Dualism and Exclusion

Bram Vaassen

Abstract

Many philosophers argue that exclusion arguments cannot exclude non-reductionist physicalist mental properties from being causes without excluding properties that are patently causal as well. List and Stoljar (2017) recently argued that a similar response to exclusion arguments is available to dualists as well, thereby challenging the predominant view that exclusion arguments undermine dualist theories of mind. In particular, List and Stoljar maintain that exclusion arguments against dualism require a premise that states that, if a property is metaphysically distinct from the sufficient cause of an effect, this property cannot be a cause of that effect. I argue that this premise is indeed likely to exclude patently causal properties, but that exclusion arguments against dualism do not require this premise. The relation that enables metaphysically distinct properties to cause the same effect in the relevant way turns out to be tighter than the relation typically posited between dualist conscious properties and their underlying physical properties. It is therefore still plausible that the latter causally exclude the former and that compelling exclusion arguments against dualism can be formulated by using a weaker exclusion premise. I conclude by proposing such a formulation.

Keywords: Mental Causation, Dualism, Non-Reductionism, Causal Exclusion, Exclusion, Exclusion Argument, Overdetermination

1 Introduction

Some philosophers argue that all effects have sufficient physical causes and that properties which cannot be reduced to the physical are therefore excluded from being causes.[1] They conclude from this that non-reductionist theories of the mind, such as dualism and non-reductionist physicalism, are false (e.g Kim, 1989; Papineau, 2002). Non-reductionist phsyicalists often contend that such *exclusion arguments* cannot exclude non-reductionist physicalist mental properties from being causes without excluding properties that are patently causal, such as being a hurricane, or having an infection, from being causes as well (e.g. Block, 2003; Woodward, 2008; Yablo, 1992). Consequently, it is a matter of dispute whether exclusion arguments pose a serious problem for non-reductionist physicalism.[2] There *is* broad agreement, however, that such arguments undermine dualism (e.g. Bennett, 2008; Loewer, 2001).[3]

List and Stoljar (2017) challenge that predominant view. They maintain that exclusion arguments against dualism require an implausible exclusion premise, which states that if a property is metaphysically distinct from the sufficient cause of an effect, this property cannot be a cause of that effect. List and Stoljar argue that this premise excludes patently causal (and non-mental) properties from being causes and must therefore be false. By doing so, they provide a more concrete formulation of the suggestion made by some dualists that, if non-reductionists have a convincing reply to exclusion worries at their disposal, the dualist must have a similar reply available as well (e.g. Koons and Bealer (2010a, p. xix–xx); Pautz (2010, p. 65)). List and Stoljar conclude that dualists have a convincing reply to exclusion arguments.

I argue that this conclusion is too hasty. Metaphysically distinct properties *can* cause the same effects, but only if they stand in a sufficiently tight relation to one

---

[1] I adopt some working assumptions from List and Stoljar (2017). First, I take causes to be sufficient for their effects given some fixed set of background conditions, thereby ignoring the possibility of so-called 'contributory' causes. Second, I assume that properties and property instances can be causal relata. Third, I move freely between claims about properties and property instances. The ideas presented here can be reformulated straightforwardly in more precise language and their plausibility is independent of the nature of causal relata. I will sometimes use the relevant qualifiers as a reminder and I will be precise where necessary.

[2] See Kim (2007) and Ney (2012) for exclusionist rejoinders.

[3] See also, interestingly, Stoljar (2008, p. 270–271).

another. Dualists typically deny that conscious properties stand in such a tight relation to their underlying physical properties. Plausibly, one can exploit this difference in tightness to formulate an exclusion argument that excludes dualist conscious properties from being causes without excluding patently causal properties from being causes. I propose such a formulation and conclude that exclusion worries for dualism persist.

2   Dualism, Exclusion and Distinctness

Setting aside some details that will not matter for our purposes, the exclusion argument that List and Stoljar target runs as follows (2017, p. 96):

(1) Some instances of physical properties are caused by instances of conscious properties.

(2) Every instance of a physical property has a sufficient physical cause at any given time t (if it has a cause at all at t).[4]

(3) Every instance of a conscious property is distinct from any instance of a physical property.

(4) If an effect has a sufficient cause at a time t, nothing distinct from its sufficient cause is a cause of that effect at t.

This set of propositions is inconsistent. (1) states that there are conscious causes for some physical effects. (2) entails that these physical effects have a sufficient physical cause. (4) states that nothing distinct from that physical cause can be a cause of that effect. Lastly, (3) states that the conscious cause *is* distinct from that physical cause. If the dualist is to secure mental causation, she will have to deny (2) or (4).

List and Stoljar take issue with (4), the exclusion premise. Before developing their criticism, they explicitly set aside the possibility that effects of conscious property instances are *genuinely overdetermined*. In cases of genuine overdetermination, there *is* a cause that is distinct from a simultaneous sufficient cause of

---

[4]I drop the time-index from here on. When talking about competing causes, I take these to be simultaneously instantiated.

its effect, but that cause is *redundant* for the occurrence of the effect. For example, two bullets simultaneously piercing a victim's heart can both be causes of the victim's death, but each one is redundant for this effect given the presence of the other. It is generally taken to be implausible that all conscious causes are redundant for their effect. Consequently the possibility of genuinely overdeterministic causation is typically set aside in mental causation debates (e.g. Bennett, 2008; Papineau, 2002). List and Stoljar maintain that, even if one focuses solely on non-overdeterministic causation, (4) is still objectionable.

In particular, they argue that, if one renders the required notion of 'distinctness' in (4) more precise, it becomes clear that this principle is implausible. They define the relevant notion of distinctness as follows (2017, p. 98):

> **Modal Distinctness** Two properties are *modally* distinct if and only if it is possible for the first to be instantiated without the second *and* vice versa.

As List and Stoljar point out, this definition is ambiguous. Its ambiguity lies in the modal strength of possibility. Is it *metaphysically* possible for one of the two properties to be instantiated without the other, like a *perpetuum mobile* is possible, but a square circle is not? Or is it *nomologically* possible for one to be instantiated without the other, like a whale-sized diamond is possible, but a *perpetuum mobile* is not?

In the current context, it is important to distinguish between these two meanings of modal distinctness. Many contemporary dualists claim that conscious properties are nomologically necessitated by physical properties in the actual world and thus explicitly *deny* that they are nomologically distinct. Such a 'naturalist' dualism has become fairly popular, and is defended by Jackson (1982); Robinson (1988); Chalmers (1996) and several contributions in Koons and Bealer (2010b), such as Pautz (2010), Hasker (2010), and Koons and Bealer (2010a, p. xvi). Consequently, it is taken to be the standard brand of dualism in mental causation debates (e.g. Stoljar (2008, p. 270); Bennett (2008); Kroedel (2015)). If the exclusion argument is to target that standard dualism, a more precise formulation of (3) should read:

> **Dualism** Every instance of a conscious property is *metaphysically* distinct from the underlying instances of physical properties.

List and Stoljar argue that, if we formulate (4) with similar precision, we can see that this exclusion premise is false. For the set of propositions to be inconsistent, a more precise formulation of (4) should read:

> **Metaphysical Exclusion** If an effect has a sufficient cause at a time t, nothing *metaphysically* distinct from its sufficient cause is a cause of that effect at t.

They argue that *Metaphysical Exclusion* excludes patently causal (and non-mental) properties from being causes. They conclude that *Metaphysical Exclusion* is false and the dualist's exclusion worries therefore disappear. I disagree. *Metaphysical Exclusion* is false, but the exclusion worries persist.

3   *Metaphysical Exclusion* is false, ...

Metaphysically distinct property instances *can* cause the same effect. Let us say that property instances that cause the same effect 'co-cause' that effect. Now consider the following case derived from List and Stoljar (2017, p. 105). A certain university is organized such that the committee delegated to make tenure decisions always consists of the most successful professors. Given this organizational structure, these professors making a negative decision simultaneously makes it the case that the university made a negative decision. If, in such a case, an applicant loses her job due to the university's decision being negative (UD), the most successful professors' decision being negative (PD) would *also* count as a cause, despite its being metaphysically distinct from UD. The example thus provides us with metaphysically distinct property instances that co-cause an effect and thereby disproves *Metaphysical Exclusion*.

Moreover, the example lines up well with the philosophical literature on *realization*. This allows us to embed the counterexample to *Metaphysical Exclusion* in an established theoretical framework and affords us a closer look at what enables the properties to co-cause in such cases.

In cases of realization, one can distinguish between the realized property, its *total* realizer and its *core* realizer (cf. Shoemaker, 2007, p. 21–22). The realized property can be any non-fundamental property, like UD. The *total* realizer of the

realized property is typically a large and complex set of properties that metaphysically necessitates the realized property and is therefore *not* metaphysically distinct from it. The total realizer of *UD* for example, will include *PD* as well as relatively permanent properties, such as the organizational structure of the university. The *core* realizer is a salient part of this total realizer, such as *PD*. This property *is* metaphysically distinct from the realized property, because its instances do not on their own suffice for instances of the realized property. For example, there are possible worlds where the professors make the same decision, but the university is organized differently and the university makes a different decision.

Despite this metaphysical distinctness, core realizers and their realized property can co-cause effects. At least, this is what we should conclude if we adopt two widely accepted heuristic principles. Let p be a property instance and P the proposition that p occurs, and similarly for q and Q. The first principle is that p non-overdeterministically causes q if and only if Q counterfactually depends on P.[5] The second principle is that counterfactual dependence is defined in accordance with Lewis (1973a):

Counterfactual Dependence Q counterfactually depends on P *iff*
$\quad$ P $\Box\!\!\rightarrow$ Q and $\frown$P $\Box\!\!\rightarrow$ $\frown$Q

where the semantics of $\Box\!\!\rightarrow$ are such that

$\quad$ P $\Box\!\!\rightarrow$ Q is true *iff* there is a possible world where P and Q hold which is closer to the actual world than any possible world where P and $\frown$Q hold (or there are no possible worlds where P holds).

In mental causation debates, it is customary to also adopt Lewis's analysis of the closeness of possible worlds, which relies on the amount and the size of the 'miracles' that separate possible worlds from the actual world.[6] Miracles are to be understood as violations of nomological laws, such as gravity locally failing

---

[5]List and Stoljar rely on this heuristic as well (2017, p. 103–104). Remember that they set aside the issue of overdetermination.

[6]See Bennett (2008); Kroedel (2015). This analysis is not uncontroversial, as it can be difficult to assess whether one miracle creates more spatiotemporal dissimilarity than another. However, the analysis does provide clear results in cases of dualist mental properties and their target effects (see section 4). List and Stoljar do not indicate that their strategy requires a different analysis of counterfactuals, and such a deviation would require a separate defense.

to attract my body to the earth. In order to determine what distance a miracle creates between two worlds, Lewis proposes the following guidelines (1979, p. 472):

1. It is of the first importance to avoid big, widespread, diverse violations of law.

2. It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.

3. It is of the third importance to avoid even small, localized, simple violations of law.

4. It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly.

According to these guidelines, the closest possible world where a realized property is excised is typically one where its core realizer is absent as well. Suppose that we are looking for the closest possible world where UD is not instantiated. At least some part of UD's total realizer will have to be absent from that world. After all, its total realizer metaphysically necessitates UD, and metaphysically impossible worlds are standardly ignored when one evaluates counterfactual dependence.[7] Therefore, this world will lack either the core realizer, i.e. PD, or some of the more permanent properties making up the total realizer, like the organizational structure of the university. *Prima facie*, it will be easier to maintain maximum match of particular fact by excising the core realizer, because changing more permanent properties, like organisational structures, is likely to result in more extensive mismatches of particular fact. For instance, changing the organisational structure of the university such that the committees are composed differently, or their decision no longer settles the university decision, is likely to affect *several* university decisions on tenure applications, rather than just this one. All of these different decisions will spread into further differences in matter of fact: new faculty gets hired, lectures are given by different professors, unsuccessful applicants move to other cities, etc. The resulting world will probably be more different from ours than a world where the organizational structure remains identical, but the most successful professors make a different decision in this specific case. Consequently,

---

[7]See Lewis (1973b, sect. 1.6) and Woodward (2008, p. 254–256).

the closest possible world where the realized property is excised typically is a world where the core realizer is absent as well.

Given that counterfactuals are evaluated by looking at the closest possible world where the antecedent is true, this means that realized properties and their core realizers will often enter into the same relations of counterfactual dependency. For example, in the university decision case, an applicant's job loss will be counterfactually dependent on both *UD* and *PD*:

   (i) UD $\Box\!\!\rightarrow$ JOB LOSS

   (ii) $\frown$UD $\Box\!\!\rightarrow$ $\frown$JOB LOSS

   (iii) PD $\Box\!\!\rightarrow$ JOB LOSS

   (iv) $\frown$PD $\Box\!\!\rightarrow$ $\frown$JOB LOSS

Hence, the example not only conforms with our intuitions and the realization literature, it is also supported by the relevant counterfactuals. Unless one is willing to disregard this evidence and maintain that realized properties are causally excluded by their core realizers, the current exclusion argument has no force against the dualist. We can reasonably conclude that (1), (2) and *Metaphysical Exclusion* cannot provide a sound argument against *Dualism*, because *Metaphysical Exclusion* is false.

4   ... but the exclusion worries persist

Remember that naturalist dualists take conscious properties and their underlying physical properties to stand in a nomological necessitation relation (cf. section 1). List and Stoljar maintain that, if these psychophysical necessitation relations are reciprocal, we should expect naturalist dualist conscious properties and their underlying physical properties to co-cause their effects. They take this claim to be supported by the relevant counterfactuals. Concerning cases where two properties F and F* nomologically necessitate one another, they say (2017, p. 104):

> [...] 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*.

However, the considerations which lead us to reject *Metaphysical Exclusion* do not support this claim. In fact, the relevant counterfactuals contradict it. Consider the following naturalist dualist example: my pain is nomologically necessitated by its underlying physical property 'phys'. Phys is in turn nomologically necessitated by my pain. Suppose further that phys is a necessary part of a nomologically sufficient condition for my wincing a moment after phys and my pain are instantiated.

Despite the reciprocal nomological necessitation relation between my pain and phys, these two property instances exhibit relevantly *different* patterns of counterfactual dependence, because the closest possible world where my pain is absent is *not* a world where phys is absent. After all, it takes but a small localized miracle in a psychophysical law to excise my pain and hold all physical facts, including the occurrence of phys, fixed. The resulting possible world will still contain phys *and* my wincing, as it still contains a sufficient physical cause for my wincing. Compare that possible world with the closest possible world where both my pain and phys are absent. We can assume that both property instances can be excised with one small, localized miracle preceding the occurrence of phys, because my pain would not have occurred in the absence of its underlying physical property. Just like the possible world lacking pain, this world only requires one small, localized *miracle.* Even so, the resulting world is further removed from actuality than the world that just lacks my pain, because it contains strictly more mismatch in *particular fact.* In particular, the resulting world lacks *both* my pain *and* phys, rather than just lacking my pain. Furthermore, the absence of phys will *spread* throughout this possible world. For example, given that the occurrence of phys is a necessary part of the sufficient condition for my wincing to occur, I will not wince in the resulting world, which makes for a further mismatch of particular fact. The closest possible world lacking both phys and pain will thus also lack my wincing, but it is *not* the closest possible world where pain is lacking, as there is a closer possible world that lacks pain but contains both phys and my wincing.

By contrast, the closest possible world where phys is absent will *lack* my wincing, because phys is a necessary part of the sufficient condition for my wincing.[8] Consequently, my wincing counterfactually depends on phys, but not on my pain:

(v) PAIN $\square\rightarrow$ WINCE

---

[8] For reasons just addressed, that possible world plausibly lacks my pain as well. However, this is not essential to my argument.

(vi) $\backsim$PAIN $\Box\!\!\rightarrow$ WINCE

(vii) PHYS $\Box\!\!\rightarrow$ WINCE

(viii) $\backsim$PHYS $\Box\!\!\rightarrow$ $\backsim$WINCE

Assuming that counterfactual dependency is required for non-overdeterministic causation, this demonstrates that my pain and phys do *not* co-cause my wincing.

5 New formulation for an old argument

We can summarize our findings as follows. The relation between realized properties and their core realizers is *tighter* than nomological necessitation in that its holding puts stronger restrictions on those nearby possible worlds where the first relatum is instantiated and the second is not. In the case of nomological necessitation, this world is but a small localized miracle away. In the case of realization, the miracle excising the realized property will make for some further mismatch of particular fact in the total realizer, which increases the departure from actuality. The counterfactuals indicate that it is exactly this further tightness that allows the relata of realization to co-cause effects. Given that naturalist dualism does not allow for such a tighter relation between conscious properties and physical properties (cf. Chalmers, 1996, p. 124–129), this means that naturalist dualist conscious properties and physical properties cannot co-cause effects.[9]

We should thus expect there to be an exclusion principle which targets naturalist dualist conscious properties without affecting metaphysically distinct property instances that are tightly related. Here is one proposal for such a principle.

First, call any asymmetric binary relation R 'tight' *iff* for any two property instances x and y, if xRy at w_1, then any world w_2, containing x but not y, is more than a small localized miracle removed from w_1. For example, the relation between a core realizer and its realized property is tight, because a miracle that excises the realized property whilst maintaining the core realizer will have to do

---

[9]Bennett (2008) argues for a similar conclusion by relying on a counterfactual test for genuine overdetermination. Keaton and Polger (2014) use cases of realization to demonstrate that realized properties are often still excluded by their realizers according to Bennett's proposal, which might therefore still be considered to impose too strong requirements on co-causes. My proposal imposes weaker restrictions on co-causes and the resulting exclusion argument thus relies on weaker assumptions.

so by excising other particular facts making up the total realizer. Therefore, a possible world lacking the realized property that contains its core realizer has to be removed from actuality by at least a small, localized miracle *and* some further mismatch in particular fact — a mismatch that, as we have seen, is likely to spread.[10]

Second, let us say that any property instance x 'merely nomologically necessitates' y at w_1 if x nomologically necessitates y and there is no tight relation between x and y at w_1.

We can now formulate the following exclusion principle:

> **Weak Exclusion** If an effect has a sufficient cause at a time t, nothing that is merely nomologically necessitated by this sufficient cause is a cause of the effect at t.

Weak Exclusion is supported by the relevant counterfactuals and is consistent with realization cases. It does *not* affect all varieties of dualism. For example, Kroedel's (2015) *supernomological* dualism requires a separate treatment, as it posits tight psychophysical laws. Nonetheless, the principle *does* affect the standard naturalist dualism that List and Stoljar set out to defend:

> **Naturalist Dualism** Every instance of a conscious property is merely nomologically necessitated by instances of physical properties.

After all, the following set of propositions generates a powerful exclusion argument:

> (1) Some instances of physical properties are caused by the instances of conscious properties.

> (2) Every instance of a physical property has a sufficient physical cause at any given time t (if it has a cause at all at t).

Naturalist Dualism Every instance of a conscious property is merely nomologically necessitated by instances of physical properties.

Weak Exclusion If an effect has a sufficient cause at a time t, nothing that is merely nomologically necessitated by this sufficient cause is a cause of that effect at t.

---

[10]Other relevant relations that will turn out to be tight on this account include: *being grounded in*, *being a determinable of*, *metaphysically supervenes on*, etc.

There might be an elegant reply to this exclusion argument as well, but that reply will require different motivations than those provided by List and Stoljar.

References

Bennett, K. (2008). Exclusion Again. In Hohwy and Kallestrup (2008), pages 280–306.

Block, N. (2003). Do Causal Powers Drain Away? *Philosophy and Phenomenological Research*, 67(1):133–150.

Chalmers, D. J. (1996). *The Conscious Mind*. Oxford University Press, Oxford.

Hasker, W. (2010). Persons and the unity of consciousness. In Koons and Bealer (2010b), pages 175–190.

Hohwy, J. and Kallestrup, J. (2008). *Being Reduced: New Essays on Reduction, Explanation, and Causation*. Oxford University Press.

Jackson, F. (1982). Epiphenomenal qualia. *The Philosophical Quarterly*, 32(127):127–136.

Keaton, D. and Polger, T. W. (2014). Exclusion, still not tracted. *Philosophical Studies*, 171(1):135–148.

Kim, J. (1989). The Myth of Nonreductive Materialism. *Proceedings and Addresses of the American Philosophical Association*, 63(3):31–47.

Kim, J. (2007). Causation and Mental Causation. In McLaughlin, B. P. and Cohen, J. D., editors, *Contemporary Debates in Philosophy of Mind*. Blackwell.

Koons, R. C. and Bealer, G. (2010a). Introduction. In Koons and Bealer (2010b), pages ix–xxxi.

Koons, R. C. and Bealer, G. (2010b). *The Waning of Materialism*. Oxford University Press.

Kroedel, T. (2015). Dualist Mental Causation and the Exclusion Problem. *Noûs*, 2:357–375.

Lewis, D. (1973a). Causation. *The Journal of Philosophy*, 70(17):556–567.

Lewis, D. (1973b). *Counterfactuals.* Oxford University Press, Oxford.

Lewis, D. (1979). Counterfactual dependence and time's arrow. *Noûs*, 13(4):455–476.

List, C. and Stoljar, D. (2017). Does the exclusion argument put any pressure on dualism? *Australasian Journal of Philosophy*, 95(1):96–108.

Loewer, B. M. (2001). From Physics to Physicalism. In Gillett, C. and Loewer, B. M., editors, *Physicalism and its Discontents.* Cambridge University Press.

Ney, A. (2012). The causal contribution of mental events. In Christopher, H. and Simone, G., editors, *New Perspectives on Type Identity: The Mental and the Physical*, pages 230–250. Cambridge University Press.

Papineau, D. (2002). *Thinking about Consciousness.* Oxford University Press, Oxford.

Pautz, A. (2010). A simple view of consciousness. In Koons and Bealer (2010b), pages 25–66.

Robinson, W. S. (1988). *Brains and People: An Essay on Mentality and its Causal Conditions.* Temple University Press.

Shoemaker, S. (2007). *Physical Realization.* Oxford University Press, Oxford.

Stoljar, D. (2008). Distinctions in Distinction. In Hohwy and Kallestrup (2008).

Woodward, J. (2008). Mental Causation and Neural Mechanisms. In Hohwy and Kallestrup (2008), pages 218–262.

Yablo, S. (1992). Mental causation. *Philosophical Review*, 101(2):245–280.