Does the exclusion argument put any pressure on dualism?

Christian List and Daniel Stoljar

To appear in the *Australasian Journal of Philosophy*

Abstract: The exclusion argument is widely thought to put considerable pressure on dualism if not to refute it outright. We argue to the contrary that, whether or not their position is ultimately true, dualists have a plausible response. The response focuses on the notion of 'distinctness' that is employed to distinguish mental and physical properties: if 'distinctness' is understood one way, the exclusion principle on which the argument rests can be denied by the dualist; if it is understood another way, the argument is not persuasive.

1. Introduction

On one very simple formulation, the exclusion argument against dualism starts from the assertion that the following theses are inconsistent:

(1) Being in pain causes me to wince.

(2) Being in phys causes me to wince (where 'phys' denotes some overall physical state that I am in).

(3) Being in pain is distinct from being in phys.

(4) If being in phys causes me to wince, nothing distinct from being in phys causes me to wince.

The dualist is then invited to agree that (1) and (2) are empirical claims that are (in the context) non-negotiable; and that (4) is a principle of causation or an instance of a principle we must accept, often called 'the exclusion principle'. The conclusion of the argument is that (3)—a thesis distinctive of traditional dualism—has to go. Formally speaking, the reasoning goes: (1), (2), and (4) are true, so (3) is not true.

This argument is widely thought to put considerable pressure on dualism if not to refute it outright (see, for example, Bennett [2003] and Kim [1993, 1998]). In this paper, we argue to the contrary that, regardless of whether their position is ultimately true, dualists have a simple response, which targets the validity of the argument. We further show that this response is an instance of a very general strategy, which extends the way non-reductive physicalists sometimes respond to a similar argument, and which is of interest in its own right.

2. Complications set aside

In formulating the exclusion argument in terms of (1)–(4), we are deliberately setting aside some complications that are not central for our analysis. We will begin by enumerating what those complications are.

First, our presentation suggests that it is properties that are causal relata, but one may think that, strictly speaking, it is not properties that cause but instantiations of properties or events. We will ignore this point in what follows. Our exposition could be modified to accommodate it, but nothing turns on it.

Second, although we do not go into this issue, we assume throughout that 'causes' means 'is a sufficient cause for', rather than merely 'is causally relevant to'. This interpretation seems in line with how exclusion arguments are classically understood in the literature. Hence, for example, claim (4) says that if being in phys is a sufficient cause for some effect, then no property distinct from being in phys is a sufficient cause for it. To be even more precise, 'sufficient cause' here means 'sufficient cause in the circumstances'. For example, being in pain may be a sufficient cause for wincing only if certain background conditions are present, such as the absence of facial muscle paralysis; we may not wish to build all these background conditions into our specification of pain itself. For simplicity, we set this complication aside, taking the conditionalization on background circumstances, where required, to be understood.[1]

---

[1] Furthermore, in keeping with classic work on the exclusion argument, one might here wish to understand causation, broadly, as 'production' or 'generation' rather than as 'difference-making' (on the distinction, see, e.g., Hall [2004]). If we understood causation as difference-making (as discussed, e.g., in List and Menzies [2009]), then it would no longer be clear that premises (1) and (2) can simultaneously hold. On a difference-making account, a necessary condition for A to cause B is, roughly speaking, that:

(First conditional)   if A were present, B would also be present, and
(Second conditional)   if A were absent, B would also be absent.

In the case of (1), these conditionals seem relatively unproblematic. But in the case of (2), it is not clear that the second conditional is true. If I wasn't in phys, I might be in some related state phys' in which I might still wince, since, plausibly, the pain-triggering conditions are multiply realizable. More formally, the nearest possible worlds in which phys is absent may be ones in which phys' is present, which is also sufficient to make me wince. Thus only the mental state of being in pain may be a difference-making cause of my wincing, while the associated bodily state of being in phys might not [List and Menzies 2009]. By contrast, if we understand causation as production or generation, the falsity of the second conditional does not undermine causation. Aside from noting these complications, we here set aside the question of what the right account of causation is, or whether (1) and

Third, we are employing a simplified version of the exclusion principle in our formulation of (4) and are setting aside some nuances that do not matter for present purposes. In particular, we take (4) to be an instance of the following general principle:

(Exclusion)   If an effect E has a sufficient cause C, then it does not have any sufficient cause C* that is distinct from C and that occurs at the same time (except in cases of genuine overdetermination).

Since cases of genuine overdetermination—the classic example being the firing-squad case—are not relevant to our analysis, we set them aside and pretend, for expositional simplicity, that (4) is true outright rather than merely true 'in general'. Strictly speaking, however, claim (4), as formulated here, is a generalization that has, rather than lacks, exceptions. Furthermore, we omit the clause 'that occurs at the same time' in our formulation of (4). We take this qualification as understood. Relatedly, we are not carefully distinguishing direct from indirect causation. However, it seems implausible that (4), as stated, could be true if 'causation' were understood to mean 'either direct or indirect causation', for if A causes B, and B causes C, it may be that A causes C but not directly.

We ignore all these complications not because we consider them unimportant, but because attending to them properly would needlessly distract us from our main point: the connection between the exclusion argument and the notion of distinctness. It is to that point that we now turn.

3. Numerical and modal distinctness

Whether there is an inconsistency in (1)–(4) depends on how 'distinctness' is interpreted in claims (3) and (4). Of course, the same thing applies, *mutatis mutandis*, to the other terms in (1)–(4); we here concentrate on 'distinctness'.[2] There are a number of different but still

---

(2) are ultimately true. Our focus is on whether the exclusion argument itself is valid, or put differently, whether there is an indeed an inconsistency in (1)–(4).

[2] For example, another issue on which one might focus is the interpretation of what it is to be a physical property, and so what phys is exactly. In the main text, we assume, for the sake of argument, that there is a clear interpretation of what this amounts to, but notoriously it is not obvious that this is so. To illustrate, notice that (2) is plausibly an instance of a more general thesis sometimes called 'closure', the thesis (roughly) that any instantiated physical property that has a (sufficient) cause has a physical cause. What, then, is a physical cause, and what, for that matter, is a physical property? Suppose we define it with respect to the physical theories,

legitimate interpretations of 'distinctness' (see Stoljar [2007]; see also Sandford [2005]). Usually the term is used to express *numerical distinctness*, where:

> (Definition)   Two properties are *numerically distinct* if and only if they are non-identical, that is, one property itself has a property that the other lacks.[3]

But the term has philosophical uses in which it expresses relations distinct from non-identity. For example, when traditional dualists say that being in pain is distinct from being in phys, they certainly mean in part that being in pain is numerically distinct from being in phys, but they also mean more than this. In particular, they mean that being in pain is *modally distinct* from being in phys, where:

> (Definition)   Two properties are *modally distinct* if and only if it is possible for the first to be instantiated and not the second *and* vice versa.

That they mean this is shown, for example, by their commitment to the (metaphysical) possibility of zombies, creatures who, in our terminology, instantiate being in phys but not being in pain; and likewise it is shown by their commitment to the (metaphysical) possibility of ghosts, creatures who, in our terminology, instantiate being in pain but not being in phys.[4]

Similarly, consider those philosophers who adhere to the Humean principle that there are no necessary connections between distinct existences. Such philosophers cannot have in mind numerical distinctness, because otherwise their principle would be refuted by cases in

---

whatever they are, that are true in the actual world. In that case, closure will be false in worlds that are exact duplicates of this world but which have twin-mass and twin-charge rather than mass and charge (where 'twin-mass' is a property that plays a similar role as mass but is otherwise different). Suppose we define it with respect to the physical theories, whatever they are, that are true in some possible world or other. In that case, it is plausible that closure is consistent with various possibilities at which intuitively physicalism is false. As we indicate, we will set aside these issues here, but for some discussion, see Stoljar [2010].

[3] This definition of numerical distinctness implies the identity of indiscernibles: if F and G each have the same properties, then F and G are identical. Strictly speaking, this principle is not needed for our subsequent argument. It may be sufficient to assume the indiscernibility of identicals: if F and G are identical, then each have the same properties. But if the notion of a property is understood broadly enough, the identity of indiscernibles is not objectionable: for example, if F has the property of being F, and G has the same property, then obviously F and G must be identical. We thank an anonymous referee for prompting us to clarify this point.

[4] Of course, some positions sometimes called 'dualism' would deny these possibilities, e.g., so-called necessitarian or emergentist variants of dualism. We set aside such variants here.

which one thing necessitates another, that is, one thing is necessarily connected to a numerically distinct thing. It is better, then, to interpret them as intending a distinct notion of distinctness, for example modal distinctness.

For a down-to-earth illustration of how numerical distinctness and modal distinctness come apart, consider the properties 'being red' and 'being coloured'. Being red is numerically distinct from being coloured; but being red is not modally distinct from being coloured, since it is impossible that something is red without being coloured.[5]

4. Fine-grainedness and the inconsistency in (1)–(4)

Returning to (1)–(4), if the interpretation of 'distinctness' in (3) coincides with the one in (4), then the inconsistency between the four claims clearly arises, assuming no equivocation elsewhere. But the inconsistency may even arise when the two interpretations of 'distinctness' do not coincide, provided a particular condition is met, as we now show. Let us begin with a preliminary definition:

> (Definition)   One interpretation of 'distinctness' is *at least as fine-grained* as a second if any two properties that count as distinct under the second interpretation also count as distinct under the first.

For instance, numerical distinctness is at least as fine-grained as modal distinctness: any two properties that are modally distinct must also be numerically distinct. In fact, we can say something stronger. Call one interpretation of 'distinctness' *more coarse-grained* than a second if that second interpretation is at least as fine-grained as the first *but not vice versa*. Then modal distinctness is more coarse-grained than numerical distinctness. While any two modally distinct properties are also numerically distinct, the converse is not true: there are many properties that are numerically distinct, but not modally so—being red and being coloured, for instance. Generally, the binary relation 'at least as fine-grained as' partially orders different interpretations of 'distinctness'. (The ordering is partial, because, logically, there is no guarantee that any two interpretations of distinctness will be comparable in terms of fineness of grain.) Going back to (1)–(4), the following result holds:

---

[5] For another, more philosophical illustration (drawing on Fine [1994]), being Socrates is numerically distinct from being Socrates' singleton, but being Socrates is not modally distinct from being Socrates' singleton, since necessarily if the first is instantiated so too is the second.

(Proposition) Even if the interpretations of 'distinctness' in (3) and (4) may come apart, the inconsistency between (1)–(4) arises if and only if the interpretation of 'distinctness' in (4) is at least as fine-grained as the one in (3), assuming no equivocation elsewhere.

The proof of this point is straightforward. Suppose the interpretation of 'distinctness' in (4) is at least as fine-grained as the one in (3). Then, if being in pain and being in phys are distinct in the sense of (3), they are also distinct in the sense of (4) and hence (4) implies that they cannot both cause me to wince, contradicting (1) and (2). Conversely, suppose the interpretation of 'distinctness' in (4) is not at least as fine-grained as the one in (3). Then we cannot rule out—logically speaking—that being in pain is distinct from being in phys in the sense of (3), but not in the sense of (4), and consequently there is no inconsistency. We now show that these observations about distinctness give rise to a very general strategy by which we can respond to the exclusion argument in several of its variants.[6]

5. The exclusion argument and non-reductive physicalism

It is instructive to begin by considering a variant of the exclusion argument (for example, in Kim [1998, 2005]) that is directed not against dualism but against the distinct position of non-reductive physicalism. This is the thesis that being in pain and being in phys are numerically distinct even though the second necessitates the first. Like the dualist, the non-reductive physicalist is presented with (1)–(4) and then invited to agree that (1) and (2) are (in the context) non-negotiable and that (4) is a principle of causation that cannot be given up. The conclusion is that (3)—a thesis distinctive of non-reductive physicalism—has to go.

Granting—for the sake of argument—that (1) and (2) are true, the non-reductive physicalist may argue that, properly understood, (1)–(4) present no inconsistency. Here is why. If non-reductive physicalism is true, (3) had better mean (3-num) rather than (3-mod):

(3-num) Being in pain is numerically distinct from being in phys.
(3-mod) Being in pain is modally distinct from being in phys.

This is because non-reductive physicalists are not saying that it is possible for someone to be in phys and yet not to be in pain; their whole point is to deny that possibility. But if it is

---

[6] It is worth emphasizing that this strategy is different from the one based on difference-making causation defended in List and Menzies [2009]. Compare footnote 1.

(3-num) rather than (3-mod) that is in play, then, in order for the inconsistency to be maintained, (4) had better mean (4-num) rather than (4-mod):

(4-num)   If being in phys causes me to wince, nothing numerically distinct from being in phys causes me to wince.

(4-mod)   If being in phys causes me to wince, nothing modally distinct from being in phys causes me to wince.

As we have noted, modal distinctness is more coarse-grained than numerical distinctness, and thus interpreting (4) as (4-mod) in the presence of (3-num) would violate our necessary and sufficient condition for the inconsistency in (1)–(4) to arise. Now, however, the non-reductive physicalist may argue that there are reasons to reject (4-num) that are not also reasons to reject (4-mod). In light of these reasons, the non-reductive physicalist can assert that what is true is only (1), (2), (3-num), and (4-mod). And from this, nothing of interest follows.[7]

What are the reasons to reject (4-num) that are not also reasons to reject (4-mod)? Well, consider Yablo's pigeon, Sophie, who is trained to peck at a red card at the exclusion of others [Yablo 1993]. A red card is produced, and Sophie pecks. As Yablo notes, most people would unhesitatingly say that the redness of the card is what caused Sophie to peck. But of course red cards are not just red; they are specific shades of red—scarlet say. Surely being scarlet is a property of the card that is causally sufficient to get Sophie to peck, at least in the context. But then, by (4-num) or the principle behind it, being red cannot be a cause. If this is a bad result, and we want both being red and being scarlet to count as causes, (4-num) is false. By way of contrast, we should note that Yablo's pigeon is no counterexample to (4-mod). Being scarlet and being red are numerically distinct but they are not modally distinct. So, as far as this example goes, we may agree with Yablo that (4-num) is false, but still maintain the truth of (4-mod).

To be sure, every example is controversial in some way or another; but it is not necessary for our purposes to defend Yablo's example in detail. Rather we will assume that it is convincing, and in consequence that the exclusion argument against non-reductive physicalism is dialectically ineffective.

---

[7] The non-reductive physicalist may also assert that the exclusion argument seems successful to so many people only because they have failed to attend to the distinction between (4-num) and (4-mod).

6. Different kinds of modal distinctness

If non-reductive physicalists can use the distinctions in distinction to counter the variant of the exclusion argument that attacks them, can dualists do something similar? To make a case that they can, we now show that, while we can and should distinguish between numerical and modal distinctness, we may also distinguish among various kinds of modal distinctness. We begin by noting that there are different interpretations of 'possibility': metaphysical possibility, nomological possibility, and so on. As is well known, some interpretations of possibility are more permissive than others, rendering more things possible. For example, metaphysical possibility is more permissive than nomological possibility; the latter requires compatibility with the laws of nature, the former does not. Accordingly, something can be metaphysically possible without being nomologically possible. Indeed, that's what dualists like Chalmers [1996] would say about zombies. On that account, there are metaphysically possible worlds in which someone could be in phys without being in pain, but those worlds are not permitted by our actual laws of nature. Generally, a coherent scenario involving departures from the laws of nature may still count as metaphysically possible, even though it is not nomologically possible; if you like, think of a universe with different natural constants that is otherwise well-behaved.

Now observe the following: there exists not only a single interpretation of 'modal distinctness', but a family of such interpretations, one for each interpretation of 'possibility' that might be used in defining modal distinctness. For example, two properties are modally distinct in a metaphysical sense if it is metaphysically possible for each to be instantiated without the other also being instantiated; two properties are modally distinct in a nomological sense if this is nomologically possible; and so on.

So far, by 'modal distinctness' we have tacitly meant 'modal distinctness in the metaphysical sense'. Let us disambiguate our terminology by saying that two properties are *metaphysically distinct* just in case they are modally distinct in the metaphysical sense, while two properties are *nomologically distinct* just in case they are modally distinct in the nomological sense. To see that these two notions do not coincide, recall for instance that, for the dualist, being in pain and being in phys are metaphysically distinct but they need not be nomologically distinct. Dualists usually hold that psychological properties are nomologically connected to physical properties, because of certain psycho-physical laws that hold in the actual world (see, for example, Chalmers [1996]).

It should be clear that, just as numerical distinctness is more fine-grained than modal distinctness, so too metaphysical distinctness is more fine-grained than nomological distinctness. More generally:

(Proposition) Whenever one interpretation of 'possibility' is more restrictive than another—such as nomological possibility as compared to metaphysical possibility—then the resultant interpretation of modal distinctness is more coarse-grained under the first, more restrictive interpretation of possibility than under the second, less restrictive one.[8]

Before proving this claim, let us give an informal gloss. What this proposition says is that under a more restrictive interpretation of 'possibility' it is harder for two properties to count as modally distinct than it is under a more permissive interpretation of 'possibility'. And so an interpretation of modal distinctness resulting from a more restrictive notion of 'possibility' is more coarse-grained than one resulting from a more permissive notion of 'possibility'.

To prove the claim formally, suppose 'possibility$_1$' is more restrictive than 'possibility$_2$', in the sense that anything that is possible$_1$ is also possible$_2$ but not vice versa, and let 'distinctness$_1$' and 'distinctness$_2$' be the corresponding interpretations of modal distinctness. To show that 'distinctness$_1$' is more coarse-grained than 'distinctness$_2$', we need to show that the latter is at least as fine-grained as the former—call this claim (i)—but not vice versa—call this claim (ii). Suppose, then, that two given properties are distinct$_1$. This means that it is possible$_1$ for each to be instantiated without the other. Since this is possible$_1$, it is also possible$_2$, and so the two properties are also distinct$_2$. This establishes claim (i). To establish claim (ii), notice that the set of worlds that are possible$_1$ is a proper subset of the set of worlds that are possible$_2$. Let us assume (without much loss of generality) that there are at least two additional worlds in the second set. We can then 'construct' two properties which are co-instantiated across all worlds that are possible$_1$ but which suitably come apart in other worlds. Consider Figure 1. Let the large oval represent all worlds that are possible$_2$, and let the smaller oval inside represent all worlds that are possible$_1$. For labelling purposes, call the worlds in the small oval the R-worlds, those in the shaded left half-disk the Q-worlds, and those in the white right half-disk the S-worlds. (The assumption that there are at least two additional worlds outside the small oval is needed to ensure that the sets of Q- and S-worlds can each be chosen

---

[8] To be precise, for this claim to hold, the less restrictive interpretation of 'possibility' must render possible at least two additional worlds compared to the more restrictive interpretation, as explained in the proof below.

to be non-empty.) Now let P_1 be a property that is instantiated in all Q-worlds and all R-worlds (and nowhere else), and let P_2 be a property that is instantiated in all R-worlds and all S-worlds (and nowhere else). Then P_1 and P_2 are not distinct$_1$, because they are co-instantiated in all worlds that are possible$_1$, but they are distinct$_2$, because it is possible$_2$ for each to be instantiated without the other (namely, in the Q-worlds and the S-worlds, respectively). This completes the proof.
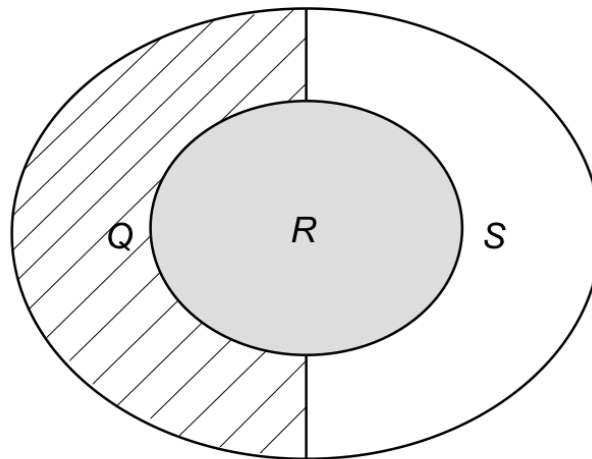
Figure 1: Some possible worlds

7. The exclusion argument and dualism

In light of the distinction between metaphysical and nomological distinctness, the dualist, like the non-reductive physicalist, may argue that (1)–(4), when properly understood, present no inconsistency. For the dualist, (3) had better mean (3-met) rather than (3-nom):

> (3-met)   Being in phys is metaphysically distinct from being in pain.
> (3-nom)   Being in phys is nomologically distinct from being in pain.

This is because dualists are not denying that psychological properties are nomologically connected to physical properties; on the contrary, as we have observed, they usually assert that they are (see, for example, Chalmers [1996]). On the other hand, if (3-met) rather than (3-nom) is in play, then, in order for the inconsistency to be maintained, (4) had better mean (4-met) rather than (4-nom):

(4-met)   If being in phys causes me to wince, nothing metaphysically distinct from being in phys causes me to wince.

(4-nom)   If being in phys causes me to wince, nothing nomologically distinct from being in phys causes me to wince.

For, otherwise, the condition for the inconsistency would be violated. But—and here is the point—the dualist may argue that there are reasons to reject (4-met) that are not also reasons to reject (4-nom). In the light of these reasons, the dualist may assert that what is true is only (1), (2), (3-met), and (4-nom). And from this, again, nothing of interest follows.[9]

8. Reasons to reject (4-met)

What are the reasons to reject (4-met) that are not also reasons to reject (4-nom)? To approach this issue, imagine that we have two properties F and F* that are nomologically connected to each other in a particularly dramatic way: they are 'nomologically co-extensive'; that is, in all possible worlds that share the same laws as our world, F and F* are co-instantiated. It does not follow from this that F and F* are numerically identical, since it might be that in a remote possible world, at which the laws that obtain at our world do not hold, something is F and not F* (or vice versa). But now suppose that as a matter of fact F is causally sufficient for some effect E. Is F* likewise causally sufficient? If we hold a version of an exclusion principle that we began with, the one that involves numerical distinctness— that is, (4-num)—the answer is no. For F and F* are numerically distinct; hence, according to (4-num), F* is not causally sufficient for E if F is. But on the other hand, this conclusion is implausible. If F and F* are nomologically co-extensive as we supposed, it may well be that both are causally sufficient for the effect. For example, to the extent that we are prepared to say, of F, that 'if it were not instantiated, E would not have happened', we should be prepared to say exactly the same thing of F*. Likewise, it may be that both properties are nomologically connected to E in the right way. More generally, if F and F* are nomologically co-extensive, the most natural thing to say is that either both are causally sufficient or neither is. (In fact, there is a perfectly natural sense in which the laws themselves may be said not to distinguish between F and F*.) It is implausible to conclude that neither property is causally sufficient; hence both are, and (4-num)—which obliges us to deny this—should be rejected.

---

[9] The dualist may also assert that the exclusion argument seems successful to so many people only because they have failed to attend to the distinction between (4-met) and (4-nom).

Now so far, you might say, so familiar. We already *had* a reason to reject (4-num); at best, what we have just done is provide a related reason for the same conclusion. However, a small variation on the point just considered gives us a reason to reject, not simply (4-num), but (4-met) as well. Imagine that there is not simply a remote possible world at which something is F but not F* but also a different but equally remote possible world at which the reverse is true. That is sufficient for showing, not simply that F and F* are numerically distinct but that they are metaphysically distinct as well. On the other hand, apart from this change, the situation is precisely the same as before. In particular, if we suppose that F is causally sufficient for E, we would be obliged by (4-met) to say that F* is not causally sufficient. And, as before, this is objectionable; since F and F* are nomologically co-extensive, a natural thing to say is that either both are causally sufficient or neither is. But it is implausible that neither is causally sufficient. Hence both are, and (4-met) should be rejected.

To reject (4-met) in this way, however, is not to reject (4-nom). What (4-nom) says is that if F causes E, then no property nomologically distinct from F causes E. But since F* is *not* nomologically distinct from F, the principle does not apply. More generally, we have a reason to reject (4-met) that is not a reason to reject (4-nom). More generally still, we have an answer to the exclusion problem from the point of view of the dualist.

9. A false principle?

It might be objected that the argument we just gave against (4-met) relies on a principle that is false. The principle asserts, or implies, that if F is causally sufficient for E, and F is nomologically co-extensive with F*, then F* is also causally sufficient for E. And that principle, so the objection goes, is subject to counterexample. Suppose that F* is the conjunctive property of being F and being such that 2+2=4. In that case, F* is nomologically co-extensive with F and yet, one might think, it is not causally sufficient for E—not because it is not sufficient in some sense or other, but because it is not even a causally relevant property. In particular, note the irrelevance of the second conjunct, 'being such that 2+2=4'.[10]

We have two things to say in response to this objection. In the first place, we are not convinced that F and F* as just defined are genuinely distinct properties at all. Here is why. For a start, F and F* are not merely nomologically co-extensive, but metaphysically so; it is metaphysically impossible for either of them to be instantiated without the other. But even

---

[10] We are indebted to an anonymous referee for raising this objection.

more importantly, 'F*' might be viewed as a more cumbersome (though extensionally equivalent) re-description of F, which simply adds to the original description a redundant conjunct, namely, a tautological one. On a plausible criterion of property individuation, 'F' and 'F*' might therefore be said to pick out the same property. Plausibly, properties should be individuated more coarsely than property descriptions: for each property, there exists an entire equivalence class of possible descriptions. For example, we would not want to treat the conjunctive property descriptions 'F and F', 'F and F and F', and so on, as picking out properties distinct from F itself. Rather, 'F', 'F and F', and 'F and F and F' are just different descriptions of the same property, some of which are more parsimonious than others. If this is right, it would not be a mistake to say that F* is causally sufficient for E, if F itself is. It would just be more cumbersome.

But in the second place, our argument against (4-met) does not rely on the principle targeted by the present objection. Recall that that principle asserts (or implies) that if F is causally sufficient for E, and F is nomologically co-extensive with F*, then F* is also causally sufficient for E. We need not take a stand on whether this principle is true. Our claim is rather two-fold: first that (4-num) should be rejected since it entails that if F is causally sufficient for E in the circumstances, then F* is *not* causally sufficient (when in fact we would not always want to rule out F*'s causal sufficiency); and second that if (4-num) should be rejected on those grounds, then (4-met) should be rejected *mutatis mutandis*.

In virtue of what might two nomologically co-extensive properties, such as F and F*, each be causally sufficient for E? That is clearly a difficult question in the philosophy of causation, but one suggestion that is consonant with our discussion is that they are both causally sufficient for E in virtue of their being nomologically connected to E in the right way, or exhibiting the right pattern of counterfactual connection to E, and so forth. Of course, there are some (very big) philosophical questions about these conditions; we will not go into them here. The important point for present purposes is this. It is often accepted that the non-reductive physicalist may say that a physical property and a mental property that is supervenient on it are both causally sufficient for some effect E on the grounds of both standing in the right nomological or counterfactual relation to E.[11] It is then hard to see why the dualist may not say

---

[11] As Yablo puts it [1992: 273], 'for all that the exclusion argument shows, mental phenomena *can* be causally relevant compatibly with the causal sufficiency of their physical bases'. Yablo carefully qualifies this claim and also distinguishes causation *simpliciter* from causal sufficiency. Causes, he says, must be '*commensurate* with their effects: roughly, they should incorporate a good deal of causally important material but not too much that is

the analogous thing, where the metaphysical modality associated with supervenience is replaced by a nomological one, with everything else remaining structurally the same, especially as we focus on worlds that have the same laws of nature as ours.

10. Lack of example?

It might also be objected that the reason for rejecting (4-met) is importantly different from the reason for rejecting (4-num) considered earlier. In that earlier case, the reason was a straightforward counterexample in the form of Yablo's pigeon, whereas in the new case, our argument is more abstract: we have not given an explicit counterexample to (4-met). However, there are two reasons why this does not undermine the plausibility of the suggested response of the dualist to the exclusion argument.

In the first place, the reason to reject (4-met) builds on the reason to reject (4-num). What Yablo's example brings out is that a version of the exclusion principle that invokes numerical distinctness is implausible precisely because the opposite of distinctness in that sense (that is, identity) is such an unforgiving relation: all it takes for F to be numerically distinct from G is for F to have one feature that G does not. But a similar point applies to metaphysical distinctness: all it takes for F to be metaphysically distinct from G is for there to be possible worlds at which F obtains and G does not and vice versa. But since the worlds in question can be as remote as you please, it becomes hard to see how their existence can have any impact on causal facts that are sensitive only to worlds nomologically similar to our own.

In the second place, while we have not relied on examples to defeat (4-met), it is not impossible to do so. We here offer just two sketchy illustrations; nothing much depends on them, however. Suppose, for example, I am in an overall physical condition that includes as a proper part the fact that a particular neuronal cluster in my brain has property P1—call this condition NP1. It seems plausible to say that my being in NP1 is causally sufficient to get me to wince; at any rate, it is true that if I were in NP1, I would wince. Now contrast this condition with a metaphysically distinct condition NP2. NP2 is exactly like NP1 except that it includes as a proper part the fact that the same neuronal cluster has property P2. NP2 is nomologically

---

causally unimportant' [1992: 274]. This might, in turn, lead us to conclude that *only* the mental property, and not the physical one, is a cause *simpliciter*, while both properties might be causally sufficient in the more permissive sense (see also List and Menzies [2009]). Be that as it may, it seems fair to suggest that, according to non-reductive physicalism, mental sufficient causes may (under certain conditions) coexist with physical sufficient causes, provided both stand in the right relationship to one another and to the effect in question.

connected to NP1 because, relative to the actual biological laws governing human brains, the co-occurrence of NP1 and NP2 is nomologically necessary. However, they are metaphysically distinct because there are more remote possible worlds—worlds with a subtly different brain biology—at which P2 is instantiated and P1 is not (and vice versa). It seems plausible to say that my being in NP2 is causally sufficient to get me to wince too; it is also true that if I were in NP2, I would wince. But that could not be true if (4-met) is true.

For another example, suppose a junior lecturer loses his job because the university rejects his application for tenure; hence the university's making this decision—a property of the university—causes the lecturer's job loss. But suppose that, as a robust empirical regularity—a 'social-scientific law'—the committee that votes on these matters always consists of the most successful professors; they are reliably chosen to serve on the committee. Then it is also true that the most successful professors' judgment—a property of the professors—causes the lecturer's job loss. In this example, the university's decision is nomologically connected to the most successful professors' judgment: first, as we have assumed, there is a robust regularity ensuring that those professors are the members of the relevant committee; and, second, we may assume that a dismissal of a junior lecturer without that committee's vote is ruled out by the university's by-laws. But despite this nomological connection, the two apparently competing causes are clearly metaphysically distinct: for example, it is metaphysically possible for other professors to serve on the committee, or for the by-laws to be different, so that the judgments of the most successful professors do not influence the decision. So (4-met) implies that the university's decision and the most successful professors' judgment cannot both cause the lecturer's job loss. If this implication is unacceptable, we have reason to deny (4-met). At the same time, there is no conflict with (4-nom), since the two properties in question are nomologically connected.

Of course, these two examples, like any, may be controversial in some way or another. But, on the face of it at least, they do seem to make the point the dualist wants to make. And in any case, our official reason for rejecting (4-met) without equally rejecting (4-nom) does not depend on either example.

11. Conclusion

We have argued that in response to the exclusion problem, a dualist could say that, whereas dualism requires psychological properties to be modally distinct from physical properties in a metaphysical sense, any plausible version of the exclusion principle refers only to properties

that are modally distinct in a nomological sense. The dualist could then go on to point out that being in pain and being in phys are modally distinct in a metaphysical sense but not in a nomological sense and that, therefore, the exclusion problem does not arise.

But it is important to emphasize two limitations on our discussion. First, we have focused on what we may call 'strong' variants of the exclusion argument, which try to establish that it is not the case that being in pain causes wincing if being in phys does. But there might be variants of the argument that try to show something less, for example, *that we have no reason to suppose* that being in pain causes wincing if being in phys does. We have said nothing about those weaker, epistemic variants of the exclusion argument. Our point is only that if one is convinced of dualism, then there is a reply available to the kind of 'strong' exclusion argument we have focused on.

Second, we have focused only on what someone convinced of dualism could say to the exclusion argument. But we have said nothing about whether anybody should be convinced of dualism. For example, dualism is usually interpreted as being committed to the view that consciousness is a fundamental element in nature, akin to basic fundamental physical features. Many philosophers find this very implausible on its face, quite apart from any technical issues raised by the exclusion argument. There is nothing in the present paper to suggest that those who find dualism implausible are wrong. Moreover, many think that the positive arguments for dualism are not convincing (see, for example, Stoljar [2006]). However, even if there is no reason to believe dualism—indeed even if dualism is false—it may also be that one of the key arguments against it has a plausible answer.

References

Bennett, Karen 2003. Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract it, *Noûs* 37/3: 471–97.

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

Fine, Kit 1994. Essence and Modality, in *Philosophical perspectives 8: Logic and Language*, ed. James Tomberlin, Atascadero: Ridgeview Publishing: 1–16.

Hall, Ned 2004. Two Concepts of Causation, in *Causation and Counterfactuals*, ed. Ned Hall and L. A. Paul, Cambridge, MA: MIT Press: 225–76.

Kim, Jaegwon 1993. *Supervenience and Mind*, Cambridge: Cambridge University Press.

Kim, Jaegwon 1998. *Mind in a Physical World*, Cambridge: Cambridge University Press.

Kim, Jaegwon 2005. *Physicalism or Something Near Enough*, Princeton: Princeton University Press.

List, Christian and Peter Menzies 2009. Non-Reductive Physicalism and the Limits of the Exclusion Principle, *Journal of Philosophy* 106/9: 475–502.

Sanford, David H. 2005. Distinctness and non-identity. *Analysis* 65/4: 269–74.

Stoljar, Daniel 2006. *Ignorance and Imagination: The Epistemic Origin of the Problem of Consciousness*, New York: Oxford University Press.

Stoljar, Daniel 2007. Distinctions in Distinction, *Being Reduced: New Essays on Causation and Explanation in the Special Sciences*, ed. Jesper Kallestrup and Jakob Hohwy, Oxford: Oxford University Press: 263–79.

Stoljar, Daniel 2010. *Physicalism*, London: Routledge.

Yablo, Stephen 1992. Mental Causation, *The Philosophical Review* 101/2: 245–80.