

What a dualist should say about the exclusion argument

Christian List and Daniel Stoljar*

8 February 2007

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 phys¹ 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 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.

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 concentrate here on ‘distinctness’.) As one of us has argued elsewhere, there are a number of different but still legitimate interpretations of ‘distinctness’ (see Stoljar 2006; see also Sandford 2005). Usually the term

* C. List, Department of Government, London School of Economics, London WC2A 2AE, U.K.; D. Stoljar, Philosophy Program, Research School of Social Sciences, Australian National University, Canberra ACT 0200, Australia.

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.

‘distinctness’ is used to express *numerical distinctness*, where 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 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*.³ 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, for otherwise their principle would be refuted by pointing out that one thing entails another, i.e., that one thing is necessarily connected to a numerically distinct thing. Better then to interpret them as intending a distinct notion of distinctness, for example modal distinctness.⁴

Returning now 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 the following condition is met. Let us say that 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’.⁵) Then the inconsistency between (1)-(4) arises whenever the interpretation of ‘distinctness’ in (4) is at least as fine-grained as the one in (3), again assuming no equivocation elsewhere. By contrast, if 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 under the interpretation

³ 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 versions of dualism. We set aside such versions here.

⁴ 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.

⁵ 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.

of ‘distinctness’ in (3), but not under the one in (4), and consequently there is no inconsistency.

These observations about distinctness may be exploited to defend a position that, while distinct from dualism, is often thought (e.g., in Kim 1998 and Kim 2005) to confront the exclusion problem too, viz., non-reductive physicalism, 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.

In response to the argument, 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.

For of course 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. On the other hand, 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.

But, notoriously there are counterexamples to (4-num) that are not counterexamples to (4-mod). In the light of these counterexamples, the non-reductive physicalist can assert that, by their lights, 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 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 is not relevant. If this is a bad result, and we want both the red and the scarlet to be causally relevant, (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).

If non-reductive physicalists can use the distinctions in distinction to counter the version of the exclusion argument that attacks them, might dualist do something similar? We think so. The key observation is that, while we can and should distinguish between numerical and modal distinctness, we may also distinguish among various kinds of modal distinctness. In particular, 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.

Now so far, by ‘modal distinctness’ we have tacitly meant ‘modal distinctness in the metaphysical sense’. So let us fix terminology by saying that two properties are *modally 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. It should be clear that, just as numerical distinctness is more fine-grained than modal distinctness, so too modal distinctness is more fine-grained than nomological distinctness. More generally, whenever one sense of ‘possibility’ is more restrictive than another—such as nomological possibility as compared to metaphysical possibility—then the resultant definition of distinctness is more coarse-grained under the first, more restrictive interpretation of possibility than under the second, less restrictive one.

In the 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-mod) rather than (3-nom):

- (3-mod) Being phys is modally distinct from being in pain.
- (3-nom) Being in phys is nomologically distinct from being in pain.

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

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

But—and here is the main point—the dualist may argue that there are counterexamples to (4-mod) that are not counterexamples to (4-nom). In the light of these counterexamples, the dualist may assert that, by their lights, what is true is only (1), (2), (3-mod) and (4-nom). And from this, again, nothing of interest follows.

What are the counterexamples to (4-mod) that are not also counterexamples to (4-nom)? Here is one. 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-mod) is true, then, if being stationary is causally relevant to your crossing, being green is not. For being stationary and being green are modally 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-mod). 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.

In summary, in response to the exclusion problem, a dualist should argue 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 should then go on to point out that being in pain and being in phys are strongly modally distinct in a metaphysical sense but not in a nomological one and that, therefore, the exclusion problem does not arise.

References

- Chalmers, D. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Kim, J. 1998. *Mind in a Physical World*. Cambridge: Cambridge University Press.
- Kim, J. 2005. *Physicalism or Something Near Enough*. Princeton: Princeton University Press.
- Fine, K. 1994. Essence and Modality. In J. Tomberlin, ed., *Philosophical perspectives 8: Logic and Language*. Atascadero: Ridgeview Publishing. 1-16.
- Sanford, D. H. 2005. Distinctness and non-identity. *Analysis* 65: 269-274.
- Stoljar, D. 2007. Distinctions in Distinction. In J. Kallestrup and J. Hohwy, eds., *Being Reduced: New Essays on Causation and Explanation in the Special Sciences*. Oxford: Oxford University Press.
- Yablo, S. 1992. Mental Causation. *The Philosophical Review* 101: 245-280.