenal concept" (122).[10]

To activate a sensory template either perceptually or imaginatively is to have a *conscious experience* (117-8). Now, this means that when one thinks about an *experience* via a perceptual or imaginative activation of a sensory template, one *uses* an experience to think *about* an experience:

> any exercise of a phenomenal concept to think about a perceptual experience will inevitably either involve that experience itself or an imaginary

[9] One may wonder what is *different* when I use a template to think about (say) an individual bird and a type of experience. Papineau suggests that different kinds of information are activated in these cases. If I use a template to think about an individual bird, I may associate with it the information that what I experience has a missing claw (116), whereas if I use it to think about an experience, I associate with it "experience-appropriate information" like: "what I'm encountering ceases when I close my eyes" (122).

Papineau denies, tentatively, that sensory templates can be used to think about *individual experiences*. His reason is that (i) the function of a sensory template is to accumulate information about a referent on encounters with it, and make that information available on reencounters, and (ii) individual experiences "do not seem to persist over time in the way required for reencounters to be possible". When a token experience departs from consciousness, it never returns – or so Papineau is inclined to think (123). This feature of the account will not matter for present purposes.

[10] As is often noted, 'concept' is used in different ways in philosophy and cognitive science. For example, the term sometimes denotes *mental particulars* in the mind or brain, and sometimes abstract constituents of Fregean Thoughts. I'm not sure Papineau's usage is entirely uniform, but for the most part he seems to use 'concept' for mental particulars. I shall take this to be the default sense of the term here.

> recreation of that experience. If we count imaginary recreations as 'versions' of the experience being imagined, then we can say that phenomenal thinking about a given experience will always *use* a version of that experience in order to *mention* that experience (124).

Herein lays the source of our persistent intuition of mind-brain distinctness, according to Papineau. Suppose I consider the physicalist hypothesis that *pain = nociceptive-specific neuronal activity*, using a phenomenal concept on the left-hand side and a material concept on the right. Since the phenomenal concept activates a version of the mentioned experience, thinking the left-hand side of the identity "will *feel* like having the experience itself" (135). But this phenomenology will be absent when I think the right-hand side. So, there is a sense in which the exercise of the material concept "leaves out" the experience at issue: it doesn't activate any "version" of it. Now, it is, Papineau claims, "all too easy to slide from this to the conclusion that, in exercising such a material concept, we are not thinking *about* the experiences themselves" (ibid.). The inference is, of course, fallacious. Material concepts may well *refer* to experiences despite not *using* them, and the causal argument establishes that some of them do. Still, it is this fallacy, dubbed "the antipathetic fallacy", that accounts for our residual dissatisfaction with materialism.[11]

I think this account has some quite attractive features. It provides an intuitively appealing explanation of how we might come to believe that a description of the world couched in entirely "material terms" fails to mention consciousness properties

[11] The *pathetic fallacy* is the attribution of desires, emotions, or other mental features to inanimate objects. (The term is apparently due to Ruskin 1856.) The *antipathetic fallacy* is, Papineau explains, the converse: "where we refuse to recognize that conscious feelings inhere in certain parts of nature, namely, the brains of conscious beings" (1993, 116).

even though there is no rational ground for that belief. We should also note that the basic idea of the account is quite simple, and that many details in Papineau's story seem to be inessential to it. The basic idea seems to be that the intuition of mind-brain distinctness arises because we have a peculiar way of thinking about experiences that has *some* kind of tight connection with having or imagining having the experiences thought about. The suggestion that we think with "activated sensory templates" that can also be used to think about features of our environment should, I believe, be regarded as *one* speculation about what such "imagistic" thinking consists in. But the basic idea of the account does not stand or fall with this particular speculation.

As one might for these reasons have expected, the basic idea of Papineau's account has appealed to many philosophers, and my guess is that its apparent promise has contributed significantly to promoting the hypothesis that we can explain away our puzzlement about physicalism as a superficial cognitive illusion. For example, something like Papineau's account seems to be part of what both Michael Tye and Brian Loar have in mind when they promote this hypothesis. Thus, Tye echoes Papineau when he says:

> Consider … the following example of a phenomenal-physical identity claim: The visual experience of red = brain state *B*. One reaction some philosophers have to claims of this sort is that they must be mistaken, since the phenomenology isn't captured by the right-hand side. From the present perspective, this reaction involves a sense/reference confusion. When we think of the referent of the designator on the left-hand side in a phenomenal

> way, we bring it under a concept that has a distinctive functional role. In reflecting on the identity claim and what is puzzling about it, the phenomenal concept we deploy is apt to trigger in us a visual image of red. In this event, if the identity is true, our brain actually goes into brain state *B*. But, of course, when we think of the referent of the designator on the right hand side *as* brain state *B*, nothing like that happens. Exercising the neurophysiological concept is not apt to trigger a visual image of red. It may then be tempting to infer that the right-hand side has left out the phenomenology of the left, that there is a huge gap that the physicalist has failed to close. This conclusion clearly does not follow, however. There is indeed a striking difference in the roles that the concepts play, in their functioning, but not (so far as is shown here) in their referents (Tye 1999, 712-3).

Similarly, Loar:

> A phenomenal concept exercised in the absence of the phenomenal quality it stands for often involves not merely a recognitional disposition but also an image. And so, as a psychological state in its own right, a phenomenal concept – given its intimate connection with imaging – bears a phenomenological affinity to a phenomenal state that neither state bears to the entertaining of a physical-theoretical concept. When we then bring phenomenal and physical-theoretical concepts together in our philosophical ruminations, those cognitive states are phenomenologically so different that the illusion may be created that their referents must be different. It is as

> though anti-physicalist intuitions rest on a resemblance theory of mental representation, as though we conclude from the lack of resemblance in our phenomenal and physical-functional conceptions a lack of sameness in the properties to which they refer (Loar 1997, 605).[12]

But despite its attractive features, I think Papineau's account is incorrect. It fails, I believe, to give the right predictions about cases where we think about some experience *not* as the brain state that it (by the physicalist hypothesis) is, but also

---

[12] These passages both suggest that the "imagistic" character of phenomenal concepts gives rise to an intuition of mind-brain distinctness. So this seems to be *part* of what Tye and Loar have in mind when they argue that the puzzle of physicalism is a cognitive illusion. But it's not all of it. Both Tye and Loar also argue that we have doubts about physicalism because we expect the physicalist identity thesis to be *explanatory* in a way that we can – on appropriate reflection – appreciate that it cannot possibly be and shouldn't be expected to be. The relevant contrast here is with "theoretical identifications" like that of water with H_2O. As noted in passing above, such identities seem to be in a sense explanatory: the properties of H_2O molecules can *explain* the superficial features that we pre-theoretically associate with water, like reflectance properties, freezing patterns, etc. In contrast, identifications of consciousness properties with physical or functional properties seem not to be similarly explanatory. But according to Tye and Loar, this "explanatory gap" is predicted and fully accounted for by a correct view of our pre-theoretical, phenomenal concepts and how they differ from our pre-theoretical concept of, say, water. Briefly stated, the idea is that while our pre-theoretical concept of water is associated with descriptive predicates – such as *freezes at 0° C* – the satisfaction of which can be inferred from truths about the behaviour of H_2O molecules, our phenomenal concepts of consciousness properties are not associated with any such descriptions. This difference in our concepts predicts that, and explains why, physicalism about consciousness is not explanatory in the way that 'water = H_2O' is.

As far as I'm aware, neither Tye nor Loar says anything about the relation between the two (alleged) facts that (i) the "imagistic" character of phenomenal concepts gives rise to an erroneous intuition of mind-brain distinctness, and that (ii) we mistakenly expect physicalism to be explanatory in a way that other theoretical identifications are. Are there two distinct illusions here, or is there some relation between them, and if so which?

Incidentally, Papineau denies that we find physicalism mysterious because it is unexplanatory. He grants *that* physicalist hypotheses are not explanatory in the way that 'water = H_2O' is, but denies that this is *why* we find them problematic. His reason is that, "As far as explanatoriness goes, mind-brain identities are no worse off than many other respectable identities" (2002, 160). For example, he argues, a claim like 'Tully = Cicero' is just as unexplanatory as any physicalist identity hypothesis; 'Tully' and 'Cicero' are no more associated with explanation-facilitating descriptions than are our phenomenal concepts. This shows that we are not in general expecting or craving that an *explanation* should accompany a true identity hypotheses. Insofar as we have such sentiments in the case of physicalism about consciousness, this is because, Papineau suggest, we have a hard time believing this view in the first place. The real source of our concern is not a missing explanation, but the felt implausibility of it all (ibid., 145-6).

Another development of the Illusion Hypothesis that resembles Papineau's account is Lycan's discussion of the "stereoptic fallacy" – a fallacy that is, according to Lycan, one source of our resistance to physicalism (1987, 76-7; 1996, 47-8).

without having or imagining having that experience. To spell this out, it will be helpful to draw on Papineau's own discussion of such a case.

Once one has acquired a sensory template with which to think about a given type of experience, one can, Papineau suggests, create a "non-sensory file to house the information that has become attached to that template" (127). As Papineau conceives of it, such a file still has a close connection with experience. It is derived from a sensory template, and hence represents a way of thinking about an experience that is available only to those who have undergone that experience.[13] But a non-sensory file is not "imagistic": it allows its possessor to think about an experience at a time without having or imagining having that experience at that time.

Papineau raises the question whether such "phenomenally derived concepts" are distinct from their originating phenomenal concepts, or whether there is just one concept here that can be exercised in two different ways.[14] It is not obvious, he says, that one should take them to be distinct:

> From some perspectives, [taking them to be distinct] might seem like double counting. In particular, it is not clear that the standard Fregean criterion of cognitive significance will tell us that there are two concepts here (120).

---

[13] So, it differs from a concept like *Aunt Brenda's favourite experience*, which represents a way of thinking about an experience that *is* available even to somebody who has not undergone the relevant experience.

[14] This formulation is a bit sloppy. The issue here is whether a *token non-sensory file* and a *token sensory template* (which are both token concepts) might be concepts of the same *type*. But I shall follow Papineau in phrasing this issue as whether phenomenally derived concepts are distinct from their originating phenomenal concepts; I trust the reader to make the required type-token disambiguation.

This hesitation seems to me well-motivated. Suppose, for example, that on two occasions, I attribute in thought some predicate, *F*, to some experience-type, *E*, and that in one case I "activate a version" of *E* when I think of *E* while in the other case I don't. It may well be *transparently* and *immediately* clear to me that in these two cases, I think about the same experience, *E*, and attribute to it the same predicate, *F*. And if so, it's not clear that any Fregean test for cognitive significance will distinguish the concepts that are employed in these cases. Papineau's considered verdict on this issue is brief, and, I think, obscure. He says:

> There is no substantial issue here. To the extent that the flow of information between the two ways of thinking is smooth, the Frege test gives us reason to say that there is only one concept. On the other hand, to the extent that there are cognitive operations that distinguish a [phenomenally] derived concept from its originating [phenomenal] concept, there is a rationale for speaking of two concepts (ibid.).[15]

However, he doesn't spell out what the "rationale" is for speaking of two concepts.[16]

Now, we can afford to leave it open whether phenomenally derived concepts *should* always be taken to be distinct from their originating phenomenal concepts. Whatever

---

[15] The quotes in this paragraph are from Papineau's discussion of *perceptual* and *perceptually derived* concepts, a distinction that parallels the one between *phenomenal* and *phenomenally derived* concepts. I think it's clear that his remarks apply to both distinctions, if to either, and so I have taken the liberty to substitute 'phenomenal' for 'perceptual' to fit the present discussion.

[16] It's not at all obvious why – or that – one should want to distinguish concepts *whenever* there is some difference in "cognitive operations". To illustrate: I can think *that 15 is the square root of 225* either passingly or in a focussed, concentrated way. There is a difference in these "cognitive operations": one is quick and casual, the other slow and focussed. But it's not clear what rationale there might be for saying that I therefore exercise different concepts of *15* or *identity* or *square root* or *225* in these two cases, as opposed to saying that the same concepts are exercised in different ways. If there is such a

one thinks about that, I think phenomenally derived concepts can be used to show that Papineau's account is incorrect.

Let me begin by pointing out a *doubtful* implication of the account. According to the account, the intuition of distinctness ought *not* to arise if I think about some experience under a phenomenally derived concept and consider the hypothesis that the experience (thus conceived) is identical with some material property (materially conceived). Papineau acknowledges this implication and urges that it's defensible:

> Note how my explanation implies that the intuition of distinctness will only arise when we are thinking with phenomenal concepts, which use the very states they mention, and not when we are thinking with phenomenally derived concepts. This seems to me quite in accord with the facts (136, note 21).[17]

I'm less certain than Papineau is that this really does accord with the facts. Do we really feel mystified about physicalism *only* when we think about experience imagistically? That seems, to me, at least doubtful.

However that may be, what is clearly unacceptable are some predictions the theory entails about when intuitions of distinctness *will* arise. Suppose I can think about pain both with a "sensory template" and with a "non-sensory file". Then I can consider an identity hypothesis that we may render as follows:

---

rationale in the case of imagistic and non-imagistic cognitive operations, that would have to be spelled out.

$$pain_{st} = pain_{nsf}$$

where the subscripts indicate the different "mental terms" I employ to think about the pain. As noted, there is an issue about whether we should count these as uses of different concepts at all.[18] By Papineau's own admission, the "flow of information" between these ways of thinking might be "smooth", and to that extent we have reason to say that the same concept is employed twice. It is plain that, in such a case, I have no intuition that the identity hypothesis must be false. On the contrary, it may be quite *evident* to me that it's *true*. However, Papineau's account predicts that an intuition of distinctness *should* arise in this case. After all, when I think one side of this identity, it "will feel like having the experience itself", but this phenomenology will be "left out" when I think the other side. And that's what gives rise to the intuition of distinctness, according to Papineau.

Papineau's account entails, I think, false predictions about further cases where one thinks about an experience not as the brain state that it (by the physicalist hypothesis) is but also without activating any version of it. Even setting phenomenally derived concepts aside, it is clear that Papineau's distinction between "phenomenal" and "material" concepts of experience is not exhaustive. For example, I can, without activating any experience, employ the concept of *my brother's most salient current experience*. I don't know which experience this is; I don't even know for sure that there is such an experience (perhaps my brother is in dreamless sleep). But I can

[17] In fact, Papineau's account has this implication only if it claims that the imagistic character of phenomenal concepts is the *only* source of the intuition of distinctness. Since the implication is acknowledged by Papineau, this is apparently what he intends.

entertain the hypothesis that my brother's most salient current experience is identical with the experience of off-white that I have right now, as I attend to the background of this word document. Papineau's account seems again to predict that an intuition of distinctness should arise in this case. The experience of off-white is "left out" when I think non-imagistically about my brother's most salient current experience. But it is again clear that no intuition of distinctness arises in this case. Of course, it is not *evident* to me that my brother's most salient current experience is the same as the experience of off-white that I have right now. In fact, I doubt it. Indeed, if I take the problem of other minds seriously, I may forever doubt it. But these doubts are different from the intuition of distinctness. They are a *lack of conviction* that two concepts co-refer. The intuition of distinctness, on the other hand, is a kind of instinctive conviction that two different concepts *don't co-refer*. There is no such conviction in this case.

Quite generally, Papineau's account seems to predict that whenever I conceive of an experience-type *E* imagistically and consider a hypothesis that *E*, thus conceived, is identical with something, *X*, that I conceive of non-imagistically, I should have an intuition that *E* must be distinct from *X*. But I don't have such an intuition in each such case.

4. Concluding remarks

I have argued that Papineau's account is incorrect. What can we infer about the mystery of consciousness from this? I want to make three concluding remarks.

---

[18] That is, there is an issue about whether we should count these *token mental terms* as instances of

First: As I tried to illustrate above, I think Papineau's account has contributed significantly to selling the hypothesis that our intuitive puzzlement about physicalism can be explained away as a superficial cognitive illusion. To this extent, I believe optimism about this hypothesis is unfounded.

Second: There may of course be other, promising ways of developing the Illusion Hypothesis.[19] However, we have uncovered at least some important constraints on such a hypothesis. A successful Illusion Hypothesis must be compatible with the wide variety of cases in which we are *not* mystified by identity claims. We are not mystified by standard "theoretical identifications", like that of water with H_2O, or identifications of individuals, like that of Clark Kent and Superman. At least, we are not mystified by them once evidence for their truth has been absorbed. Moreover, we are not mystified in many cases where terms or concepts *of consciousness properties* are claimed to co-refer. For example, we are quite prepared to accept that a consciousness property conceived of "imagistically" may be identical with a consciousness property conceived of under some descriptive term like 'my brother's most salient current experience', and again, that a consciousness property conceived of imagistically may be identical with some consciousness property conceived of non-imagistically but (somehow) directly, as they are under Papineau's "phenomenally derived concepts".

---

different *types*. See footnote 14 above.

[19] For one alternative development of the hypothesis (explored by Tye and Loar), and some criticism of it (from Papineau), see footnote 12 above.

Third: Even if a promising version of the Illusion Hypothesis should be hard to come by, this does not by itself suggest that physicalism is false. Physicalism says that there is a single set of consciousness/material properties where we might have thought there were two distinct sets of properties. If we are mystified about this, this must be due to how we apprehend or think about this one phenomenon. As we have seen, the Illusion Hypothesis is *one* way of developing this general line of thought. An alternative view is that our current concepts or theories don't adequately capture whatever phenomena we need to grasp in order to understand consciousness.

Acknowledgments

For helpful comments and discussion, I thank Peter Nilsson, Iris Oved, David Papineau, Bertil Strömberg, and Jesper Östman.

References

Balog, Katalin (manuscript). Acquaintance and the mind-body problem. Online at http://pantheon.yale.edu/%7Ekb237/Web%20publications/Acquaintance.pdf.

Johnston, Mark (1997). Manifest kinds. *The Journal of Philosophy*, 94, 564-83.

Levine, Joseph (2001). *Purple Haze: The Puzzle of Consciousness*. Oxford: Oxford UP.

Loar, Brian (1990/1997). Phenomenal states. Originally published in J. Tomberlin (ed.), *Philosophical Perspectives, 4: Action Theory and Philosophy of Mind*. Atascadero, CA: Ridgeview, 1990. Reprinted in revised form in N. Block et al. (eds.), *The Nature of Consciousness: Philosophical Debates*, Cambridge, Mass.: MIT Press, 1997. Page references are to the 1997 version.

Lycan, William (1987). *Consciousness*. Cambridge, Mass.: MIT Press.

——— (1996). *Consciousness and Experience*. Cambridge, Mass: MIT Press.

McGinn, Colin (1989). Can we solve the mind-body problem? *Mind*, 98, 349-66.

Melnyk, Andrew (2002). Papineau on the intuition of distinctness. In *SWIF Philosophy of Mind Review*, 4. Online at http://www.swif.uniba.it/lei/mind/forums/melnyk.htm.

Nagel, Thomas (1974). What is it like to be a bat? *Philosophical Review*, 83, 435-50.

——— (2000). The psychophysical nexus. In P. Boghossian and C. Peacocke (eds.), *New Essays on the A Priori*. Oxford: Oxford UP.

Papineau, David (2007). Phenomenal and perceptual concepts. In T. Alter and S. Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford: Oxford UP.

——— (2002). *Thinking about Consciousness*. Oxford: Oxford UP.

——— (1993). *Philosophical Naturalism*. Oxford: Blackwell.

Perry, John (2001). *Knowledge, Possibility, and Consciousness*. Cambridge, Mass.: MIT Press.

Ruskin, John (1856/2005). Of the pathetic fallacy. In *Modern Painters, vol. 3: Of Many Things*. Chestnut Hill, MA: Adamant Media Corporation. Also online at http://www.ourcivilisation.com/smartboard/shop/ruskinj/.

Sturgeon, Scott (1994). The epistemic view of subjectivity. *The Journal of Philosophy*, 91, 221-35.

——— (1998). Physicalism and overdetermination. *Mind*, 107, 411-32.

Tye, Michael (1999). Phenomenal consciousness: the explanatory gap as a cognitive illusion. *Mind*, 108, 705-25. Reprinted as chapter 2 of Tye, *Consciousness, Color, and Content*. Cambridge, Mass.: MIT Press, 2000.