CANADIAN JOURNAL OF PHILOSOPHY, 2017
VOL. 47, NO. 1, 20–42
http://dx.doi.org/10.1080/00455091.2016.1218254

Routledge
Taylor & Francis Group

Mental causation, compatibilism and counterfactuals

Dwayne Moore

Philosophy Department, University of Saskatchewan, Saskatoon, Canada

ABSTRACT

According to proponents of the causal exclusion problem, there cannot be a sufficient physical cause and a distinct mental cause of the same piece of behaviour. Increasingly, the causal exclusion problem is circumvented via this compatibilist reasoning: a sufficient physical cause of the behavioural effect necessitates the mental cause of the behavioural effect, so the effect has a sufficient physical cause and a mental cause as well. In this paper, I argue that this compatibilist reply fails to resolve the causal exclusion problem.

ARTICLE HISTORY Received 10 November 2015; Accepted 26 July 2016

KEYWORDS Mental causation; nonreductive physicalism; compatibilism; causal exclusion; counterfactuals

According to proponents of the causal exclusion problem, there cannot be a sufficient physical cause and a distinct mental cause of the same piece of behaviour. This causal exclusion problem is largely motivated by the causal exclusion principle, which is a principle stipulating that 'no single event can have more than one sufficient cause occurring at any given time …' (Kim 2005, 42). This causal exclusion principle is in turn substantially supported by the necessity argument: if an effect already has a sufficient cause, additional causes of the effect are unnecessary (Kim 1998, 44–45), hence excludable. Increasingly, the necessity argument is circumvented via this compatibilist reasoning: a sufficient physical cause of the behavioural effect necessitates the mental cause of the behavioural effect, and vice versa, so the effect necessarily has a sufficient physical cause and a mental cause as well (Bennett 2003; Kallestrup 2006, 473; Marras and Yli-Vakuri 2008, 125; Walter 2008, 678; Carey 2011, 258–259; Arnadottir and Crane 2013, 255; Kroedel 2015b). In this paper, I argue that this compatibilist reply insufficiently circumvents the necessity argument, hence fails to completely undermine the causal exclusion principle, and fails to solve the causal exclusion problem.

CONTACT Dwayne Moore dwayne.moore@usask.ca

© 2016 Canadian Journal of Philosophy

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

This paper is divided into five sections. After briefly outlining the Causal Exclusion Problem (§1), I demonstrate how the causal exclusion problem is motivated by the Causal Exclusion Principle, which is substantially undergirded by the Necessity Argument (§2). Then, after defining Compatibilism, and narrowing my scope to that subset of compatibilists that operate within a counterfactual theory of causation (§3), I outline the counterfactualist compatibilist reply to the necessity argument (§4). I then introduce significant difficulties that counterfactualist compatibilism faces (§5).

1. The causal exclusion problem

According to a common though not universal presentation, the causal exclusion problem is the conjunction of the following four individually plausible, but (seemingly) jointly inconsistent principles:

(1) *The principle of mental causation*: some physical effects have mental causes.[1]
(2) *The principle of irreducibility*: mental causes are distinct from physical causes.[2]
(3) *The principle of physical causal completeness*: 'if a physical event has a cause at *t*, it has a sufficient physical cause at *t*' (Kim 2009, 38).
(4) *The principle of causal exclusion*: 'no single event can have more than one sufficient cause occurring at any given time …' (Kim 2005, 42).[3]

While each of these principles is individually plausible, they seem to form an inconsistent tetrad. In brief, how can behavioural effects have no more than a single sufficient physical cause, while they must simultaneously possess distinct mental causes as well?

There are those who resolve the causal exclusion problem by rejecting one of the principles constituting the causal exclusion problem. The reductionist argues that the evidence in favour of (1) mental causation; (3) physical causal completeness; and (4) causal exclusion is conclusive. So, (2), irreducibility must be rejected (Kim 2005, 125; Ney 2007). Notoriously, however, mental causes are multiply realizable and intuitively distinct from physical causes, so many find the principle of irreducibility 'obviously true' (Bogardus, 2013, 446). The epiphenomenalist claims that the overwhelming evidence in favour of (2) irreducibility; (3) physical causal completeness; and (4) causal exclusion renders (2), mental causation, false (Robinson 2004, 158; Lyons 2006). But, mental causation is pre-theoretically intuitive, and serves as a foundation for moral responsibility and epistemic justification, so many think that abandoning mental causation amounts to 'the end of the world' (Fodor 1989, 77). The substance dualist contends that (1) mental causation; (2) irreducibility; and (4) causal exclusion are true, so (1), physical causal completeness, can be abandoned (Foster 1989; Meixner 2008). However, due to twentieth-century scientific advances,

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

most think the principle of physical causal completeness is 'fully established' (Papineau 2001, 33). There is also an expanding tribe of scholars, increasingly called the 'compatibilists', who embrace (1) mental causation; (2) irreducibility; and (3) physical causal completeness, thereby rejecting (4) causal exclusion in some way (Bennett 2003; Sider 2003). Pre-theoretically though, presuming that behavioural effects, on a global scale, have more than a single sufficient cause is 'a bit bizarre' (cp. Schiffer 1987, 148; Harbecke 2008, 28; Kim 2009, 45) or even 'absurd' (Kim 1993, 281).

2. Motivating the causal exclusion principle

While each of these positions deserve (and receives) its due attention, it is the latter compatibilist position that is of present concern. In order to properly evaluate the compatibilist position, it is helpful to first marshal together the arguments often provided in favour of the principle they reject, namely, the causal exclusion principle.

The literature contains at least five arguments in favour of the causal exclusion principle, though I will only expand upon the last one. First, there is an appeal to intuition: Jaegwon Kim claims that the causal exclusion principle is 'virtually an analytic truth' (Kim 2005, 51), or, is *prima facie* obvious. This is less an argument than an assertion that the causal exclusion principle carries intuitive force, as some argue that the principles of mental causation and/or irreducibility are supported by brute intuitive strength. Second, the parsimony argument: as a general scientific value, we should 'get by with the fewest possible entities' (Kim 1989, 98), so if one cause is sufficient, we should get by without positing additional causes (cp. Malcolm 1968; Goldman 1969; Kim 1989, 98), so we should exclude additional causes. Third, and most commonly, the coincidence argument: overdetermination is a rare coincidence—barns infrequently burn down by the simultaneous occurrence of a short circuit and a dropped match. Mental causation, however, is ubiquitous. Overdetermination, if true of mental causation, would be ubiquitous and these massive amounts of coincidence are unacceptable (cp. Schiffer 1987, 148; Kim 1998, 53; Kim 2006, 558; Lim 2013, 670–671; Roche 2014, 817–818; Kroedel 2015a, 367). Fourth, the additivity argument: if causation involves energy/momentum transfer, and one sufficient cause transfers all the requisite energy/momentum to the effect, then a second cause would either transfer surplus energy/momentum to the effect (pushing the effect twice as hard and/or far) or be incapable of transferring energy/momentum to an effect on account of the fact that it already possesses all its requisite energy/momentum (Kim 1998, 53–55; Sider 2003, 3–4; Carey 2011, 253–254; Audi 2013, 460; Paul and Hall 2013, 147).

The fifth argument, the necessity argument, is perhaps the most problematic argument of all. The necessity argument is especially problematic because it, unlike other arguments, indicates that abandoning the causal exclusion principle

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

ultimately leads to the failure of one or both of the principles of physical causal completeness or mental causation. Unfortunately, the necessity argument is underdeveloped in the literature, so I will presently attempt to shine some light on the necessity argument. Consider the following passage as an initial articulation of the necessity argument:

> If this physiological event is indeed sufficient for the climbing, the climbing should occur whether or not any other event (such as beliefs and desires) occurred. That is, no other event should be necessary for the occurrence of the climbing … If [*p*] is sufficient for a later event [*p**], then no event [*m*] occurring at the same time as [*p*] and wholly distinct from it is necessary for [*p**]. (Kim 1989, 82; cp. Kim 1998, 44–45; Sider 2003; Carey 2011, 254; Engelhardt 2015, 207)

According to Kim, if the physical cause *p* is a sufficient cause, then the physical cause *p* suffices to, or is all that one needs to, ensure that *p** will occur all by itself. So, the distinct event *m* is not necessary for *p** to occur. Rather, *m* is 'otiose' (Kim 2006, 558), or 'superfluous' (Corry 2013, 33). Or, it suffers from 'redundancy' (Harbecke, 2013, 65) and 'dispensability' (Lim 2013, 670), and hence can be excluded.

The necessity argument presupposes that causal necessity and causal sufficiency stand in the following relations: if *a* alone is a sufficient cause for *b*, *c* is unnecessary as a cause for *b*; and, if *a* is necessary as a cause of *b*, *c* alone is an insufficient cause of *b*. This proposed relation between sufficient causes and necessary causes is controversial, so it will be defined, defended and nuanced below. But to begin, here is a classic expression of this intuition from Myles Brand:

> If there are two sets of distinct conditions, each necessary for *b*, then neither set alone is sufficient for *b*, for if only one of these sets occurs, then *b* will not occur, since there will be conditions required for *b*'s occurrence that did not take place. Similarly, if there are two sets of distinct conditions, each sufficient for *b*, then neither set is necessary for *b*, for it is enough for one of these two sets of conditions to occur in order for *b* to occur; neither set is such that it is required that it occur in order for *b* to occur'. (cp. Brand and Swain 1974, 358–359; Brand 1976, 41; Marras 2007, 319; Turner 2008, 262; Moore 2012, 322)

According to Brand, if we say that *b only needs* conditions *a*, then we cannot simultaneously say that *b* also needs conditions *c* (i.e. that *b does not only need* conditions *a*). And, if we say that *b needs* conditions *a*, then we cannot also say that *b* only needs conditions *c* (i.e. that *b* does *not need* conditions *a*).

This inaugural articulation of the necessity argument requires the following clarifications and refinements. First, the proposed relationship that exists between *a*'s sufficiency for *b* and *c*'s unnecessity for *b* is not a relationship of logical necessity and logical sufficiency. It is possible for *a* to be logically sufficient for *b*, while *c* is necessary for *b*. To slightly modify an example from David Sanford, *a* is the poet looking at a bird in a prime number of ways that is more than 12 but less than 14. From this it is logically necessary that *c*, the poet looks at a bird in a prime number of ways that is more than 11 but less than 15. At the

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

same time, *a* is logically sufficient for the fact that *b*, the poet looks at the bird in 13 ways and *c* is logically necessary for the fact that *b* (Sanford 1975, 109; cp. Engelhardt 2012, 237). Nor is the proposed relationship that exists between *a*'s sufficiency for *b* and *c*'s unnecessity for *b* a relationship of necessary conditions and sufficient conditions. Imagine that you shoot a gun, necessitating a 'bang' sound, but the gunshot is sufficient for some murder, while the sound is a necessary condition for the murder as well.

Rather, the problematic relationship that exists between *a*'s sufficiency for *b* and *c*'s unnecessity for *b*, is a relationship of causal necessity and causal sufficiency. It is not possible for *a* to be a sufficient cause of *b*, while at the same time maintaining that *c* is necessary as a cause (that is, as a partial cause, as a sufficient cause or as any other type of cause) of *b*. To return to the prior example, the gun blast is a sufficient cause of the murder, and though the sound coming from the gun blast is a necessary condition for the murder, it is not necessary that the sound be a cause of the murder. In fact, intuitively, it is not. Here is a perhaps unfortunate way of making the point: assume that *a* causes *b*, and that *b* requires 100 causal units in order for it to occur. If *a* delivers 100 causal units to *b*, then no causal unit contribution is needed from *c*. If *a* necessarily gives rise to another condition *c*, then *c* may be necessarily present for *b*, but this does not change the fact that *b* does not need any causal contribution from *c*. Further elaboration elsewhere clarifies that Brand has this in mind as well (cp. Brand and Swain 1974).

Before noting the problems that the necessity argument poses to the mental causation debate, it is worth pointing out that the principle of physical causal completeness indicates that some physical event *p* is a sufficient cause of the behavioural effect *p**. This follows from the fact that *p** 'Has a sufficient physical cause' (Kim 2009, 38), and it captures the fact that the physical world is causally complete in itself. As Kim says, 'there is no need to go outside the physical domain to find a cause … of a physical event' (Kim 2005, 16).

Not only does the principle of physical causal completeness indicate that some *p* is a sufficient cause of *p**, but it also shows that some *p* is necessary as a cause of *p**. This follows from the above definition of physical causal completeness as well. The fact that *p** 'has a sufficient physical cause' entails that *p** has a physical cause. Some even define the principle of physical causal completeness without reference to the causal sufficiency of the physical cause, to emphasize the necessity of some physical cause: 'if a physical event has a cause at *t*, then it has a physical cause at *t*' (Kim 2005, 15; cp. Kroedel and Schultz 2016, 8). Imagine, for example, that Joe's brain is connected to an exceptionally powerful fMRI machine that detects all brain activity. Joe snaps his finger. Might the fMRI machine detect that no brain activity occurred? Might some disembodied soul have inaugurated the finger snapping without any physical cause at all? It is common for those that endorse physical causal completeness to answer no to these types of possibilities (cp. Crisp and Warfield 2001, 314; Kim 2005, 46–50).

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

Presumably, this is because physical causal completeness requires behavioural effects to have some physical cause.[4]

The principle of mental causation indicates that some mental event *m* is necessary as a cause of *p**. This follows from the fact that the principle of mental causation says some physical effect *p** has a mental cause *m*. It is possible for physical effects to lack mental causes—the tree falling causes the dirt to scatter, the melting ice causes the river to flood, etc. But, mental causation clearly does not occur in these cases. It is also possible for physical effects to be preceded by necessarily occurring mental events—*m* as an epiphenomenal shadow accompanies some sufficient cause *p* of *p**. But, mental causation clearly does not occur in these cases either. But, if the principle of mental causation is true on an occasion, some *m* is necessarily a cause of *p**. If *m* is merely a necessarily present event prior to *p**, or worse yet, does not occur at all, then mental causation fails in this instance. So, in order for mental causation to be true in this instance, some *m* is necessary as a cause of *p**.

Some argue that, in the interests of parity, the principle of mental causation also indicates that *m* is a sufficient cause of *p** as well (Crane 1995, 231; Bennett 2003, 481). This does not follow from the definition of mental causation given above, and I do not endorse or require this stronger principle of mental causation—in fact, I argue against it. But some others do, so it will be discussed throughout the paper. These applications of the definitions of necessary causation and sufficient causation to the principles of physical causal completeness and mental causation are admittedly brief. Once the counterfactual model of causation is in place (§3), I shall bolster these definitions with their appropriate counterfactual tests.

Having set the stage, I now turn to four problems that the necessity argument raises in the mental causation debate. First: if *p* is a sufficient cause of *p**, *m* is unnecessary as a cause for *p**—if *p* is all the cause needed for *p**, some *m* cannot also be needed as a cause of *p**. But, by the principle of mental causation, some *m* is necessary as a cause for *p**, or this is not a case of mental causation. So, if the principle of mental causation is true in this instance, some *m* is necessary as a cause of *p**, but the necessity argument indicates that it is not, so the principle of mental causation is false. This first version of the necessity argument appears in the literature on several occasions (cp. Kim 1989, 82; Kim 1998, 44–45; Kim 2005, 48–49; Moore 2012, 323; Corry 2013, 33).

The second problem merely inverts the first: according to the principle of mental causation, in cases of mental causation, some *m* is necessary as a cause of *p**. If some *m* is necessary as a cause of *p** in cases of mental causation, then *p* is not a sufficient cause of *p** on those occasions where *m* is necessary as a cause as well, which violates the principle of physical causal completeness. For example, Sam steals a car, and, to secure the principle of mental causation, his decision is necessarily a cause of this effect. So, the physical cause is not all that is needed to cause this theft. This second problem makes several appearances

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

in the literature (Bennett 2003, 481; Moore 2012, 328). As noted, problems one and two serve as opposite sides of the same coin. Problem one indicates that the causal sufficiency of *p* for *p** (from physical causal completeness) renders *m* unnecessary as a cause of *p** (falsifying mental causation), while problem two indicates that the requirement that *m* be a cause of *p** (from mental causation) renders it false that *p* is all the cause needed for *p** (falsifying physical causal completeness).

Problem three: according to some, *m* is a sufficient cause of *p**, so some *p* is unnecessary as a cause of *p**. But, by the principle of physical causal completeness, some *p* is necessary as a cause of *p**. So, in order for the principle of physical causal completeness to be true, some *p* is necessary as a cause of *p**, but the necessity argument shows that it is not, so physical causal completeness is false. For example, Judy is a disembodied soul that decides to make turkey for supper, so she does. But, physical causal completeness indicates that some physical cause is necessary for this behaviour—she cannot make turkey without some accompanying neural stimulus. Yet, as it turns out, her disembodied decision is sufficient for her behaviour, so she could make turkey even without a neural cause—another result that is hard to swallow. This third problem is invoked on several occasions as well (Kim 1998, 44–45; Kim 2005, 48–49; Moore 2012, 323; Roche 2014).

Problem four merely inverts this issue: the principle of physical causal completeness indicates that some *p* is necessary as a cause of *p**, which renders *m* an insufficient cause of *p**. That is, since the principle of physical causal completeness indicates that some physical event must cause *p**, it is not the case that *m* alone and by itself, as a disembodied soul, can cause *p**. So, *m* is not a sufficient cause of *p**. This conclusion poses problems for those, such as Karen Bennett, who maintain that the principle of mental causation requires *m* to be a sufficient cause of *p**. Bennett gestures at, but ultimately rejects, this final version of the problem (Bennett 2003, 481). And, as noted, problems three and four serve as opposite sides of the same coin as well. Problem three indicates that the causal sufficiency of *m* for *p** (possibly from mental causation) renders *p* unnecessary as a cause of *p** (falsifying physical causal completeness), while problem four indicates that the requirement that some *p* be a cause of *p** (from physical causal completeness) renders it false that *m* is all the cause needed for *p** (possibly falsifying mental causation).

3. Counterfactualist compatibilism

Compatibilism, a doctrine coined by Terence Horgan (1997, 166), and labelled a movement by Bennett (2003, 473) suggests that physical effects have sufficient physical causes and distinct mental causes. To secure compatibilism, one must reject or substantially nuance the principle of causal exclusion. The principle of causal exclusion is, however, supported by, among other things, the necessity

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

argument (§2). So, to secure compatibilism, one must reject or nuance the causal exclusion principle in a manner that responds to, among other things, the necessity argument.

Compatibilists typically overcome the necessity argument by positing an especially tight relation between the physical cause and mental cause of the effect (cp. Yablo 1992; Pereboom 2002; Bennett 2003; Shoemaker 2007), where this tight relation ensures that the mental cause necessarily accompanies a physical cause of the effect, and vice versa. Therefore, contra the necessity argument, the mental cause and physical cause of the effect necessarily occur after all. In the next two sections, I discuss how this strategy plays out for one particular subset of compatibilists, namely, counterfactualist compatibilists.

Counterfactualist compatibilists (hereafter simply called 'compatibilists') endorse compatibilism via emphasis on a counterfactual model of causation (Loewer 2002, 658–660; Bennett 2003; Kallestrup 2006, 473; Marras and Yli-Vakuri 2008, 125; Walter 2008, 678; Carey 2011, 258–259; Arnadottir and Crane 2013, 255; Roche 2014; Kroedel 2015b). Thus, to understand their position, it is important to highlight the relevant details of the counterfactual model of causation. The counterfactual analysis of causation, while rooted in a brief remark from David Hume, is popularized by David Lewis (1986 and 1973a). For Lewis, for an event p to cause event p^* there is (a chain of) causal dependence from p to p^*, where the (chain of) causal dependence from p to p^* obtains when there is (a chain of) counterfactual dependence from p to p^* (Lewis 1986, 166–167; Lewis, 1973a, 563). Counterfactual dependence occurs when the following two statements are true:

(I) Had p occurred, p^* would have occurred: $p \mathrel{\square\!\!\rightarrow} p^*$.
(II) Had p not occurred, p^* would not have occurred: $\sim p \mathrel{\square\!\!\rightarrow} \sim p^*$.

For Lewis, (I) is automatically true, since p and p^* are actual events. So, establishing the truth/falsity of (II) is paramount. The truth/falsity of counterfactuals is established by appealing to Lewis' possible worlds semantics, according to which counterfactuals are nonvacuously true when the closest possible world where the antecedent and the consequent are both true is closer than any possible world where the antecedent is true but the consequent is false. Counterfactuals are false when the closest possible world where the antecedent and the consequent are both true is further from any possible world where the antecedent is true but the consequent is false. Counterfactuals are vacuously true when the antecedent is false in all possible worlds. For example, for the conditional 'Had the square been round, then the Yankees would have won the world series', the antecedent is metaphysically impossible, so it is false in all possible worlds, so it is impossible to locate the requisite worlds where the antecedent is true while the consequent is true or false. So, rather than being strictly true or strictly false, they are considered vacuously true. For ease of discussion, I shall refer to the former types of counterfactuals as true (rather than as non-vacuously true), the

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

second types of counterfactuals as false and the latter types of counterfactuals as vacuous (rather than as vacuously true).

The truth/falsity of counterfactuals, then, requires a mechanism for discerning possible world proximity. Here it is: comparative closeness amongst possible worlds relies on the assumption that possible worlds are weakly ordered with respect to overall similarity of particular fact and natural law. According to Lewis, the furthest worlds are the worlds with large-scale violations of natural law. Closer worlds are worlds without large-scale violations of natural law, but worlds with significantly differing particular fact across time and space. Closer still are the worlds with a small-scale violation of law gives rise to minute distinctions of particular fact. The closest world is the actual world, which has no variation in particular fact or natural law (Lewis 1979, 472).

Within this counterfactual model of causation, the necessity argument reveals itself, among other places, in cases of overdetermination. In cases of overdetermination, more than a single sufficient cause brings about the same effect at the same time. Since either event is a sufficient cause of the effect, the other event is unnecessary as a cause of the effect. To use a classic example, two assassins shoot Smith at the same time, so Smith's death is overdetermined by bullet *a* firing and bullet *b* firing. The following counterfactual test establishes this fact:

(III) Had bullet *a* fired without bullet *b* firing, the death *c* would have occurred: $(a \,\&\, \sim b) \,\Box\!\!\rightarrow c$.
(IV) Had bullet *b* fired without bullet *a* firing, the death *c* would have occurred: $(b \,\&\, \sim a) \,\Box\!\!\rightarrow c$.

If both of these counterfactuals are true, Smith's death is overdetermined by two sufficient causes, indicating that each bullet firing is unnecessary as a cause of Smith's death. That is, the bullet *a* firing suffices to cause Smith's death because we can remove bullet *b* firing, leaving only bullet *a* firing, and Smith's death still occurs. Similarly, the bullet *b* firing suffices to cause Smith's death because we can remove the bullet *a* firing, leaving only the bullet *b* firing, and Smith's death still occurs. Likewise, the bullet *a* firing is not needed to cause Smith's death because we can remove the bullet *a* firing and Smith's death still occurs. Nor is the bullet *b* firing needed as a cause of Smith's death because we can remove the bullet *b* firing and Smith's death still occurs.

Similarly, on the counterfactual account, the necessity argument, applied to the case of mental causation, takes the following form (cp. Mills 1996, 107; Kim 1998, 44–45; Bennett 2003, 480; Kim 2005, 46–49; Carey 2011, 257; Lim 2013, 672; Kroedel 2015a, 363):

(V) Had *p* without *m* occurred, *p** would have occurred: $(p \,\&\, \sim m) \,\Box\!\!\rightarrow p^*$.
(VI) Had *m* without *p* occurred, *p** would have occurred $(m \,\&\, \sim p) \,\Box\!\!\rightarrow p^*$.

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

If both of these counterfactuals are true, then the behavioural effect *p** is overdetermined, since the individual sufficiency of each cause renders the other cause individually unnecessary.

The pieces are now in place to add counterfactual tests to the application of the concepts of necessary causation and sufficient causation to the principles of mental causation and physical causal completeness. As stated above, if mental causation is true in this instance, *m* is necessary as a cause of *p**. The event *m* is necessary as a cause of *p** if counterfactual (V) is false—if we take only *m* away and *p** does not occur, then clearly *m* is needed for *p** to occur. The event *m* is unnecessary as a cause of *p** if counterfactual (V) is true—if we can wipe only *m* off the face of the universe and *p** still occurs, then clearly *m* is not needed to bring about *p**.[5] So, the principle of mental causation is false if counterfactual (V) is true.

Here is an objection: some compatibilists believe that counterfactual dependency is sufficient for causation. Counterfactual dependency is established via the truth of counterfactuals such as (II). So, if $\sim m \; \Box\!\!\rightarrow \sim p^*$ is true, then mental causation is established, regardless of whether counterfactual (V) is true as well (Kroedel 2015a, 359–361; Kroedel 2015b, 842–844). While counterfactual dependency may be a necessary condition for causation, it cannot be a sufficient condition for causation. After all, a long-standing criticism of the counterfactual account of causation is that it is possible to establish counterfactual dependency among causally unrelated events. For example, the counterfactual 'Had the gunshot sound not occurred, the death would not have occurred' is true, though the gunshot sound is not a cause of the death. Similarly, many worry that $\sim m \; \Box\!\!\rightarrow \sim p^*$ may be true, while it is also true that *m* is epiphenomal (Kim 1998, 45; Esfeld 2010, 101–102; Roche 2014). The way to overcome this problem is to insist that the non-truth of counterfactual (V) is also a necessary condition for the truth of mental causation.[6]

Similarly, as mentioned above, physical causal completeness requires that some *p* is necessary as a cause of *p**. Some *p* is necessary as a cause of *p** in this instance if counterfactual (VI) is false—if we wipe only any physical events away, and *p** does not occur, then clearly some *p* is needed for *p** to occur. Some *p* is unnecessary as a cause of *p** if counterfactual (VI) is true—if we can wipe only the physical events off the face of the universe and *p** still occurs, then clearly some *p* is not needed to bring about *p**. So, physical causal completeness is false if counterfactual (VI) is true.

Physical causal completeness also required that *p* is a sufficient cause of *p**. The event *p* is a sufficient cause of *p** if, among other things, counterfactual (V) is true—if we keep only *p*, and *p** still occurs, then clearly *p* is all that was need to cause *p**. The event *p* is insufficient as a cause of *p** if counterfactual (V) is false—if we keep only *p*, and *p** does not occur anymore, then clearly *p* was not all that was needed to cause *p**. So, physical causal completeness is false if counterfactual (V) is false. Likewise, some said the principle of mental

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

causation requires that *m* is a sufficient cause of *p**. The event *m* is a sufficient cause of *p** if, among other things, counterfactual (VI) is true—if we keep only *m*, and *p** still occurs, then clearly *m* is all that is needed to cause *p**. The event *m* is insufficient as a cause of *p** if counterfactual (VI) is false—if we keep only *m*, and *p** does not occur anymore, then clearly *m* is not all that is needed to cause *p**. So, mental causation may be false if counterfactual (VI) is false.

Here is an objection: this argumentation may rely upon a definition of sufficient causation that is questionable, and perhaps even unorthodox. To be sure, the definition of sufficient causation is variously articulated. As Elizabeth Anscombe indicates: 'Now "sufficient condition" is a term of art whose users may therefore lay down its meaning as they please' (Anscombe 1971, 90). At least four definitions of sufficient causation are present in the relevant literature, though I shall only focus on two. First: *p* is a sufficient cause of *p** if *p* is 'enough' (Anscombe 1971, 91) for *p**, so *p* 'alone and by itself' (Campbell 2004, 153) causes *p**, which 'implies that no other condition is necessary' (Marras 2007, 319) for *p**. Call this individual sufficiency. According to the counterfactual analysis, *p* is individually sufficient for *p** if counterfactual (V) is true. In the argumentation above, I assume that individual sufficiency is a necessary condition for sufficient causation.

But, perhaps individual sufficiency is not a necessary condition for sufficient causation because some other type of sufficient causation is a sufficient condition for sufficient causation. Here is a second definition of sufficient causation that may be preferable: *p* is a sufficient cause of *p** if the material conditional $p \rightarrow p^*$ is true. Call this nomological sufficiency. It is notoriously difficult to analyse nomological sufficiency in counterfactual terms. Most naturally, it seems that *p* is nomologically sufficient for *p** when the counterfactual $p \mathrel{\square\!\!\rightarrow} p^*$ is true. But, as noted above, this test fails, since *p* and *p** are stipulated as actual events, so the counterfactual is automatically true, even of unrelated events (Ruben 1981, 40). Here is another option: *p* is nomologically sufficient for *p** when the logically equivalent counterfactual $\sim p^* \mathrel{\square\!\!\rightarrow} \sim p$ is true (Marc-Wogau 1962, 221–222; Tranøy 1962, 241). This unfortunately renders the causal relation symmetric (Ruben 1981, 40). How about: *p* is nomologically sufficient for *p** when the counterfactual $\sim p \mathrel{\square\!\!\rightarrow} (p \mathrel{\square\!\!\rightarrow} p^*)$ is true (Ruben 1981, 240; Rasmussen 1982, 209; Mills 1996, 106–107). This test solves both the 'automatically true' issue and the 'causal symmetry' issue, though it is not without its detractors (Kroedel 2015a, 373). Thomas Kroedel argues that *p* is nomologically sufficient for *p** if the material conditional $p \rightarrow p^*$ is true in a suitable range of possible worlds (Kroedel 2015a, 366). Relatedly, Jeff Engelhardt argues that *p* is nomologically sufficient for *p** if $p \mathrel{\square\!\!\rightarrow} p^*$ is true in this world and relevant nearby possible worlds (Engelhardt 2012, 237). Either of these last two definitions can serve as a suitable definition of nomological sufficiency.

Perhaps this nomological sufficiency is a sufficient condition for sufficient causation. In other words, perhaps *p* is sufficient for *p** if $p \mathrel{\square\!\!\rightarrow} p^*$ is true in a

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

suitable range of possible worlds, and it is not necessary for counterfactual (V) to be true.[7] As Kroedel remarks, 'Thus, the falsity of [VI] may well be compatible with *m*'s being a sufficient cause of [*p**]' (Kroedel 2015a, 366). Or, Engelhardt seems to share a similar general view: '*A* does not need to suffice for *K* in isolation, however, in order to suffice for *K*. Rather, it need only be the case that $A \mathrel{\square\!\!\rightarrow} K$ holds in actuality and relevant nearby possible worlds, as is the case' (Engelhardt 2012, 237).

There are reasons to believe, however, that nomological sufficiency, while an important condition on sufficient causation, is not a sufficient condition for sufficient causation. First, nomological sufficiency is a relic of the regularity theory of causation, where the counterfactual account of causation is presently presumed, so it is unclear that nomological sufficiency can be the appropriated for use as sufficient causation in this context. But, perhaps it is possible to gerrymander nomological sufficiency into the counterfactual analysis? Perhaps. I tried to do so above, but the task is difficult. Second, nomological sufficiency is commonly rejected for a number of reasons. Most relevantly, nomological sufficiency fails to track causation, let alone causal sufficiency. Kim, for example, argues that a nomological sufficiency relation obtains between a series of shadows from a moving car, but these shadows are not causally efficacious (Kim, 2007, 231–232). And, to return to previous examples, the gun's sound is nomologically sufficient for the death, but the sound is not causally efficacious, let alone causally sufficient, for the death. Third, even granting this definition of sufficient causation, *p* is only nomologically sufficient for *p** if the material conditional $p \rightarrow p^*$ is true in a suitable range of possible worlds. But surely one of the worlds included in this suitable range of possible worlds is the one world that is actually most relevant to the causal exclusion problem, namely, the (p & $\sim m$)-world. This must be the case because the principle of physical causal completeness states that in the circumstance where *p* and *m* both occur, *p* is nevertheless a sufficient cause for *p**, which, in this context, must mean that *p* is able to do the work alone and by itself. For these reasons, while nomological sufficiency may be a necessary condition for sufficient causation, it is not a sufficient condition for sufficient causation. Hopefully this is uncontroversial, as even Kroedel grants that, while sufficient causation involves nomological sufficiency, nomological sufficiency 'Had better not be a sufficient condition for causal sufficiency' (Kroedel 2015a, 373).[8]

4. The compatibilist reply

Given these results, the compatibilist has a straightforward though daunting task. Namely, the compatibilist must secure the principles of mental causation and physical causal completeness, over and against the facts that mental causation is false if (V) is true but physical causal completeness is false if (V) is false, and

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

mental causation may be false if (VI) is false but physical causal completeness is false if (VI) is true.

The compatibilist reply starts by highlighting a relevant disanalogy between the two assassins case and the mental causation case. The two assassins case involves independent sufficient causes, which yields the result that neither bullet is individually necessary as a cause for Smith's death. The mental causation case, however, involves dependent causal processes, which yields the result that the two causal processes necessitate one another. As Barry Loewer explains: 'In the two assassins case the two causes are metaphysically (and nomologically) independent. In the latter case *m* depends on *N* [i.e. *p*] since it is metaphysically (or physically) entailed by it' (cp. Crisp and Warfield 2001, 135; Loewer 2002, 658; Melnyk 2003, 168; Bontly 2005; Walter 2008, 678; Segal 2009, 83; Carey 2011, 261).

This dependency relation between *m* and *p* follows from the following plausible compatibilist assumptions. The compatibilist typically argues that mental events strongly supervene upon physical events, meaning that *p* determines that *m* occurs and *m* depends upon some *p*. Given this strong supervenience relation, it may be metaphysically impossible for *p* to occur without giving rise to *m* (Kallestrup 2006, 473; Marras and Yli-Vakuri 2008, 125; Walter 2008, 678). Moreover, compatibilists typically endorse physicalism, according to which all events, physical and mental, are caused/determined by physical events, so *m* depends on some *p*. Since *m* depends upon some *p*, it is at least nomologically impossible for *m* to occur without some physical subvenience base. Note, however, that compatibilists also argue that mental events are multiply realizable, so *m* could occur without *p*, so long as another physical event p_1 realizes *m*.

The compatibilist intuitions that some *p* necessitates *m* and that *m* implies that some *p* occurs, combined with the compatibilist acceptance of physical causal completeness, which states that *p** implies that some *p* occurred, together entail that *p* cannot precede *p** without *m* preceding *p** as well (Kroedel 2015a, 359; Lewis, 1973, 32). These assumptions also entail that *m* cannot precede *p** without *p* preceding *p** as well. Let the following expression of the argument stand for many:

> You could not delete one of them [i.e. the mental or physical cause] from a given context without thereby deleting the other. If such a relation were to hold … then this is an excellent reason for denying that mental causes and their realizers overdetermine their effects. (cp. Crane 1995, 64–66; Loewer 2002, 658–660; Kroedel 2008, 126; Walter 2008, 678; Arnadottir and Crane 2013, 255)[9]

This solution reveals itself in a counterfactual analysis as well. Namely, on compatibilism, it is not true that *p* without *m* occurs and *p** still occurs, so counterfactual (V) is not true, so it is not established that *m* is unnecessary as a cause of *p**, so it is not established that mental causation is false. Similarly, it is not true that *m* without some *p* occurs and *p** still occurs, so counterfactual (VI) is not

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

true, so it is not established that *p* is unnecessary as a cause of *p**, so it is not established that physical causal completeness is false.

This is too quick, it requires the following details. The compatibilist has two ways to show that counterfactual (VI) is not true. Namely, the compatibilist can show that counterfactual (VI) is false or vacuous. The compatiblist has two strategies for establishing the falsity or vacuity of counterfactual (VI). First, the Replacement Strategy. The nearest possible world where *m* without *p* occurs, and *p** still occurs is not the world of disembodied souls where no *p* occurs and *m* still causes *p** alone. After all, that world is a distant world where both the principles of physical causal completeness and supervenience are false. Rather, the closest possible world where *m* without *p* occurs is the world where *p* is replaced by a slightly different realizer p_1. This is a nearby possible world that only has a slight change in particular fact (cp. Crisp and Warfield 2001, 314; Loewer 2002, 658–660; Block 2003, 136; Raatikainen 2010, 358; Lim 2013, 673; Roche 2014).

It is common to argue that the *m* and p_1 world is nevertheless a *p** world. My desire for a beer, realized by slightly different neurons, still causes me to reach for the beer. Indeed, some argue that this secures the principle of mental causation, as the *m* without *p* world is still a *p** world, indicating that *p** counterfactually depends on *m* (LePore and Loewer 1987, 639–640; Mills 1996, 109; Loewer 2002, 658; Kallestrup 2006, 475–476; Zhong 2011, 134). This may be a mistake, however. While this move shows that *p** counterfactually depends upon *m*, it also shows that *p** does not counterfactually depend upon *p*. After all, if *p* does not occur and *p** still occurs, *p** is not counterfactually dependent upon *p* and hence *p* is not a cause of *p** on the counterfactual analysis (Harbecke 2014, 371; Kroedel 2015a, 363–365).

The world where *m* is replaced by p_1 may actually be the world where *p** is replaced by p_1^* as well. My desire for a beer, realized by slightly different neurons, causes a slightly different reaching for the beer. This view not only secures the counterfactual dependency of *p** on *p*, but it is also a more consistently applied view on event fragility. That is, if one is prepared to say that a few different neural firings changes the neural event from *p* to p_1, then one should also be prepared to say that the resulting few different muscle fibre activations changes the behavioural effect from *p** to p_1^*. The result is that the nearest possible world where *m* occurs without *p* is a world where *m* occurs, and *p* is replaced by p_1, and p_1 causes a slightly different behavioural effect p_1^*. Thus, counterfactual (VI) is false. It is false that 'Had *m* without *p* occurred, *p** would have occurred' since the nearest possible world is the world where *m* without *p* occurs and p_1^* rather than *p** occurs. Since counterfactual (VI) is false, it is false that *p* is unnecessary as a cause of *p**, so it is false that physical causal completeness fails.

Notice, however, that this move renders it false that *m* is a sufficient cause of *p**. After all, the counterfactual 'Had *m* without *p* occurred, *p** would have occurred' is false, since the nearest possible world where *m* without *p* occurs is a world where *p** does not occur, indicating that the presence of *m* alone does

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

not guarantee that *p** occurs (cp. Zhong forthcoming, 10). This point is problematic for those who argue that the mental cause is a sufficient cause. Moreover, some, including some compatibilists, argue that the Replacement Strategy is problematic for other reasons as well. As Lewis himself acknowledges, the closest possible world may not be the most relevant possible world (Lewis 2000, 190; Paul and Hall 2013, 161). Consider this counterfactual: had John not said all those mean things to Jenny, Jenny would not have left. The closest possible world is the world where John says very slightly fewer mean things, and Jenny still leaves, which fails to establish the counterfactual dependence of Jenny's departure on John's meanness. But, the most relevant possible world is the world where John says none of those mean things, and Jenny stays, which establishes counterfactual and causal dependence.

For these reasons, many compatibilists suggest the Excision Strategy is more appropriate (Harbecke 2014, 366; Bennett 2003, 482; cp. Lewis 2000, 190; Collins et al., 2004, 21; Paul and Hall 2013, 161). According to the excision strategy, we do not replace the relevant event with a similar event, rather we completely delete the relevant event, replacing it with a vacuum. Thus, the nearest possible (*m* & ~*p*)-world is the world where *p* is excised, by a 'metaphysical hole-puncher' (Bennett 2003, 482), and replaced with empty space. In the nearest possible world where *p* is excised, the principle of physical causal completeness is still true, so *p** would not occur, and supervenience is still true, so *m* would not occur either (Harbecke 2014, 371). The former condition establishes that counterfactual (VI) is false. Had *p* not occurred, while *m* somehow still occurred, *p** would still not have occurred, so counterfactual (VI) is false. The latter condition establishes that counterfactual (VI) may in fact be vacuous. Given that *m* can only exist if some *p* is present, but we have excised *p* without replacing it with another p_1, *m* cannot occur either. So, there is no possible world where *m* occurs without *p*, rendering counterfactual (VI) vacuous. If the counterfactual is not vacuous in this way, then it is still false for the previously mentioned reason.

The results of the excision and replacement strategies are similar. In both cases, the counterfactual (VI) is false or vacuous. The truth of (VI), however, is required in order to establish that *p* is unnecessary as a cause, so the principle of physical causal completeness is not shown to be false. There is a general consensus that these arguments are persuasive. Even Kim, who was once critical of this type of compatibilist response, acknowledges that his original worries are 'not quite right and at best incomplete' (Kim 2005, 46).

As for counterfactual (V), the compatibilist can show that this counterfactual is vacuous by appealing to the Vacuity Strategy. Here it goes: *p* metaphysically necessitates *m* (by Supervenience), so the counterfactual (p & ~*m*) $\Box\!\!\rightarrow$ *p** is vacuous. Or, there are no metaphysically possible worlds where *p* occurs without *m*, so it is impossible to evaluate whether *p*, without *m*, would cause *p** (cp. Loewer 2002, 658; Bennett 2003, 479; Kallestrup 2006, 472; Walter 2008, 678–679; Carey 2011, 257; Shapiro 2012, 522; Arnadottir and Crane 2013, 255).

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

Hence, it is impossible to establish that *m* is unnecessary as a cause of *p**, so it is impossible to establish that the principle of mental causation is false.

In summary, the compatibilist loses mental causation if (V) is true, but loses physical causal completeness if (V) is false. So, the compatibilist argues that (V) is vacuous, thereby preserving both mental causation and physical causal completeness. And, the compatibilist may lose mental causation if (VI) is false, and loses physical causal completeness if (VI) is true. So, the compatibilist argues either that (VI) is vacuous, or will have to argue that mental causation is not lost if (VI) is false.

5. Problems with compatibilism

Unfortunately, these compatibilist victories come with heavy prices. Beginning with counterfactual (VI), the compatibilist shows that *p* is not unnecessary as a cause of *p** by showing that $(m \,\&\, \sim p) \,\square\!\!\rightarrow p^*$ is false or vacuous. For several reasons, it is unlikely that counterfactual (VI) is vacuous. First, all one needs to do in order to show that the antecedent is not metaphysically impossible is to adopt a replacement strategy. According to the replacement strategy, the nearest world where *m* occurs without *p* is the easily accessible world where *m* occurs with a slightly different realizer p_1. If the compatibilist rejects this suggestion, and holds fast to the excision strategy, counterfactual (VI) is probably not vacuous either. Presumably, the disembodied soul world where *m* occurs without some *p*, though false, is nevertheless metaphysically possible. If so, counterfactual (VI) is not vacuous. To reject this possibility is to argue that physicalism is necessarily true. Some would find it question begging to rule out the possibility that disembodied souls are at least metaphysically possible.

For these reasons, the compatibilist may instead prefer to argue that counterfactual (VI) is false. The downside: it is the truth of this counterfactual that establishes what some take to be a central aspect of the principle of mental causation. For some, the principle of mental causation says that *m* is a sufficient cause of *p**. The event *m* is a sufficient cause of *p** if *m* alone, without *p*, can cause *p**, or, if counterfactual (VI) is true. How can we say that *m* alone is all we need to cause *p**, when it is not true that *m* all by itself causes *p**? On compatibilism, (VI) is not true, so the causal sufficiency of the mental cause cannot be established.

This result serves as an objection to those compatibilists who maintain that *m* must be a sufficient cause of *p**. This result is neither surprising, nor need it undermine compatibilism. It is not surprising because compatibilism is the view that *m* only exists if some *p* exists (by supervenience and physicalism), and that some *p* must cause *p** (by physical causal completeness), so it is not surprising to learn that compatibilism does not state that *m*, without some *p*, is sufficient to cause *p**. While the compatibilist probably cannot argue that *m* is a sufficient cause of *p**, the compatibilist need not argue that *m* is a sufficient cause of *p** anyway (Arnadottir and Crane 2013, 259; Morris 2015, 438). The principle of

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

mental causation states that 'some physical effects have mental causes', it does not state that *m* is causally sufficient for *p**. Moral responsibility is grounded if agents cause actions, not if agents, without physical influence qua disembodied souls, cause actions. Common sense indicates that mental events play some causal role, not that mental events, without physical influence as if in disembodied souls, cause behavioural effects.

As for counterfactual (V), the compatibilist shows that *m* is not unnecessary as a cause of *p** by showing that (*p* & ~*m*) □→ *p** is vacuous. However, it is the truth of this counterfactual that establishes that the principle of physical causal completeness is true. The principle of physical causal completeness says that *p* is a sufficient cause of *p**. A necessary condition for *p* being a sufficient cause of *p** is that *p* alone, without *m*, can cause *p**, or, that counterfactual (V) is true. How can we say that *p* alone is all we need to cause *p** when it is not true that *p* all by itself causes *p**? On compatibilism, (V) is vacuous, so the principle of physical causal completeness cannot be established as true. This does not amount to the falsity of physical causal completeness, but it indicates there is no way to prove that physical causal completeness is true. As Kim complains, 'we would not know what it could mean for both putative causes to be independently causally sufficient for the effect' (Kim 2005, 46). Or, as Daniel Lim frames it: 'How … are we to verify that every physical event at time *t* has an independent, sufficient physical cause at *t* if the physical causes we are interested in can never occur without their non-physical counterparts' (cp. Ritchie 2005, 127; Lim 2011, 3; Moore 2012, 327; Harbecke 2014, 370–371).

Not only does the vacuity strategy render it impossible to prove that physical causal completeness is true, but the vacuity strategy probably indicates that physical causal completeness is false. The principle of physical causal completeness says that *p is* a sufficient cause of *p**, so *p can* cause *p** without *m*, so counterfactual (V) *is* true. Yet according to the vacuity strategy, *p* causing *p** without *m* cannot occur, so *p* is not a sufficient cause of *p**, since (V) is vacuous. The physical cause *p* is only a sufficient cause of *p** if *p* is all that is needed to cause *p**. Yet it is not true that *p** occurs when only *p* occurs, since *p** only occurs when *p* and *m* occur. As Bennett indicates, 'it certainly sounds as though *p* needs m's help' (Bennett 2003, 481; cp. Marras 2007, 319). If *p** also needs *m* to occur, how can *p* be all that is needed for *p** to occur?

I take the former problem with the vacuity strategy to be unpleasant—philosophers generally, but not necessarily, avoid embracing doctrines that are metaphysically impossible to prove. I take the latter problem with the vacuity strategy to be deadly. If the principle of physical causal completeness fails, then compatibilism may fail (since physical causal completeness is one of the principles constituting the doctrine of compatibilism) and compatibilism loses attractiveness (since many philosophers want to hold on to the principle of physical causal completeness).

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

In summary, those attempting to motivate the causal exclusion problem argue that the sufficiency of the physical cause for the behavioural effect renders the mental cause unnecessary as a cause of the behavioural effect, and vice versa. Compatibilists respond that there obtains an especially tight relation between *m* and *p*, such that *m* and *p* are both necessary as causes for the behavioural effect. In so doing, however, the compatibilist also renders it impossible to establish that *m* and *p* are sufficient causes of the behavioural effect—the latter being a problem that is equal to or greater than the original difficulty. My conclusion is not that compatibilism is now a defunct position. Rather, viable avenues of response remain for the compatibilist, but more work needs to be done. Compatibilism, as currently construed and defended, has not fully disentangled itself from the grips of the causal exclusion problem.[10]

Notes

1. I leave aside the possibility, advanced by the so-called autonomists, that mental causation is secured if mental causes only have mental effects (Jackson 1982, 133; Gibbons 2006). I also leave aside the suggestion that mental causation is secured if mental causes have only mental or higher level physical effects, rather than microphysical effects (Zhong 2014, 349–350). While I have some sympathy for these positions, they raise questions about whether the mental causes of these mental and/or higher level physical effects are excluded by the subvenient bases of these mental and/or higher level physical effects (Kim 2005, 39ff). In this paper, I focus on the distinct though related questions about whether the mental causes of lower level physical effects are excluded by the lower level physical causes of these lower level physical effects.
2. I leave open the possibility that mental causes are mental events or mental properties of events.
3. More fully, according to the principle of causal exclusion: 'no single event can have more than one sufficient cause occurring at any given time—unless it is a genuine case of causal overdetermination' (Kim 2005, 42). For Kim, genuine cases of causal overdetermination occur when two independent causal processes converge on the same effect. Neither Kim, nor the compatibilist, argues that mental causation is an instance of this independent overdetermination. So, I bracket out the caveat that permits independent overdetermination, leaving only the contentious claim that no single event can have more than one sufficient cause occurring at any given time.
4. It is worth noting, as will be argued below, that physical causal completeness does not indicate that the specific physical cause *p* is necessary as a cause of *p**, since multiple realizability entails that p_1 rather than *p* might cause *p** (i.e. $\sim p \;\Box\!\!\rightarrow \sim p^*$ may be false). On the contrary, physical causal completeness only indicates that a physical cause is necessary as a cause of *p** (i.e. $\sim \cup p \;\Box\!\!\rightarrow \sim p^*$ must be true).
5. The suggestion is that we remove only *m* to properly analyse the causal contribution of *m* and *p* in the case where *m* and *p* and *p** all occur in the actual world. This suggestion follows from the classic case of overdetermination, where we discern whether Smith's death is overdetermined by removing the bullet *a* firing while keeping the bullet *b* firing, to discern whether the bullet *a* firing is

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

a sufficient cause of Smith's death and whether the bullet *b* firing is necessary as a cause of Smith's death. After all, little is revealed by removing both bullets and Smith's death. We still would not know whether the bullet *a* firing was sufficient or necessary or irrelevant, or whether the bullet *b* firing was sufficient or necessary or irrelevant, or whether Smith's death was overdetermined or jointly caused. Likewise, it reveals little to remove both *m* and *p* then notice that *p** does not occur. This is worth mentioning because some try to solve the mental causation problem by arguing that the nearest possible world where *m* does not occur is the world where neither *p* nor *p** occur either, thereby establishing that *m* is a cause of *p** (Kroedel 2015a). I rule out this possibility by focusing attention on the nearest relevant possible world (cp. Bernstein 2016, 21), which is the world where *m* is removed but *p* remains, or, where only *m* is removed.

6. There are several other reasons to refrain from the view that counterfactual dependency is sufficient for causation. First, for reasons discussed below, this leaves compatibilism back in the unenviable position of being unable to establish that *p* is a sufficient cause of *p**. Second, this move suffers from the Too Many Causes problem. If *m* is established as a cause of *p** because $\sim m \mathrel{\Box\!\!\rightarrow} \sim p^*$, then a wide array of other necessarily present, seemingly epiphenomenal, events must also be considered causes of *p**. For example, the bullet's firing causes the victim to die, but the object's firing also necessary occurs, so it is a cause. And, the object's being a colour necessarily occurs, so it is a cause, and the object's being a shape necessarily occurs, so it is a cause and the object's making a sound necessarily occurs, so it is a cause, etc. Third, if one accepts the currently flourishing possibility that impossible worlds exist (Brogaard and Salerno 2013; Bjerring 2014), then it is possible to re-establish the difficulties raised below. Namely, the impossible world in which *p* occurs without *m*, yet *p** still occurs is closer to our world than the more distant impossible world where *p* occurs without *m*, yet *p** does not occur. After all, the former impossible world only violates logical law, while the latter impossible world violates both logical law and physical causal completeness. This being the case, while it is logically impossible to imagine *p* without *m*, it is nevertheless reasonable that *p* without *m* would still cause *p**, since the compatibilist also wants to endorse physical causal completeness. So, while *m* is necessarily present, it remains unlikely that *m* is necessarily a cause of *p**.
7. This is a tempting move, as it may permit the bulk of the difficulties raised below against the compatibilist to dissipate. Here is how: if nomological sufficiency is all that is required for sufficient causation, then it is possible for the mental event to be necessary for the effect, while the physical event is nomologically sufficient for the effect. After all, while $(p \,\&\, \sim m) \mathrel{\Box\!\!\rightarrow} p^*$ is vacuous, rendering it unestablished that *m* is unnecessary as a cause of *p**, it is also the case that $p \mathrel{\Box\!\!\rightarrow} p^*$ is true in the actual world and nearby worlds, as *p* necessitates *p** in all the possible worlds where physical causal completeness holds, rendering *p* nomologically sufficient for *p**. This solution fails, however, once it is established that nomological sufficiency is not a sufficient condition for sufficient causation.
8. Here are the two additional definitions of sufficient causation: *p* is a sufficient cause for *p** if *p* is the minimal set of individually necessary causes that are jointly sufficient for *p** (cp. Mackie 1975; Taylor 1976, 298; Paul and Hall 2013, 14). On the counterfactual analysis, *p* is one of the necessary causes for *p** if the counterfactual 'had only *p* not occurred, *p** would not have occurred' is true, and, from above, *p* is the sufficient as a cause for *p** if the counterfactual 'had

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

only *p* occurred, *p** would have occurred' is true, where *p* contains the entire plurality of necessary causes. Call this absolute sufficiency, and it poses problems for compatibilism as well. Namely, the minimal set indicates that any events that are not individually necessary causes are left out of the sufficient cause. Now, either *m* is included in the minimal set *p* or *m* is excluded from the minimal set *p*. If *m* is included in the minimal set, then physical causal completeness is false, since the sufficient cause is not entirely physical. If *m* is excluded from the minimal set, then *m* is unnecessary as a cause for the effect. After all, the minimal set consists of all of the individually necessary causes, so if *m* is excluded from the minimal set, it cannot be an individually necessary cause of *p** and the principle of mental causation is false. Here is the another definition of sufficient causation: *p* is a sufficient cause of *p** if *p*, in the circumstances, or, holding other facts the same, causes *p** (Mills 1996, 115; Bennett 2003, 490; Crane 2004, 234–235; Arnadottir and Crane 2013, 259–260). With respect to the counterfactual model, *p* is a cause of *p** when, in the nearest possible world in which *p* does not occur, *p** does not occur. But, in the nearest possible world, by definition, the circumstances remain almost entirely fixed. So, holding everything else equal, *p* is sufficient for *p**, given that removing *p* removes *p**, and keeping *p* keeps *p**. Call this circumstance sufficiency, but it cannot be a sufficient condition for sufficient causation. After all, it is probable that numerous synchronous neural firings cause a behavioural effect. If sufficient causation is only sufficient in the circumstances, then removing a small neural cluster within this larger neural process prevents the effect from occurring, so it is a sufficient cause for the effect. At the same time, removing another small neural cluster within the larger neural process also prevents the effect from occurring, so it is a sufficient cause for the effect as well, etc. As it turns out, each behavioural effect has a multitude of sufficient neural causes—and this is true before even considering the additional mental cause. This picture is not consistent with the causal exclusion problem as traditionally imagined.

9. It is worth pointing out that compatibilists use this metaphysical backdrop to argue that the coincidence argument for the causal exclusion principle is unpersuasive. Recall, the coincidence argument for the causal exclusion principle stipulates that it is a rare coincidence when multiple causal processes converge on the same effect, but mental causation is ubiquitous, so it would be too coincidental for mental causes and physical causes to continuously converge on behavioural effects. The compatibilist argues, on the contrary, that physical causes of behavioural effects metaphysically necessitate mental causes of behavioural effects, which replaces the air of coincidence with the systematic expectation that physical causes and mental causes continuously converge on behavioural effects (Funkhouser 2002, 338; Sider 2003, 722; Walter 2008, 678; Arnadottir and Crane 2013, 261–262; Kroedel 2015a, 368).
10. I would like to thank two anonymous referees for valuable comments and suggestions. I would also like to thank several audience members and the commentators at the 2015 Canadian Philosophical Association conference and the 2015 Western Canadian Philosophical Association conference for valuable feedback on earlier incarnations of this paper.

Notes on contributor

Dwayne Moore is an assistant professor of Philosophy at the University of Saskatchewan.

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

References

Anscombe, E. 1971. *Causality and Determination*. London: Cambridge University Press.

Arnadottir, S., and T. Crane. 2013. "There is No Exclusion Problem." In *Mental Causation and Ontology*, edited by S. Gibb, and R. Ingthorsson, 248–265. Oxford: Oxford University Press.

Audi, P. 2013. "Causation, Coincidence, and Commensuration." *Philosophical Studies* 162 (2): 447–464.

Bennett, K. 2003. "Why the Exclusion Problem Seems Intractable, and How." *Just Maybe, to Tract it', Noûs* 37: 471–471.

Bernstein, S. 2016. "Overdetermination Underdetermined." *Erkenntnis*, 81 (1): 17–40.

Bjerring, J. 2014. "On Counterpossibles." *Philosophical Studies* 168: 327–353.

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

Bogardus, T. 2013. "Undefeated Dualism." *Philosophical Studies* 165 (2): 445–466.

Bontly, T. 2005. "Exclusion, Overdetermination, and the Nature of Causation." *Journal of Philosophical Research* 30: 261–282.

Brand, M. 1976. "Introduction: Defining Causes." In *The Nature of Causation*, edited by M. Brand, 1–44. Urbana: University of Illinois Press.

Brand, M., and M. Swain. 1974. "Causation and necessary and sufficient conditions: Reply to Hilpinen." *Philosophical Studies* 25 (5): 357–364.

Brogaard, B., and J. Salerno. 2013. "Remarks on Counterpossibles." *Synthese* 190 (4): 639–660.

Campbell, S. 2004. *Flaws and Fallacies in Statistical Thinking*. Mineola, NY: Courier Dover Publications.

Carey, B. 2011. "Overdetermination and The Exclusion Problem". *Australasian Journal of Philosophy* 89 2: 251–262.

Collins, J., N. Hall, and L. Paul. 2004. "Causation and Counterfactuals." In *Causation and Counterfactuals*, edited by J. Collins, N. Hall, and L. A. Paul, 1–59. Cambridge: MIT Press.

Corry, R. 2013. "Emerging from the Causal Drain." *Philosophical Studies* 165 (1): 29–47.

Crane, T. 1995. "Mental Causation." *Aristotelian Society Supplementary Volume* 69: 211–254.

Crane, T. 2004. "Summary of Elements of Mind and Replies to Critics." *Croatian Journal of Philosophy* 4 (11): 223–240.

Crisp, T., and T. Warfield. 2001. "Kim's Master Argument." *Noûs* 35: 304–316.

Engelhardt, J. 2012. "Varieties of Multiple Antecedent Cause." *Acta Analytica* 27: 231–246.

Engelhardt, J. 2015. "What is the Exclusion Problem?" *Pacific Philosophical Quarterly* 96: 205–232.

Esfeld, M. 2005. "Mental Causation and Mental Properties." *Dialectica* 59 (1): 5–18.

Esfeld, M. 2010. "Causal Overdetermination for Humeans?" *Metaphysica* 11 (2): 99–104.

Fodor, J. 1989. "Making Mind Matter More." *Philosophical Topics* 17 (1): 59–79.

Foster, J. 1989. "A Defence of Dualism." In *The Case for Dualism*, edited by J. Smythies, and J. Beloff. Charlottesville: University of Virginia Press.

Funkhouser, E. 2002. "Three Varieties of Causal Overdetermination." *Pacific Philosophical Quarterly* 83: 335–351.

Gibbons, J. 2006. "Mental Causation without Downward Causation." *Philosophical Review* 115: 79–103.

Goldman, A. 1969. "The Compatibility of Mechanism and Purpose." *The Philosophical Review* 78: 468–482.

Harbecke, J. 2008. *Mental Causation*. Frankfurt: Ontos Verlag.

Harbecke, J. 2013. "Mental Causation and the New Compatibilism." *Abstracta* 7 (1): 55–71.

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

Harbecke, J. 2014. "Counterfactual Causation and Mental Causation." *Philosophia* 42 (2): 363–385.
Horgan, T. 1997. "Kim on Mental Causation and Causal Exclusion." *Nous Supplement: Philosophical Perspectives* 11: 165–184.
Jackson, F. 1982. "Epiphenomenal Qualia." *The Philosophical Quarterly* 32 (127): 127–136.
Kallestrup, J. 2006. "The Causal Exclusion Argument." *Philosophical Studies* 131: 459–485.
Kim, J. 1989. "Mechanism, Purpose, and Explanatory Exclusion." *Philosophical Perspectives* 3: 77–108.
Kim, J. 1993. *Supervenience and Mind*. Cambridge: Cambridge University Press.
Kim, J. 2007. "Causation and Mental Causation." In *Contemporary Debates in Philosophy of Mind*, edited by B. P. McLaughlin, and J. D. Cohen. Oxford: Blackwell.
Kim, J. 1998. *Mind in a Physical World*. Cambridge: MIT Press.
Kim, J. 2005. *Physicalism, or Something Near Enough*. Princeton: Princeton University Press.
Kim, J. 2006. "Emergence: Core ideas and issues." *Synthese* 151: 547–559.
Kim, J. 2009. "Mental Causation." In *Oxford Handbook of Philosophy of Mind*, edited by B. McLaughlin, A. Beckermann, and S. Walter, 29–52. Oxford: Oxford University Press.
Kroedel, T. 2008. "Mental Causation as Multiple Causation." *Philosophical Studies* 139: 125–143.
Kroedel, T. 2015a. "Dualist Mental Causation and the Exclusion Problem." *Noûs* 49 (2): 357–375.
Kroedel, T. 2015b. "A Simple Argument for Downward Causation." *Synthese* 192 (3): 841–858.
Kroedel, T., and M. Schultz. 2016. "Grounding Mental Causation." *Synthese*, 193 (6): 1909–1923.
LePore, E., and B. Loewer. 1987. "Mind Matters." *Journal of Philosophy* 84: 630–642.
Lewis, D. 1973a. "Causation." *The Journal of Philosophy* 70: 556–567.
Lewis, D. 1973b. *Counterfactuals*. Oxford: Blackwell Publishers.
Lewis, D. 1979. "Counterfactual Dependence and Time's Arrow." *Noûs* 13 (4): 455–476.
Lewis, D. 1986. *On the Plurality of Worlds*. Oxford: Blackwell.
Lewis, D. 2000. "Causation as Influence." *The Journal of Philosophy* 97: 182–197.
Lim, D. 2011. "Exclusion, Overdetermination, and Vacuity." *Southwest Philosophy Review* 27 (1): 57–64.
Lim, D. 2013. "Why Not Overdetermination?" *The Heythrop Journal* 54 (4): 668–677.
Loewer, B. 2002. "Comments on Jaegwon Kim's Mind and the Physical World." *Philosophy and Phenomenological Research* 65: 655–662.
Lyons, J. 2006. "In Defense of Epiphenomenalism." *Philosophical Psychology* 19: 767–794.
Mackie, J. 1975. "Causes and Conditions." In *Causes and Conditionals*, edited by E. Sosa, 245–264. Oxford: Oxford University Press.
Malcolm, N. 1968. "The Conceivability of Mechanism." *The Philosophical Review* 77: 45–72.
Marc-Wogau, K. 1962. "On Historical Explanation." *Theoria* 28 (3): 213–233.
Marras, A. 2007. "Kim's Supervenience Argument and Nonreductive Physicalism." *Erkenntnis* 66: 305–327.
Marras, A., and J. Yli-Vakuri. 2008. "The Supervenience Argument." In *Tropes, Universals and the Philosophy of Mind*, edited by S. Gozzano, and F. Orilia, 101–134. Frankfurt: Ontos Verlag.
Meixner, U. 2008. "New Perspectives for a Dualistic Conception of Mental Causation." *Journal of Consciousness Studies* 15: 17–38.
Melnyk, A. 2003. *A Physicalist Manifesto*. Cambridge: Cambridge University Press.
Mills, E. 1996. "Interactionism and Overdetermination." *American Philosophical Quarterly* 33 (1): 105–115.

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press

Moore, D. 2012. "Causal Exclusion and Dependent Overdetermination." *Erkenntnis* 76 (3): 319–335.
Morris, K. 2015. "Against Disanalogy Style Responses to the Exclusion Problem." *Philosophia* 43: 435–453.
Ney, A. 2007. "Can an Appeal to Constitution Solve the Exclusion Problem?" *Pacific Philosophical Quarterly* 88: 486–506.
Papineau, D. 2001. "The Rise of Physicalism." In *Physicalism and its Discontents*, edited by C. Gillett, and B. Loewer, 3–36. Cambridge: Cambridge University Press.
Paul, L., and N. Hall. 2013. *Causation*. Oxford: Oxford University Press.
Pereboom, D. 2002. "Robust Nonreductive Materialism." *The Journal of Philosophy* 99: 499–531.
Raatikainen, P. 2010. "Causation, Exclusion, and the Special Sciences." *Erkenntnis* 73 (3): 349–363.
Rasmussen, S. 1982. "Ruben on Lewis and Causal Sufficiency." *Analysis* 42 (4): 207–211.
Ritchie, J. 2005. "Causal Compatibilism—What Chance?" *Erkenntnis* 63 (1): 119–132.
Robinson, W. 2004. *Understanding Phenomenal Consciousness*. Cambridge: Cambridge University Press.
Roche, M. 2014. "Causal Overdetermination and Kim's Exclusion Argument." *Philosophia* 42: 809–826.
Ruben, D. 1981. "Lewis and the Problem of Causal Sufficiency." *Analysis* 41 (1): 38–41.
Sanford, D. 1975. "Causal Necessity and Logical Necessity." *Philosophical Studies* 28: 103–112.
Schiffer, S. 1987. *Remnants of Meaning*. Cambridge: MIT Press.
Segal, G. 2009. "The Causal Inefficacy of Content." *Mind and Language* 24: 80–102.
Shapiro, L. 2012. "Mental Manipulations and the Problem of Causal Exclusion." *Australasian Journal of Philosophy* 90 (3): 507–524.
Shoemaker, S. 2007. *Physical Realization*. Oxford: Oxford University Press.
Sider, T. 2003. "What's so bad about overdetermination?" *Philosophy and Phenomenological Research* 67: 719–726.
Taylor, R. 1976. "Causation." In *The Nature of Causation*, edited by M. Brand, 279–305. Urbana: University of Illinois Press.
Tranøy, K. 1962. "Historical Explanation: Causes and Conditions." *Theoria* 28 (3): 234–249.
Turner, M. 2008. *The Mind-Body Problem: Knot Or Not?* Xlibris Corporation.
Walter, S. 2008. "The Supervenience Argument, Overdetermination, and Causal Drainage: Assessing Kim's Master Argument." *Philosophical Psychology* 21: 673–696.
Yablo, S. 1992. "Mental Causation." *The Philosophical Review* 101: 245–280.
Zhong, L. 2011. "Can Counterfactuals Solve the Exclusion Problem?" *Philosophy and Phenomenological Research* 83 (1): 129–147.
Zhong, L. 2014. "Sophisticated Exclusion and Sophisticated Causation." *Journal of Philosophy* 111: 361–380.
Zhong, L. forthcoming. "Semantic Normativity and Semantic Causality." *Philosophy and Phenomenological Research*, Online First.

https://doi.org/10.1080/00455091.2016.1218254 Published online by Cambridge University Press