Does the exclusion argument put any pressure on dualism?

Christian List and Daniel Stoljar*

August-October 2014

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’ as it occurs in the argument: if 'distinctness' is understood one way, the exclusion principle on which the argument is founded can be denied by the dualist; if it is understood another way, the argument is not persuasive.

§1. 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\(^1\) causes me to wince.
3. Being in pain is distinct from being in phys.
4. If being in pain causes me to wince, nothing distinct from being in pain 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.\(^2\) Formally speaking, the reasoning goes: (1), (2) and (4), so not-(3). This argument is widely thought to put considerable pressure on dualism if not to refute it outright. In this paper we argue to the contrary that, whether or not 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. The first thing to note is that not all responses to the exclusion argument target its validity. For example, whether premises (1) and (2) are true depends, in part, on how we understand causation. In particular, it depends on whether we understand it, broadly, as ‘production’ or as ‘difference-making’. If we understand causation as difference-making, then 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. 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). Proponents of the exclusion argument may therefore prefer to adopt an alternative account of causation as ‘production’ or ‘generation’, under which the falsity of the second conditional does not undermine causation; the details do not matter for present purposes. We here set aside the question of what the right account of causation is, or whether (1) and (2) are true. As noted, we focus on whether the exclusion argument itself is valid, or put differently, whether there is an indeed inconsistency in (1)-(4).

§3. 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
There are a number of different but still legitimate interpretations of ‘distinctness’ (see Stoljar 2007; see also Sandford 2005). Usually the term ‘distinctness’ is used to express numerical distinctness, where:

(Definition) Two properties are numerically distinct if and only if they are non-identical.

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.

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 which one thing entails another, i.e., 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.

§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 say that:
(Definition) One interpretation of ‘distinctness’ is *at least as fine-grained* as another if any two properties that count as distinct in the second sense also count as distinct in the first.

The binary relation ‘at least as fine-grained as’ partially orders different interpretations of ‘distinctness’. We further say that one interpretation of ‘distinctness’ is *more coarse-grained* than another if the second interpretation is at least as fine-grained as the first and not vice versa. Then:

(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 more coarse-grained than the one in (3). Then it is possible for being in pain to be 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.

§5. It is instructive to begin by considering a variant of the exclusion argument (e.g., in Kim 1998 and Kim 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.

Again setting aside the issue of whether (1) and (2) are true, as discussed above, the non-reductive physicalist may argue that, properly understood, (1)-(4) present no inconsistency. For the non-reductive physicalist, (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 it is possible for someone to be in phys and yet not be in pain; their whole point is to deny that possibility.

But if it is (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 pain causes me to wince, nothing numerically distinct from being in pain causes me to wince.

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

In the terms of the present paper, 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 counterexamples to (4-num) that are not counterexamples to (4-mod). In light of these counterexamples, 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.

What are the counterexamples to (4-num) that are not counterexamples to (4-mod)? One plausible example is Yablo’s pigeon, Sophie, who is trained to peck at a red card at the exclusion of others (see 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 the red and the 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 hang on to the truth of (4-mod).

To be sure, every example is controversial in some way or another; it may be that Yablo’s particular example does not convince everyone. Still, it is difficult to shake the feeling that there is some example that will make the point the non-reductive physicalist wants to make. If so, it is difficult to see the exclusion argument against non-reductive physicalism as dialectically effective.

§6. 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, notice that, while we can and should distinguish between numerical and modal distinctness, we may also distinguish among various kinds of modal distinctness. In particular:
(Observation) 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
nomologically sense. 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.6

To prove this claim, 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 \(Q\)-worlds, those in the shaded left half-disk the \(R\)-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 \(R\)- and \(S\)-worlds can each be chosen to be non-empty.) Now let \(P_1\) be a property that is instantiated in all \(Q\)-worlds and all \(R\)-worlds, and let \(P_2\) be a property that is instantiated in all \(Q\)-worlds and all \(S\)-worlds. 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 \(R\)-worlds and the \(S\)-worlds, respectively). This completes the proof.

Figure 1
§7. In light of the distinction between modal 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 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, they usually assert that they are (e.g., 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 pain causes me to wince, nothing metaphysically distinct from being in pain causes me to wince.
- (4-nom) If being in pain causes me to wince, nothing nomologically distinct from being in pain 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 counterexamples to (4-met) that are not counterexamples to (4-nom). In the light of these counterexamples, 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.

§7. What are the counterexamples to (4-met) that are not also counterexamples to (4-nom)? We sketch three illustrative examples.

§7. 1. You cross the road because the traffic stopped; hence being stationary (a property of the traffic) is causally relevant to your crossing. But it is also true that you cross the road because the (pedestrian-facing) streetlight is green; hence being green (a property of the light)
is also relevant to your crossing. Of course, the colour of the light is nomologically connected
to the motion of the traffic. It is a law (or so we suppose) that drivers stop when they see a red
light, which, in turn, is connected with a pedestrian-facing green one. On the other hand, if
(4-met) is true, then, if being stationary is causally relevant to your crossing, being green is
not. For being stationary and being green are metaphysically distinct. To the extent that this
result is unacceptable (because being stationary and being green were both causally relevant
to your crossing the road), we have a reason to deny (4-met). On the other hand, the example
gives us no reason at all to deny (4-nom) because the two properties at issue in this example
are connected by law.

§7.2. Suppose I am in an overall physical condition that includes as a proper part the fact
that a particular nerve n in my big toe has property P1—we may 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
numerically distinct condition NP2. NP2 is exactly like NP1 except that it includes as a
proper part the fact that the particular nerve n in my big toe has property P2. It seems
plausible to say that my being in NP2 is causally sufficient to get me to wince too; it is also
ture that if I were in NP2, I would wince. In short, what we have here are two numerically
distinct conditions which are causally sufficient to get me to wince. Moreover, there is no
reason to suppose that NP1 and NP2 are in causal competition. If NP1 is causally sufficient to
get me to wince, there is no reason to say that NP2 is not also were I to be in it. Of course
there would be such a reason if (4-num) were true; but we have already supposed it is not. At
any rate, if we say that NP1 is causally relevant to get me to wince, (4-num) will tell us that
NP2 is not. And since this is objectionable we should reject (4-num).

So far NP1 and NP2 are examples of properties that are numerically distinct which are
not in causal competition; and this adds nothing to what we have not seen already. However,
suppose now that NP1 and NP2 are not only numerically distinct, but also metaphysically distinct; that is, there are metaphysically possible worlds in which the nerve in my big toe has property P1 but not P2 and vice versa; hence there are metaphysically possible worlds in which I am in NP1 but not NP2 and vice versa. However, suppose now that as a matter of fact I am in both conditions and moreover that at it is a law of nature that I am: suppose it is a law that nerves satisfying property P1 also satisfy P2 and vice versa. It remains the case that NP1 and NP2 are not in causal competition but now it is the case that they are nomologically connected and metaphysically distinct. This suggests that we have a counterexample to (4-met) which is not a counterexample to (4-nom).

§7. 3. 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 regularly elected 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. But although the university’s decision is nomologically connected to the most successful professors’ judgment, the two are clearly metaphysically distinct (it is metaphysically possible for other professors to serve on the committee). 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 examples are just illustrative, and they may each be controversial in some way or another. But again, it is difficult to shake the feeling there must be some example that
will make the point the dualist wants to make. And if so, it is hard to see that the exclusion argument does put the pressure on dualism it is usually thought to do.

§ 8. 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, 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 one and that, therefore, the exclusion problem does not arise.

It is important to emphasize to 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 extremely implausible on its face, quite apart from any technical issues raised by the exclusion argument. There is nothing in the present paper to say that those who find dualism implausible are wrong. Moreover, many think that the positive arguments for dualism are not convincing (see, e.g., 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


* C. List, Departments of Government and Philosophy, London School of Economics; D. Stoljar, School of Philosophy, Australian National University. An earlier version of this paper, titled “What a dualist should say about the exclusion argument”, was circulated in February 2007.
1 Where ‘phys’ denotes some overall physical state that I am in.

2 In setting out the argument this way, we are deliberately ignoring some complications. First, we are not being very careful about causal relata. One may think that, strictly speaking, it is not properties that cause but instantiations of properties or events. Second, we are not carefully distinguishing direct from indirect causation: it seems implausible that (4) could be true if ‘causation’ were understood broadly 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. Third, we are pretending that (4) is true outright rather than merely true ‘in general’: (4), or the principle behind it, suggests that genuine overdetermination—the classic example is the firing squad case—is being ruled out a priori, but since this is implausible, (4) must be a generalization that has, rather than lacks, exceptions. We ignore these complications here 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.

3 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 property that has a cause has a physical cause. What then is a physical cause? Suppose we define it with respect to the physical theories, 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 a that plays a similar role as mass but is otherwise different). Suppose we define it with respect 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).

4 That they mean this is shown by their commitment to the possibility of zombies, creatures who, in our terminology, instantiate being in phys but not being in pain. Of course, some positions sometimes called ‘dualism’ would deny this possibility, e.g., so-called necessitarian or emergentist variants of dualism. We set aside such variants here.

5 Numerical distinctness and modal distinctness are distinct. 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. Similarly, to vary Fine’s 1994 example, 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.

6 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.