

COUNTERFACTUALS, OVERDETERMINATION AND MENTAL CAUSATION

Simona Aimar

Abstract. The Exclusion Problem (EP) for mental causation suggests that there is a tension between the claim that the mental causes physical effects and the claim that the mental does not overdetermine its physical effects. In response, Karen Bennett (2003, 2008) puts forward an extra necessary condition for overdetermination: if one candidate cause were to occur but the other were not to occur, the effect would still occur. She thus denies one of the assumptions of EP, the assumption that if an effect has two sufficient causes, it is overdetermined. If sound, her argument does two things: it solves EP, and it shows how to use counterfactuals in order to make the notion of overdetermination precise. However, the argument is not sound.

I

The Exclusion Problem. Consider the following claims:

- (1) A mental event is sometimes the sufficient cause of a physical event.¹ (Mental Causation)
- (2) Every physical event has a physical event as its sufficient cause.² (Closure)
- (3) Mental events are distinct from physical events. (Distinction)
- (4) The effects of mental events are not systematically overdetermined. (Non-overdetermination)
- (5) If an event has at least two sufficient causes, and one cause is not an effect of the other cause, then the event is overdetermined. (Exclusion)

As Jaegwon Kim (1998) has elegantly shown, all of (1)-(5) are plausible enough. Taken together, however, they appear to be in tension. Suppose my decision to drink causes

¹ Although I assume that the *relata* of causation are events, nothing I shall say strictly depends on this assumption.

² In this sort of argument, the notion of a sufficient cause is not meant to suggest that a cause brings about the effect all by itself; roughly, the idea is that a cause brings about the effect given that certain background conditions hold (see §ii).

my opening the fridge (from (1)). My opening the fridge is also due to the occurrence of a neural process in my head (from (2)). But this neural process is distinct from my decision to have a drink (from (3)). Now, if this one action of fridge opening is a standard action, it is not overdetermined (from (4)). But then, we have to choose what causes it: is it my decision to drink, or the corresponding neural process? 'Choose whichever, but just one!' (from (5)). Thus the fact that the mental and the physical are both sufficient causes of the same effect seems at odds with the idea that this effect is not overdetermined. This is the Exclusion Problem (EP).

A standard response to EP is to drop the Mental Causation claim, (1) (see Kim 1998, ch. 6). On this view, the physical trumps the mental: its bringing about the relevant effect leaves the mental with nothing to do. However, the idea of abandoning mental causation is at odds with common sense. Thus many wish to solve EP otherwise.

In two interesting papers, Karen Bennett (2003, 2008) contends that in order to solve EP we should reject the Exclusion claim, (5). (5) sets up a necessary and sufficient condition for overdetermination: if an effect has two simultaneous sufficient causes, it is overdetermined. Yet, Bennett claims, there is at least one extra necessary condition for overdetermination:

Bennett's extra condition for overdetermination: An effect e is overdetermined by a cause c_1 and a cause c_2 only if the counterfactuals

$$(O_1) (c_1 \wedge \neg c_2) > e$$

$$(O_2) (\neg c_1 \wedge c_2) > e$$

are both (non-vacuously) true.

Suppose that two gunwomen shoot at the same victim. If the victim dies of two bullets, and her death is overdetermined, both the counterfactuals (O₁) and (O₂) come out (non-vacuously) true. If either gunwoman had not shot but the other had, the victim would still have died. According to Bennett, this suggests that the non-vacuous truth of instances of (O₁) and (O₂) is a necessary condition for over-determination.

Having assumed this, Bennett argues that putative cases of mental causation are counterexamples to (5). At its bare bones, her argument is this:

(6) The non-vacuous truth of instances of (O₁) and (O₂) is a necessary condition for overdetermination.

(7) In mental causation cases, two sufficient causes bring about the same effect, but the relevant instances of (O₁) and (O₂) are not both (non-vacuously) true.

Therefore,

(8) Mental causation cases are counterexamples to (5), the claim that if an event has two sufficient causes, and one cause is not an effect of the other cause, it is overdetermined.

If sound, this argument does two things. First, it shows that there is an alternative to the standard solution to EP. Second, it relates counterfactuals to the notion of overdetermination in a neat and *prima facie* plausible way. Many now accept that Bennett has succeeded in establishing these two things.³ But her argument fails. I contest first premiss (7), and then premiss (6).

II

Against Premiss (7). Premiss (7) of Bennett's main argument says that in putative cases of mental causation at least one instance of (O₁) or (O₂) comes out false. To show this, she considers the following instances of (O₁) and (O₂):

$$(O_1^*) (m \wedge \neg p) > e$$

$$(O_2^*) (\neg m \wedge p) > e,$$

where p is a physical event (presumably a neural process), m is a mental event which supervenes on p (say, my decision to open the fridge), and e is a physical event of which both p and m seem to be sufficient causes (say, my opening the fridge). According to Bennett, (O₂^{*}) is false: it is not the case that if p were to occur but m were not to occur, e would still occur. (O₂^{*}) is an instance of (O₂) for an arbitrary case of mental causation. So by showing that (O₂^{*}) is false Bennett establishes premiss (7) of her overall argument, the claim that in mental causation cases at least one instance of (O₁) or (O₂) comes out false.

But *why* does Bennett think that (O₂^{*}) is false? In order to see this, we first need to consider how she understands the way physical events underlie mental events, and the notion of a sufficient cause. As to the first point, Bennett holds that the occurrence of a physical event p can underlie a mental event m without metaphysically necessitating, by itself, its occurrence. But she thinks that there is a minimal conjunction of physical conditions which hold in the actual world and which, together with p 's occurrence, metaphysically necessitate m 's occurrence.⁴ Moreover, without p 's occurrence, these conditions do not metaphysically necessitate m 's occurrence. These conditions include, Bennett says, 'the laws of nature and other facts about the outside world [i.e. the actual world]' (2003, p. 486).

As to the second point, Bennett understands the notion of a sufficient cause which is in play in EP thus: 'Something can count as causally sufficient for an effect even though it requires some set of background conditions and intermediaries in order to bring about that effect' (2003, p. 490). One way of cashing out this thought is to say that a cause can be sufficient for an effect even if, for it to cause the effect, a certain set of conditions must hold. On this view, the background conditions for a cause to bring about the effect are *necessary* for the cause to bring about the effect. This is how Bennett understands

³ See Eagle (2007), Haug (2010), Carey (2010).

⁴ The conjunction is 'minimal' in the sense that nothing which is strictly weaker than the conjunction does, together with p , necessitate m .

background conditions (2003, p. 490). So in EP a cause is sufficient for an effect even if, for it to cause the effect, a certain set of conditions must hold.

With these assumptions in place, Bennett argues for the falsity of (O₂^{*}) thus:

The conditions that *must hold* for *p* to bring about *e*—physical conditions—are basically the same as the conditions in which *p* necessitates *m*. So if *p* were to occur without *m*, those conditions would not hold—and *p* would not, or at least might not, cause *e*. (Bennett 2003, pp. 488–9)

The argument is stated concisely, but the idea seems to be this:

(9) There is a minimal set of physical conditions together with which *p* metaphysically necessitates *m*.

(10) The conditions that must hold for *p* to cause *e* are the same as the conditions in which *p* metaphysically necessitates *m*.

Hence,

(11) If *p* were to occur without *m*, at least one of the conditions that must hold for *p* to cause *e* would not hold.

Therefore,

(12) If *p* were to occur without *m*, it would not cause *e*.

And since (12) is true, Bennett concludes, (O₂^{*}) is false.

Yet, Bennett's argument for (12) is unsound. I reject premiss (10). (10) corresponds to the first claim in the quote: 'the conditions that must hold for *p* to bring about *e*—physical conditions—are basically the same as the conditions in which *p* necessitates *m*'. One issue with this claim is that it is too loose. Bennett tells us that some conditions must hold for the cause to bring about the effect, but does not point out which sense of 'must' she has in mind. However, I shall not attempt to make Bennett's claim more precise.⁵ Rather, I shall argue that on any plausible reading of (10), (10) is false.

I offer the following counterexample to (10):

PING&PONG: Bill is afraid of dogs. Each time he sees a dog, he thinks that that dog scares him, and then closes his eyes (in order not to see the scary dog). At a moment *t* Bill has two dogs in front of him, Ping and Pong. Ping is positioned right behind Pong, in such a way that Bill cannot see Ping. So at *t* Bill sees (only) Pong and thinks that that dog—i.e. Pong—scares him. At a slightly later moment *t'*, Bill closes his eyes (in order not to see Pong).

⁵ Neither metaphysical nor nomological necessities would do. For we seem to accept that the same cause can (metaphysically or nomologically) bring about the same type of effect under different conditions. So Bennett needs a modal notion which is weaker than nomological necessity.

PING&PONG is a putative case of mental causation. We have a physical effect, Bill's closing his eyes, and two candidate causes, a mental event m and the underlying neural process p . The relevant mental event m is a visual demonstrative thought: Bill's thinking that that dog (Pong) scares him. Given a plausible form of externalism, towards which Bennett herself expresses sympathy (2003, p. 484), in PING&PONG the occurrence of Bill's thought that that dog scares him depends upon the presence of the demonstrandum—Pong. Hence, by (9), Pong's presence is one of the conditions in which p which necessitates Bill's thought. So, given her commitment to (10), Bennett must say that Pong's presence in front of Bill at t is also one of the conditions which must hold for p to cause Bill's closing his eyes at t' .

However, Pong's presence in front of Bill is not one of the conditions which must hold for p to make Bill close his eyes. For consider a world w^* where everything is exactly as it actually is except that at t Pong is not in front of Bill, and so Bill instead sees Ping. In w^* , the neural process that underlies Bill's thought is plausibly the same neural process p that underlies the thought which Bill actually has—the thought that that dog (Pong) scares him. Moreover, at w^* Bill still closes his eyes at t'' . After all, there he sees Ping, and we know that his noticing *any* dog is enough to make him close his eyes. But then, at w^* , p still makes Bill close his eyes. Therefore, (10) is false.

Without (10), there is no sound argument for (12). Without (12), Bennett cannot establish that if p were to occur but m were not to occur, e would not occur. So she has not established that (o2*) is false. But then, we have no reason to accept that in mental causation cases at least one instance of (o1) and (o2) comes out false. That is, we have no reason to accept premiss (7) of her overall argument. Nevertheless, in the next section I will grant Bennett (7), and take issue with the other premiss of her main argument, premiss (6).

III

Against Premiss (6). Premiss (6) of Bennett's main argument says that the non-vacuous truth of instances of the counterfactuals (o1) and (o2)— $(c_1 \wedge \neg c_2) > e$ and $(\neg c_1 \wedge c_2) > e$, respectively—is a necessary condition for overdetermination. I will now show that this is false.

Let me first say something more about overdetermination. Bennett takes the following cases to be paradigm cases of overdetermination: firing squads simultaneously shooting at the same victim; two children simultaneously throwing a rock at a window; your alarm clock sounding at just the same moment as a jackhammer starts up outside your window (Bennett 2008, p. 287). All these cases, she suggests, involve the following two features: they are such that (a) the same effect has (at least) two distinct causes; and (b) 'there are two completely distinct transfers of energy involved' (2008, p. 287).⁶ Bennett also seems to accept that meeting (a) and (b) is a necessary and sufficient condition for

⁶ For an account of the distinction between 'transfer of energy' views of causation and counterfactual views, see Hall (2004).

overdetermination.⁷

Now consider the following case:

GLASS. There are two gunwomen and a victim. If gunwoman-1 does not shoot, a device detects this and raises a bullet-proof glass in front of the victim. Otherwise, the device does nothing. Both gunwomen shoot simultaneously. The device does nothing. Two bullets get to the victim. The victim dies.

GLASS meets condition (a), having two distinct causes. Here the putative causes are the shootings of two gunwomen—two distinct events. GLASS also meets condition (b), involving two distinct transfers of energy. The victim dies from two bullets which the two gunwomen shoot independently of each other—similarly to Bennett’s firing squads case. And since GLASS meets conditions (a) and (b), by Bennett’s criteria, it is a case of overdetermination.

However, according to Bennett’s extra condition for overdetermination, GLASS is *not* a case of overdetermination. Bennett’s extra condition for overdetermination, remember, says that we have overdetermination only if the relevant instances of (o1) and (o2) are non-vacuously true. Now consider the following instances of (o1) and (o2):

(G1) If gunwoman-1 were to shoot and gunwoman-2 were not to shoot, the victim would die;

(G2) If gunwoman-2 were to shoot and gunwoman-1 were not to shoot, the victim would die.

Given that GLASS is a case of overdetermination, if Bennett’s extra condition for overdetermination were sound, then both (G1) and (G2) *should* come out (non-vacuously) true. All is fine with (G1): if gunwoman-1 shoots but gunwoman-2 does not shoot, the device does nothing; a bullet gets to the victim, and the victim dies. However, (G2) comes out false: if gunwoman-2 were to shoot and gunwoman-1 were *not* to shoot, the bullet-proof glass would step in and save the victim. Hence, GLASS does *not* meet Bennett’s extra condition for overdetermination. But as we have seen, GLASS *is* a case of overdetermination. So Bennett’s extra condition for overdetermination is to be rejected. But then, (6) is false.

IV

Conclusion. Bennett’s argument against the Exclusion claim fails. First, Bennett gives us no good reason to accept that mental causation cases fail to meet her extra condition

⁷ A note on exegesis. I understand (a) and (b) as sufficient conditions for overdetermination because Bennett appears to accept that the paradigmatic cases of overdetermination qualify as overdetermination cases by virtue of meeting conditions (a) and (b) (Bennett 2008, p. 287). Kim (2005, 2007) seems to do the same.

for overdetermination. Second, her extra condition for overdetermination is open to counterexamples. So Bennett's appeal to counterfactuals does not help us to make the notion of overdetermination precise, nor to solve EP. There may be alternative ways of using counterfactuals in order to achieve such goals. But the failure of Bennett's attempt suggests that this is not an easy path to walk.⁸

References

Bennett, Karen 2003: 'Why the Exclusion Problem Seems Intractable, and How, Just Maybe, To Tract It'. *Noûs*, 37(3), pp. 471–97.

———2008: 'Exclusion Again'. In Jesper Kallestrup and Jakob Hohwy (eds.), *Being Reduced: New Essays on Reductive Explanation and Special Science Causation*, pp. 280–305. Oxford: Oxford University Press.

Carey, Brandon 2010: 'Overdetermination and The Exclusion Problem'. *Australasian Journal of Philosophy*, pp. 251–62.

Eagle, Antony 2007: 'Pragmatic Causation'. In Huw Price and Richard Corry (eds.), *Causation, Physics and the Constitution of Reality: Russell's Republic Revisited*, pp. 156–90. Oxford: Oxford University Press.

Hall, Ned 2004: 'Two Concepts of Causation'. In John Collins, Ned Hall and L. A. Paul (eds.), *Causation and Counterfactuals*, pp. 225–65. Cambridge, ma: mit Press.

Haug, Matthew C. 2010: 'The Exclusion Problem Meets the Problem of Many Causes'. *Erkenntnis*, 73(1), pp. 55–65.

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

———2005: *Physicalism, or Something Near Enough*. Princeton, nj: Princeton University Press.

———2007: 'Causation and Mental Causation'. In B. P. McLaughlin and J. Cohen (eds.), *Contemporary Debates in Philosophy of Mind*, pp. 227– 42. Oxford: Blackwell.

⁸ For conversations about various drafts of this paper, I'd like to thank Arif Ahmed, Daniel Morgan, John Hawthorne, Barbara Vetter and Lee Walters. Thanks also to Keble College, the Carnegie Trust and the Fondazione Casa Rurale di Trento for their financial support.