

Dualist Mental Causation and the Exclusion Problem

THOMAS KROEDEL
Humboldt University of Berlin

Abstract

The paper argues that dualism can explain mental causation and solve the exclusion problem. If dualism is combined with the assumption that the psychophysical laws have a special status, it follows that some physical events counterfactually depend on, and are therefore caused by, mental events. Proponents of this account of mental causation can solve the exclusion problem in either of two ways: they can deny that it follows that the physical effect of a mental event is overdetermined by its mental and physical causes, or they can accept that the physical effect is overdetermined but claim that this is unproblematic because the case is sufficiently dissimilar to prototypical cases of overdetermination.

1. Introduction

Dualism cannot accommodate mental causation, or only at unacceptable costs. If dualists do not want to give up the claim that some physical effects have mental causes altogether, they either have to reject that all physical effects have physical causes or accept that physical effects with mental causes are overdetermined in the way the deaths of firing squad victims are. This is the common view. It is false. As this paper argues, dualists can endorse the efficacy of the mental without great sacrifice.

The plan is as follows. Section 2 introduces the dualist position and Lewis's account of counterfactual conditionals, which is important because counterfactuals will play a crucial role in the causal claims that interest us. Section 3 argues that dualists can explain mental causation if they assume that the psychophysical laws have a special status and that counterfactual dependence is sufficient for causation. Sections 4 and 5 address the exclusion problem. Section 4 argues that dualists can reject the claim that mental and physical causes overdetermine their behavioral effects. Section 5 argues that, even if this kind of overdetermination is accepted, it is not objectionable. Section 6 discusses various objections to the preceding arguments, and Section 7 concludes.

2. Preliminaries

Dualists hold that mental events and properties are not physical. They hold, that is, that mental events and properties are neither identical to nor realized by physical events and properties (see Lowe 2003: 145). I shall confine my attention to

scientifically-minded dualists who take it to be metaphysically contingent but at least a matter of natural law that mental events and properties are accompanied by such-and-such physical events and properties (see Chalmers 1996). These dualists will say, for instance, that it is metaphysically contingent that someone who is in pain is in some physical state or other and likewise metaphysically contingent that someone who is in a given physical state is in pain. Still, they hold, it is at least a matter of natural law that someone is in pain if their c-fibers are firing; that someone is in pain if their x-fibers (which are not actually present in humans) are firing; and so on.¹

In the following, I shall mainly talk about the psychophysical relation in terms of events, as most of the discussion is going to be about broadly causal relations, which are most naturally stated in terms of events. I believe that, in our context, there is not much to choose between talking about events and talking about property instantiations, since the former are intimately related to the latter.² But it would not matter if I erred here: all the arguments to follow could be stated in terms of properties and their instances instead. For events, then, our dualists' commitment is as follows: for each token mental event m , there are simultaneous actual and possible token physical events p_1, p_2 , etc. (call them m 's *physical bases*) such that

- (i) if m occurs, then one of p_1, p_2, \dots occurs;
- (ii) for each p_i out of p_1, p_2, \dots , if p_i occurs, then m occurs;
- (iii) the conditionals in (i) and (ii) are metaphysically contingent.

I shall assume that our dualists take instances of (i) and (ii) to be entailed by certain psychophysical laws, which therefore inherit the contingency of those instances.

Let us now turn to counterfactuals. When evaluating counterfactual conditionals, I shall assume Lewis's (1973b) truth conditions, according to which a counterfactual 'If ϕ were the case, then ψ would be the case' ($\phi \Box \rightarrow \psi$) is true if and only if either there are no possible worlds where ϕ is true (the case of *vacuous truth*) or there is a world where both ϕ and ψ are true which is closer (that is, more similar overall) to the actual world than any worlds where ϕ is true while ψ is false. I shall also follow Lewis in taking the 'might' conditional 'If ϕ were the case, then ψ might be the case' ($\phi \Diamond \rightarrow \psi$) to be equivalent to the negation of 'If ϕ were the case, then ψ would not be the case'. Thus, 'If ϕ were the case, then ψ might be the case' is true if and only if there is a world where both ϕ and ψ are true which is at least as close to the actual world as any worlds where ϕ is true while ψ is false. Lastly, I shall—provisionally, at least—accept Lewis's account of how aspects of similarity determine the relation of closeness or comparative overall similarity that is in play when counterfactuals are evaluated in standard contexts:³ A world where a large-scale violation of the actual laws of nature (a 'big miracle') occurs is always less similar overall to our world than a world without such large-scale violations. Among worlds that are on a par with respect to large-scale violations of the actual laws of nature, a world that matches the actual world perfectly in particular fact throughout a larger region of space-time than another one is always more similar overall to the actual world than the latter. Among worlds that are on a par with

respect to all of the above, a world where a small-scale violation of the actual laws of nature (a ‘small miracle’) occurs is always less similar overall to our world than a world without such small-scale violations.

With this apparatus in place, let us turn to mental causation.

3. Mental Causation

My headache caused me to take an aspirin. This claim sounds as natural as any. A dualist too can make it. More importantly, a dualist can provide a rigorous argument for it. Nothing depends on the specifics of headaches and our reactions to them, so let m be some actually occurring mental event and b its actually occurring later behavioral effect (‘putative behavioral effect’, if you like, to quell any suspicion of begging the question). The argument has a complicated part with the conclusion that if m had not occurred, then b would not have occurred, and a simple part with the conclusion that m caused b . Let us start with the complicated part:

- (1) If none of m ’s physical bases had occurred, then b would not have occurred.
($\sim\text{UP} \square \rightarrow \sim\text{B}$)⁴
 - (2) If m had not occurred, then none of m ’s physical bases would have occurred.
($\sim\text{M} \square \rightarrow \sim\text{UP}$)
 - (3) If none of m ’s physical bases had occurred, then m would not have occurred.
($\sim\text{UP} \square \rightarrow \sim\text{M}$)
-
- (4) If m had not occurred, then b would not have occurred.
($\sim\text{M} \square \rightarrow \sim\text{B}$)

The logical validity of this argument is beyond reproach. Counterfactuals are not generally transitive, so (4) does not follow from (1) and (2) alone. Roughly speaking, the truth of (1) and (2) leaves it open that the closest worlds where m does not occur do not coincide with the closest worlds where none of m ’s physical bases occurs, thus leaving it open that in the closest worlds where m does not occur, b still occurs, which would falsify (4). However, together with premise (3), (2) guarantees that the closest worlds where m does not occur coincide with the closest worlds where none of m ’s physical bases occurs. Since by (1) the latter are worlds where b does not occur, so are the former.⁵

Premise (1), I take it, is uncontroversial. In the actual world, m is accompanied by its actual physical base (call it p_1). According to Lewis’s account of closeness, the closest worlds where the antecedent of (1) is true are as follows. History matches that of the actual world until just before the time of the actual occurrence of p_1 , when a small miracle prevents the occurrence of p_1 or any of m ’s other physical bases. Thence, lawful evolution does not lead to the occurrence of b .⁶ So in the closest worlds where none of m ’s physical bases occurs, b does not occur either. Hence, (1) is true.

Premise (2) is the trickiest. In fact, on the face of it, it might seem false. Compare the following two types of worlds for closeness to the actual world: In worlds of

type one, the physical laws are violated while the psychophysical laws are not, so that m 's failure to occur implies the failure of any of its physical bases to occur. In worlds of type two, the psychophysical laws are violated while the physical laws are not, so that m 's actual physical base still occurs, but m does not. It might seem that type-two worlds are closer to the actual world than type-one worlds. While, it might be held, the respective types are on a par as far as violations of law are concerned, there is vastly more match of particular fact to the actual world in the type-two worlds, for type-two worlds match the actual world perfectly from the time of m 's actual occurrence onwards. Type-one worlds, by contrast, cannot equal this match; owing to the failure of any of m 's physical bases to occur, they lawfully evolve into a different future.⁷ In type-two worlds, the antecedent of (2) is true while its consequent is false; in type-one worlds, both are true. If type-two worlds come out closer to the actual world than type-one worlds, as seems plausible, (2) is false.

Dualists can resist this line of reasoning, however. They claim that the relation between mental events and physical events is contingent. Specifically, they claim that it is contingent that the occurrence of a physical base of m implies the occurrence of m . The psychophysical laws that entail such contingent implications must be contingent as well. But nothing forces dualists to accept that psychophysical laws are modally on a par with ordinary laws of nature, such as the laws of physics. They are within their rights to claim that psychophysical laws could not have failed as easily as the other laws. They can claim, in other words, that worlds where the psychophysical laws are violated are further from actuality than any worlds where only the ordinary laws are violated.⁸ Lewis's account of the closeness or similarity relation does not make provisions for a special status of the psychophysical laws. This is not surprising, since Lewis himself was a materialist (see 1994). His account can easily be modified to accommodate the distinction, however.⁹ Then the new principal criterion for overall similarity to the actual world is that none of the actual psychophysical laws be broken, so that a world where a violation of the actual psychophysical laws occurs is always less similar overall to our world than a world without such violations. (Call such a violation a *psychophysical miracle*.) The new principal criterion can be grafted onto Lewis's original account. Thus, a world where a violation of the actual psychophysical laws occurs is always less similar overall to our world than a world without such violations; among worlds that are on a par with respect to violations of the actual psychophysical laws, a world where a large-scale violation of the ordinary actual laws of nature occurs is always less similar overall to our world than a world without such large-scale violations; and so on. On the modified account of overall similarity, type-one worlds come out more similar overall to the actual world than type-two worlds since they involve no violation of the psychophysical laws. Hence, on the modified account, (2) is true.

Given the modified Lewisian account of overall similarity, premise (3) comes out true too. Worlds where the antecedent holds in the absence of a psychophysical miracle are closer to the actual world than any worlds where such a miracle takes

place. But if the actual psychophysical laws are intact at a world where none of m 's physical bases occurs, m does not occur there either. So in the closest worlds where none of m 's physical bases occurs, m does not occur; hence (3) is true.¹⁰

So much for the complicated part of the argument. (At least for the time being; I shall address some objections in Section 6.) The simple part is short in the telling. Say that event e *counterfactually depends* on event c if and only if e would not have occurred had c not occurred.¹¹ Counterfactual dependence between distinct events is sufficient for causation: if some event counterfactually depends on another, distinct event, the latter causes the former (see Lewis 1973a, 2004). Events m and b are distinct. By (4), b counterfactually depends on m . Therefore, m causes b .¹²

Two further remarks about counterfactuals and causation are in order. First, the claim that counterfactual dependence is sufficient for causation, which we just used in the simple part of the argument, is subject to the qualification that the relevant counterfactuals be evaluated neither in a “backtracking” nor a “back-and-then-forward” way (Lewis 2004: 78). Perhaps on such a back-and-then-forward evaluation the counterfactual ‘If the barometer had not fallen, then (the air pressure would not have fallen earlier and so) there would not have been a storm’ is true. But we should not conclude that the barometer reading caused the storm. Lewis’s (1979) truth-conditions plus his account of overall similarity are not supposed to yield the result that any backtracking or back-and-then-forward counterfactuals come out true, so that counterfactual dependence is sufficient for causation without qualification. I shall address an objection to the effect that Lewis does not succeed in ruling out true backtracking or back-and-then-forward evaluations of counterfactuals in Section 6. For now, I would merely like to note that *if* Lewis’s theory succeeds in yielding no true backtracking or back-and-then-forward evaluations of counterfactuals, so will the combination of his truth-conditions with the modified account of overall similarity, where psychophysical laws feature in the new principal criterion of similarity. Our psychophysical laws are synchronic, so holding them fixed by itself never requires changing the past.

Second, while counterfactual dependence is merely a sufficient condition for causation and not a necessary one, in standard cases effects counterfactually depend on their causes. It seems plausible that m 's actual physical base, p_1 , causes b , and that there is nothing unusual about this causal relation. Hence, it seems plausible that

- (5) If p_1 had not occurred, then b would not have occurred.¹³
 $(\sim P_1 \square \rightarrow \sim B)$

Claim (5) will become relevant for some of the arguments in later sections.

We have seen that, assuming dualism, the critical condition for establishing that behavioral events counterfactually depend on, and hence are caused by, mental events is that worlds where the actual psychophysical laws are violated are always less similar overall to our world than worlds without such violations, irrespective of violations of ordinary laws of nature. Call the conjunction of this condition and the

position of dualism as specified in the previous section *super-nomological dualism*. Then the upshot so far is this: while other varieties of dualism may struggle at the task, super-nomological dualism can explain mental causation.

4. Denying Exclusion

It is one thing to argue that some mental events cause physical events; it is another to show that they do so in an unobjectionable way. In the previous section, I have done the former on behalf of the dualist; in this and the following section, I shall do the latter.

These days every account of mental causation needs to address the so-called exclusion problem.¹⁴ This problem arises from the following five principles:

(DISTINCTNESS) All mental events are distinct from physical events.

(COMPLETENESS) Every physical event that has a cause at all has a sufficient physical cause.

(EFFICACY) Some mental events have physical effects.

(EXCLUSION) If an event has more than one simultaneous sufficient cause, then it is overdetermined.

(NON-OVERDETERMINATION) The effects of mental events are not systematically overdetermined.

Each of these principles seems plausible. Yet together they form an inconsistent pentad,¹⁵ so at least one of them has to be rejected. Since dualists hold that mental events can occur without their physical bases and vice versa, they cannot deny (DISTINCTNESS). As was argued in the previous section, dualists can defend the claim that some mental events have physical effects. Hence, they can uphold (EFFICACY) and resist epiphenomenalism. Rejecting (COMPLETENESS) seems hopeless, not just from a general scientific point of view, but also in the case at hand. It is hard to deny that m 's actual physical base, p_1 , causes the behavioral effect b , so our model of mental causation fails to generate a counterexample to (COMPLETENESS).

Rejecting (EXCLUSION) or rejecting (NON-OVERDETERMINATION) are the only options left for dualists then. Karen Bennett (2008) has argued that dualists have to accept (EXCLUSION). In the remainder of this section, I shall argue that this is not the case and that rejecting (EXCLUSION) is a viable option for dualists; indeed, (EXCLUSION) will be shown to be inconsistent with the dualist account of mental causation. In the following section, I shall argue that, even if it should turn out that dualists have to accept (EXCLUSION), they could still reject (NON-OVERDETERMINATION).

According to (EXCLUSION), an event with more than one simultaneous sufficient cause is overdetermined. Whatever overdetermination is, it at least requires that the overdetermined event would still have occurred if only one of the overdetermining events had occurred.¹⁶ In our case, this means that if b is overdetermined by m and p_1 , then the following two counterfactuals are true:

(O₁) If m had occurred without p_1 , then b would still have occurred.
 (M & $\sim P_1 \square \rightarrow B$)

(O₂) If p_1 had occurred without m , then b would still have occurred.
 ($\sim M$ & $P_1 \square \rightarrow B$)

Bennett (2003, 2008) and Eugene Mills (1996) endorse the stronger requirement that (O₁) and (O₂) must be non-vacuously true in order for m and p_1 to overdetermine b . Bennett (2008: 287–292) argues that dualists have to concede the non-vacuous truth of both (O₁) and (O₂), so that—pending the satisfaction of whatever further necessary conditions for overdetermination there are—dualists have to accept (EXCLUSION).

I agree with Bennett that dualism implies the non-vacuous truth of (O₂). According to our dualists, there are worlds where p_1 occurs while m does not, although this requires a violation of the actual psychophysical laws. Given the modified Lewisian account of overall similarity presented in the previous section, worlds that do not in addition involve any ‘ordinary’ miracles are closer to the actual world than any worlds that do. Assuming that b lawfully follows from the previous physical state, which includes p_1 , b still occurs in the closest worlds where p_1 occurs without m . In other words, the closest worlds where the antecedent of (O₂) is true are just like the actual world except that the occurrence of m is removed by a psychophysical miracle. Hence (O₂) is non-vacuously true.

The case of (O₁) is different, however. What we should think about various claims related to (O₁) depends on what we should think about the nature of m ’s actual physical base, p_1 , and about the nature of the behavioral event b . These events might or might not be very *fragile*—that is, they might not or might easily have occurred in a different time and manner.¹⁷ I shall argue that, on all reasonable assumptions about the fragility of p_1 and b , (O₁) is false.

Suppose that event p_1 is very fragile. Then it takes comparatively little for it not to occur. Perhaps p_1 is essentially a c-fiber firing at a rate between 99 and 101 Hz; then my c-fibers’ firing at a rate of 102 Hz instead of the actual rate of, say, 100 Hz would have been enough for p_1 not to occur. If p_1 is very fragile, a different physical base of m might have taken its place if it had not occurred:

(6) If p_1 had not occurred, then some other physical base of m might have occurred instead.
 ($\sim P_1 \diamond \rightarrow \cup P$)

What counterfactuals about the relation between m ’s physical bases and b are true depends on the fragility (or lack thereof) of b as well. Suppose first that b is *not* very fragile. Let us assume that I’m actually slowly reaching for an aspirin with my right hand; then the assumption that b is not very fragile would allow that b would still have occurred if I had, say, quickly reached for an aspirin with my left hand. It seems that if b is not very fragile while p_1 is very fragile, then b would still have occurred if some alternative physical base of m had occurred instead of p_1 :

(7) If some other physical base of m had occurred instead of p_1 , then b would still have occurred.

$(\sim P_1 \ \& \cup P \ \Box \rightarrow B)$

For it seems that if my c-fibers had fired only slightly differently, I would still have taken an aspirin, albeit presumably somewhat differently (by reaching faster than I actually did, say); this would still have sufficed for b to occur if b is not very fragile. Claims (6) and (7) logically imply:

(8) If p_1 had not occurred, then b might still have occurred.

$(\sim P_1 \ \Diamond \rightarrow B)$

The inference rule used here is that which licenses the inference from $\chi \ \Diamond \rightarrow \phi$ and $\chi \ \& \ \phi \ \Box \rightarrow \psi$ to $\chi \ \Diamond \rightarrow \psi$, which is logically valid.¹⁸ In Section 3, we saw that it is plausible that b counterfactually depends on p_1 :

(5) If p_1 had not occurred, then b would not have occurred.

$(\sim P_1 \ \Box \rightarrow \sim B)$

By the definition of the ‘might’ conditional, (5) is inconsistent with (8), however. So if we assume that p_1 is very fragile (which yields (6)) while b is not very fragile (which yields (7)), it follows (via (8)) that b does not counterfactually depend on p_1 . Contrapositively, if we want to uphold the claim that b counterfactually depends on p_1 , we have to reject either the assumption that p_1 is very fragile or the assumption that b is not very fragile. I take it that the plausibility of the claim that b counterfactually depends on p_1 outweighs that of either assumption. We can therefore conclude that the sub-case where p_1 is very fragile while b is not does not obtain. This leaves us with the case where p_1 is not very fragile and the case where p_1 and b are both very fragile.

Suppose that p_1 and b are both very fragile. In this case, it seems, b might have failed to occur if a physical base of m other than p_1 had occurred. For instance, I might have reached faster for the aspirin if an alternative physical base of m had occurred; in that case, b would not have occurred if b is very fragile. So we can reject (7), and there is no obstacle to the joint truth of (5) and (6).¹⁹ Let us leave the counterfactual relation between m ’s physical bases and b for a moment and consider that between the physical bases and m itself. What would have been the case if p_1 had been replaced by an alternative physical base of m ? It takes an ‘ordinary’ miracle to replace the occurrence of p_1 with that of an alternative physical base. On both the original and the modified account of similarity, m occurs in the closest worlds where p_1 is thus replaced. For removing m from these worlds requires another miracle—a particularly shuntable psychophysical miracle according to the modified account—and it detracts from the match of particular fact with the actual world, where m occurs. Hence the following is true:

(9) If some other physical base of m had occurred instead of p_1 , then m would still have occurred.

$(\sim P_1 \ \& \cup P \ \Box \rightarrow M)$

Claims (5), (6), and (9) are inconsistent with (O_1) . To see this, note first that, by the inference rule we used above, (6) and (9) logically imply

- (10) If p_1 had not occurred, then m might still have occurred.
 $(\sim P_1 \diamond \rightarrow M)$

By another application of the same rule, (10) and (O_1) logically imply (8); as we saw, (8) contradicts (5). In sum, (5) (6), (9), and (O_1) are jointly inconsistent. In other words (5), (6), and (9) logically imply that (O_1) is false.²⁰

Suppose, then, that p_1 is not very fragile. In this case, a situation where m is accompanied by a physical base other than p_1 will require a more drastic physical change than a slightly different firing rate of my c-fibers. Bringing about such a situation may well require some large-scale tempering with my nervous system—perhaps it requires that x-fibers, which are not present in humans, be implanted in my brain. There is no guarantee that the easiest way of implanting them leaves all the outgoing connections intact; it might yield a situation where b , the actual behavioral effect of my c-fiber firing, is absent. This seems plausible even if b is not very fragile. Thus, we get:

- (11) If m had occurred with some other physical base instead of p_1 , then b might not have occurred.
 $(M \ \& \ \sim P_1 \ \& \ \cup P \ \diamond \rightarrow \sim B)$

Further, if m had occurred in the absence of p_1 , some other physical base of m would have occurred instead:

- (12) If m had occurred without p_1 , then some other physical base of m would have occurred instead.
 $(M \ \& \ \sim P_1 \ \square \rightarrow \cup P)$

Claim (12) is true according to the modified account of similarity. If we face the choice between antecedent-worlds of (12) where some other physical base of m occurs and antecedent-worlds of (12) where none does, the former worlds come out closer to the actual world according to the modified Lewisian account since they do not involve a psychophysical miracle.²¹ Now (11) and (12) logically imply

- (13) If m had occurred without p_1 , then b might not have occurred.²²
 $(M \ \& \ \sim P_1 \ \diamond \rightarrow \sim B)$

However, by the definition of the ‘might’ conditional, (13) is true if and only if (O_1) is false. So (O_1) is false.

To summarize, given that b counterfactually depends on p_1 , we can rule out the case where p_1 is very fragile while b is not very fragile. In all other cases, (O_1) is false. Therefore, (O_1) is false. If overdetermination requires the truth of both (O_1)

and (O₂) (non-vacuous or otherwise), dualists can reject (EXCLUSION) and deny that cases of mental causation are cases of overdetermination.

Bennett's own argument against the consistency of dualist mental causation with a denial of (EXCLUSION) is that such a denial would threaten the causal sufficiency of *m* for *b* (see 2008: 289). Whether there is such threat depends on what we understand by causal sufficiency. We saw in the previous section that, according to the dualist account, *m* causes *b* by virtue of *b*'s counterfactual dependence on *m*. This counterfactual dependence is clearly not undermined by the falsity of (O₁). So presumably causal sufficiency is supposed to be something different from causation by virtue of counterfactual dependence. I do not think that dualists are well-advised to give up causation by virtue of counterfactual dependence in favor of some other notion of causation (see Section 6 below). But even if we grant that mental causation involves such a notion for the sake of argument, there need be no conflict with the falsity of (O₁).

I take it that *m*'s being causally sufficient, in a *sui generis* sense, for *b* requires that, with a certain degree of necessity, *m*'s occurrence implies *b*'s occurrence. In other words, *m*'s being causally sufficient for *b* requires that the material conditional 'If *m* occurs, then *b* occurs' hold in a suitable range of possible worlds.²³ In what range of worlds? Not in all worlds, unless there is never sufficient causation between distinct events. Not in all worlds where the actual laws of nature hold, unless we make the sufficient causes so big as to include a cross-section of the effect's past light cone (see Loewer 2007: 253–254). Presumably, the range of worlds comprises those worlds sufficiently similar to the actual world in that various background conditions required for the occurrence of the effect obtain. We have not settled the question of how fragile *m*'s actual physical base is. As things stand, it might well not be very fragile. In this case, the closest worlds where *m* occurs in the absence of its actual physical base (in some of which *b* does not occur if (O₁) is false) would be rather dissimilar to the actual world. It is at least an open question whether these worlds are within the range of worlds where *m*'s occurrence needs to materially imply *b*'s occurrence in order for *m* to be a sufficient cause of *b*.²⁴ Thus, the falsity of (O₁) may well be compatible with *m*'s being a sufficient cause of *b*.

5. Denying Non-Overdetermination

In the previous section I argued that dualists are not committed to the overdetermination of behavioral effects by simultaneous mental and physical causes. If you are not convinced, never mind. Even if dualists *are* committed to the overdetermination of behavioral effects, they can deny that this kind of overdetermination is objectionable. In other words, even if dualists accept (EXCLUSION), they can still deny (NON-OVERDETERMINATION). Or so I shall argue.

Strictly speaking, a failure of the arguments from the previous section need not amount to a concession of *b*'s overdetermination by *m* and *p*₁, since conditions (O₁) and (O₂) were merely presented as necessary conditions for

overdetermination. What does overdetermination require in addition to the truth of (O_1) and (O_2)? There is no agreement about this. ‘Overdetermination’ is a technical term for which different theorists have stipulated different meanings.²⁵ However, insisting on a comparatively demanding notion of overdetermination that requires much besides the truth of (O_1) and (O_2) and thus insisting on the falsity of (NON-OVERDETERMINATION) would be a Pyrrhic victory for dualists. As I shall explain in more detail below, the rationale behind the (EXCLUSION)-cum-(NON-OVERDETERMINATION) part of the exclusion problem is that cases where behavioral events have simultaneous mental and physical causes would be similar to prototypical cases of overdetermination such as deaths by firing squads. Such cases can be specified without appeal to any specific characterization of overdetermination. Thus, the condition that behavioral effects are overdetermined is merely an intermediary step in the presentation of the exclusion problem, which could be omitted in principle.²⁶ Since there is nothing to be gained for dualists by insisting on any specific characterization, I shall concede not merely the truth of (O_1) and (O_2) for the sake of argument, but also the truth of any other conditions required by a reasonable characterization of overdetermination.²⁷

Assume, then, that m and p_1 overdetermine b and that (O_1) and (O_2) are true. What would be objectionable about this? The standard answer is that it would make cases of mental causation like firing squad cases where two shooters simultaneously fire at their victim, who is simultaneously hit by both bullets, each of which would have sufficed to kill him. Such cases exist, the argument goes, but they have a number of features whose presence in all cases of mental causation would be highly implausible. For instance, deaths by firing squad are rare; mental causation, by contrast, is a common phenomenon. (If it exists, that is, but by our assumption of (EFFICACY) it does.) Even if we set the worry about commonness aside, in firing squad cases the two overdetermining events independently bring about the effect. It would be a strange coincidence if this were the case whenever there is mental causation.²⁸

Schematically, this line of reasoning can be put as follows:

- (i) If x is an F , then x is like a prototypical F .
- (ii) Prototypical F s are G s.
- (iii) x is not a G .

(iv) x is not an F .

Premise (i) is ambiguous, however. Similarity comes in different aspects. ‘Being like a prototypical F ’ can mean being like a prototypical F with respect to being an F , or it can mean being like a prototypical F with respect to being a prototypical F . Take an x that is an F , but not a prototypical F . Then on the first reading of ‘being like a prototypical F ’, (i) is true, but (iv) does not follow from (i)–(iii). On the second reading, (i) is false. On either reading, the argument is unsound.

In the case of mental causation, it is claimed that, if mental causation involves overdetermination, cases of mental causation are like firing squad cases, which in turn are prototypical cases of overdetermination. On one reading of this claim, it is true, but all that is said is that cases of mental causation are cases of overdetermination. It does not follow that mental causation has any of the potentially problematic features that firing squads have, such as being rare. (Don't prototypical things have to be common for the kind of thing they are prototypical for? Or at least more common for the kind of thing they are prototypical for than very atypical members of this kind?—No. Perhaps award-winning Alsatians are prototypical mammals, but this is perfectly consistent with their being vastly outnumbered by whales.) On the other reading of the claim that cases of mental causation are like firing squad cases, what is said is that cases of mental causation share the prototypical features of overdetermination that firing squad cases have. This claim, however, can be denied without contradiction.

Thus, it is consistent with mental causation's involving overdetermination that cases of mental causation have little in common with prototypical cases of overdetermination such as firing squad cases. Consistency, of course, is not the same as plausibility. But more can be said for the claim that cases of mental causation are rather dissimilar to prototypical cases of overdetermination. First, on the dualist account of mental causation presented here, it is no coincidence that those behavioral effects that have mental causes also have physical causes, since the two causes are tied together by psychophysical laws, which could not have failed as easily as ordinary laws of nature.²⁹ Second, on this account, the behavioral effect *b* counterfactually depends on its mental cause *m* and also counterfactually depends on the actual physical base of its mental cause, *p*₁. This is compatible with the assumed truth of (O₁) and (O₂). All that is required is, first, that the closest worlds where *m* occurs without *p*₁ (where *b* occurs by (O₁)) be further from actuality than the closest worlds where *p*₁ does not occur (where *b* does not occur, by its counterfactual dependence on *p*₁), and, second, that the closest worlds where *p*₁ occurs without *m* (where *b* occurs by (O₂)) be likewise further from actuality than the closest worlds where *m* does not occur (where *b* does not occur, by its counterfactual dependence on *m*). By contrast, in prototypical cases of overdetermination such as firing squad cases, the overdetermined event does not counterfactually depend on each of the overdetermining events: if one of the shooters had not fired, the victim would still have died because the other shooter would still have fired. The counterfactual dependence of the behavioral effect on both its mental and its physical cause marks out cases of mental causation as very atypical cases of overdetermination. Therefore it should not come as a surprise that mental causation can involve overdetermination in an unobjectionable way, so that (NON-OVERDETERMINATION) can be rejected.

I should perhaps repeat that in this section I have assumed overdetermination merely for the sake of argument. If the argument from the previous section is sound, the dualist has no need to admit that behavioral effects are overdetermined by their mental and physical causes. But should they turn out to be thus overdetermined after all, no harm would befall the dualist.

6. Objections and Replies

First objection

The solution to the exclusion problem presented here draws on counterfactual dependence as a sufficient condition for causation. Yet the exclusion problem is most pressing for a view of causation as *production*, or causal ‘oomph’, which involves continuous spatio-temporal processes and the transfer of energy between cause and effect. Counterfactual dependence is not sufficient for this kind of causation.³⁰

Reply

If this is the case, so much worse for causation as production. Dualists should be allowed to endorse an account of causation that allows them to solve the exclusion problem.³¹ Moreover, the sufficient condition for causation in terms of counterfactuals that we have used is independently plausible.

Second objection

The exclusion problem needs to be formulated in terms of causation as production because our notion of agency is intimately bound up with it (see Kim 2007: 236). When my headache causes me to take an aspirin, this means that my headache *produced* the bodily movement that constitutes my taking of an aspirin.

Reply

As it stands, this is more a gesture than an argument. An advocate of a counterfactual account of causation can respond by claiming that our notion of agency is at least as intimately bound up with the notion of counterfactual dependence, since this allows our minds to *make a difference* to what we do: if I had not had the headache, I would not have taken an aspirin (see Loewer 2007: 255).

*Third objection*³²

There are counterexamples to the sufficiency of counterfactual dependence for causation. A nuclear bomb explodes in the center of our town. The blast first destroys my house; a fraction of a second later, it destroys your house, which is further away from the center. The counterfactual ‘If my house had not been destroyed, then your house would not have been destroyed’ is true according to Lewis. It takes a big miracle to prevent the destruction of my house, but only small miracle (a tiny malfunction in the fuse, say) to prevent the nuclear explosion. Hence, in the closest worlds where my house is not destroyed, the nuclear bomb does not explode and your house is not destroyed either. But clearly the destruction of my house does not cause the destruction of your house.

Reply

The counterfactual that expresses the counterfactual dependence of the destruction of your house on the destruction of my house is evaluated in a back-and-then-forward way here: if my house had not been destroyed, then *the nuclear bomb would not have exploded before and so* your house would not have been destroyed.

We saw in Section 3, first, that we may take counterfactual dependence to suffice for causation only if the relevant counterfactuals are not evaluated in a back-and-then-forward way and, second, that Lewis claims that his truth-conditions plus his account of overall similarity do not yield any true back-and-then-forward evaluations of counterfactuals. So the real target of the objection is the latter claim.

Proponents of Lewis's theory of counterfactuals should respond by claiming that the relevant counterfactual is false because it would merely take a small miracle to prevent the destruction of my house. Thus, the closest worlds where my house is not destroyed match the actual world until just before the time at which my house is destroyed in the actual world. In particular, in these worlds the nuclear bomb still explodes and, later, your house is still destroyed. Why should we say that the miracle that prevents the destruction of my house (given that the nuclear explosion occurs) is a small one? According to Lewis (1986b: 55–56), what distinguishes big miracles from small ones is that the former are spread out more broadly and that they have parts (themselves small miracles) that are varied.³³ Given that the nuclear explosion occurs, certain laws need to be broken throughout a spatial volume around my house for a short while in order to shield my house from the blast. While admittedly this volume has a substantial size, it is not spread out in the sense of being scattered.³⁴ Further, the parts of this miracle are all alike, as they all involve violations of the same laws (namely, whichever laws need to be broken in order to prevent the radiation, heat, and impact of the explosion from reaching my house).

We can, however, uphold the explanation of dualist mental causation presented here even if this defense of Lewis's claim that his similarity account yields no true back-and-then forward evaluations of counterfactuals fails. We merely need to weaken the claim that counterfactual dependence as assessed according to Lewis (or according to our modified Lewisian account) is sufficient for causation to the claim that counterfactual dependence as assessed according to Lewis (or our modified Lewisian account) is sufficient for causation *if* the relevant counterfactual is not evaluated in a back-and-then-forward (or backtracking) way. Our case—more precisely, counterfactual (4)—meets this qualification. We saw that, in the closest worlds where the antecedent of (4) is true, neither *m* nor any physical base of *m* occurs according to the modified account of overall similarity. Whatever the precise definition of a small miracle, it seems that the occurrence of a physical base of *m* can be prevented by a small miracle while holding fixed past history until just before the time at which *m* occurs in the actual world. So the truth of (4) is not due to a back-and-then-forward evaluation. (Since by assumption *b* occurs later than *m*, it is obviously not true due to a backtracking evaluation either.)

Fourth objection

There is still a specific worry about the sufficiency of counterfactual dependence for causation when simultaneous events are concerned. Presumably, the dualist will accept that *m* counterfactually depends on *p*₁ and that *p*₁ counterfactually depends on *m*. Causation is transitive: if *c* causes *e*, and *e* causes *f*, then *c* causes *f*. So *m*

causes itself (assume $c = f = m, e = p_1$); by the same token, p_1 causes itself (see Kistler 2006: 48–49). But this is highly implausible.

Reply

Even if these counterfactual dependences are granted, there are a range of responses available. First, one could weaken the claim that counterfactual dependence between distinct events is sufficient for causation to the claim that counterfactual dependence between distinct *and non-simultaneous* events is sufficient for causation. The weaker claim would still suffice to establish that m causes b , but it would no longer yield the result that m causes p_1 or that p_1 causes m . Second, one could deny the transitivity of causation, as has become increasingly popular among theorists of causation (see, e.g., Hitchcock 2001). Third, one could weaken the transitivity principle so that it requires c , e , and f to be pairwise distinct events. Since no event is distinct from itself, this weakened principle will no longer generate cases of self-causation (see Glynn and Kroedel forthcoming, § 5, n. 26). Fourth, one could bite the bullet and concede the self-causation of events m and p_1 , but insist that it is harmless since they also have other causes (see Lewis 1986a: 212–213). A dualist can pick whichever option best fits her views about causation.

Fifth objection

You assumed on behalf of the dualist that the psychophysical laws have a privileged status in the similarity criteria for worlds. Correspondingly, you assumed that these laws could not have failed as easily as the ordinary laws of nature. Assuming such a special modal status for psychophysical laws is distinctly ad hoc.

Reply

First, assuming a distinct modal status for psychophysical laws might be more congenial to dualism than it initially seems. Dualists hold that the mind is special, so they may well hold that the mind is modally special. More specifically, they may hold that a special modal status of the psychophysical laws has independent epistemological virtues. Perhaps it is easier to imagine electricity without magnetism than to imagine my body without my mind. If so, this could be straightforwardly explained if the physical laws that link magnetism to electricity could have failed more easily than the psychophysical laws that link my mind to my body. Second, even if the assumption that the psychophysical laws have a special modal status is made without independent motivation, it may be worthwhile in order to save mental causation, at least for those independently convinced of the truth of dualism. If astrophysicists are allowed to posit dark matter to save their convictions about gravity, why shouldn't dualists be allowed to posit a special modal status for the psychophysical laws to save their conviction that there is mental causation? Jaegwon Kim, for one, has argued for reductive physicalism from the existence of mental causation (see, e.g., 1998, 2005). If this general kind of argument is acceptable, it should likewise be acceptable for dualists to fine-tune their metaphysics of mind and adopt super-nomological dualism in order to accommodate mental causation.

7. Conclusion

I have argued that dualists can explain mental causation by claiming that behavioral events counterfactually depend on mental events. This counterfactual dependence can in turn be derived if a special modal status of the psychophysical laws is assumed. In response to the exclusion problem, dualists can either deny that behavioral effects are overdetermined by their mental and physical causes or deny that such overdetermination would be objectionable.

Mutatis mutandis, the account of mental causation presented here is available to non-reductive physicalists as well. For instance, the argument from (1)–(3) to (4) from Section 2 would still be sound if the relationship between a mental event and its physical bases were metaphysically necessary, rather than merely modally privileged. In this case, we would not even need to modify Lewis's original account of overall similarity. The dualist account of mental causation presented here recommends itself as a general solution to the problem of mental causation.³⁵

Notes

¹ The position thus characterized is an instance of what Stoljar (2008: 270) calls "traditional dualism".

² This is trivially the case if events *are* property instantiations, as Kim (1976) holds.

³ See Lewis 1979. Lewis assumes determinism when developing his account, and I shall follow him in making this assumption. I ignore his priority (4) (*ibid.*, 472), as it is going to be irrelevant for our purposes.

⁴ A note on the symbolism: A capital letter stands for the proposition that the event referred to by the corresponding lower-case letter occurs. **P** stands for the set of the propositions P_i , which in turn state the occurrence of the physical bases p_i . $\cup \mathbf{P}$ stands for the proposition that at least one member of **P** is true.

⁵ On the validity of the inference rule, see Lewis 1973b: 33. The inference would still be valid if (3) were replaced with the corresponding 'might' conditional (see Lewis 1973c: 433). The explanation given in the main text is rough insofar as 'the closest worlds where such-and-such is true' might not be well-defined; see Lewis 1973b: 20–21. Nonetheless, for simplicity I shall continue sometimes to use this notion.

⁶ Assuming, as seems plausible, that *b* is not overdetermined by (inter alia) further *physical* events.

⁷ See Loewer 2001a: 51–52 for an argument along these lines.

⁸ This is not to say that there are no worlds where the psychophysical laws are violated while the ordinary laws are not. On the contrary, there had better be such worlds, by the dualists' lights at least, if zombies are supposed to be metaphysically possible.

⁹ A number of authors have suggested different modifications of Lewis's account of similarity recently, including Dunn (2011), Kment (2006), and Woodward (2003: 133–145).

¹⁰ For a discussion of (3) in the context of non-reductive physicalism, see Kallestrup 2006: 473.

¹¹ Alternatively, one might define counterfactual dependence so that it also requires the truth of the counterfactual 'If *c* had occurred, then *e* would have occurred'. On the common assumption that the actual world is closer to itself than any other world, this condition is redundant since it is automatically true given that *c* and *e* actually occur. The account of mental causation by List and Menzies (2009) rejects this assumption.

¹² Other accounts of mental causation that invoke counterfactuals include Baker 1993, Horgan 1989, Lepore and Loewer 1987, List and Menzies 2009, Loewer 2007, and Lowe 2003. Kallestrup 2006 and Zhong 2011 critique such accounts as endorsed by non-reductive physicalists. Kroedel 2008 argues for the counterfactual dependence of behavioral events on mental events on the basis of a special modal

status of the psychophysical relation, but does not take into account the details of the Lewisian similarity criteria.

¹³ For discussion of conditionals closely related to (5), see Kallestrup 2006: 476, Loewer 2007: 257, Yablo 1992: 274–280, and Zhong 2011: 138.

¹⁴ The problem, which is due to Malcolm (1968), has been much discussed by Kim (e.g., 1998, 2005, 2007). For a recent overview, see Bennett 2007.

¹⁵ At least they do so on a reading that equates causation in (EFFICACY) with sufficient causation in (COMPLETENESS) and (EXCLUSION), and overdetermination in (EXCLUSION) with systematic overdetermination in (NON-OVERDETERMINATION). I shall not quibble over these possible obstacles to the inconsistency of the five principles.

¹⁶ See Lewis 1986a: 193–200. Aymar (2011: 474–476) disagrees.

¹⁷ The terminology is due to Lewis (1986a).

¹⁸ The inference from (i) $\chi \diamond \rightarrow \phi$ and (ii) $\chi \& \phi \Box \rightarrow \psi$ to (iii) $\chi \diamond \rightarrow \psi$ is valid if the inference from (i) and (ii') $\chi \& \phi \diamond \rightarrow \psi$ to (iii) is valid. For (i) and (ii) logically imply (i) and (ii'): if (i) is true, there is a possible world where both χ and ϕ are true, so (ii) is non-vacuously true if true, in which case (ii') follows from (ii). By the definition of the 'might' conditional, the inference from (i) and (ii') to (iii) is valid if and only if the inference from $\chi \diamond \rightarrow \phi$ and $\sim(\chi \& \phi \Box \rightarrow \sim\psi)$ to $\sim(\chi \Box \rightarrow \sim\psi)$ is, which is valid if and only if the inference from $\chi \diamond \rightarrow \phi$ and $\chi \Box \rightarrow \sim\psi$ to $\chi \& \phi \Box \rightarrow \sim\psi$ is, which in turn is valid if and only if the inference from $\chi \diamond \rightarrow \phi$ and $\chi \Box \rightarrow \psi$ to $\chi \& \phi \Box \rightarrow \psi$ is, which is valid according to Lewis 1973c: 433.

¹⁹ This is not to say that proponents of counterfactual accounts of causation should generally conceive of events as very fragile (see Lewis 1986a: 196–199), but this point does not matter in the present dialectic.

²⁰ Since (5) by itself is consistent with (O₁) (more on this in the following section), using (5) as a premise in an argument against (O₁) does not beg the question.

²¹ For further discussion of (12) in the context of non-reductive physicalism, see Loewer 2001b: 319–320, Bennett 2003: 481–482, and references therein.

²² The inference from (11) and (12) to (13) has the form of an inference from $\chi \& \phi \diamond \rightarrow \psi$ and $\chi \Box \rightarrow \phi$ to $\chi \diamond \rightarrow \psi$. Given the definition of the 'might' conditional, this inference is valid if and only if the inference from $\sim(\chi \& \phi \Box \rightarrow \sim\psi)$ and $\chi \Box \rightarrow \phi$ to $\sim(\chi \Box \rightarrow \sim\psi)$ is, which is valid if and only if the inference from $\chi \Box \rightarrow \sim\psi$ and $\chi \Box \rightarrow \phi$ to $\chi \& \phi \Box \rightarrow \sim\psi$ is, which in turn is valid if and only if the inference from $\chi \Box \rightarrow \psi$ and $\chi \Box \rightarrow \phi$ to $\chi \& \phi \Box \rightarrow \psi$ is. The premises and conclusion of the latter inference are all vacuously true if there is no possible world where χ is true. If there is such a world, premise $\chi \Box \rightarrow \phi$ logically implies $\chi \diamond \rightarrow \phi$, which together with the other premise $\chi \Box \rightarrow \psi$ logically implies the conclusion $\chi \& \phi \Box \rightarrow \psi$ according to Lewis 1973c: 433.

²³ This had better not be a *sufficient* condition for causal sufficiency lest the well-known problems of effects and epiphenomena recur; see Lewis 1973a: 556–557 and Moore 2012: 330–331.

²⁴ Mills (1996: 106) endorses a different necessary condition for *m*'s being a sufficient cause of *b*, viz., the condition that if *m* had not occurred, *then if it had*, *b* would have occurred. This condition is consistent with the falsity of (O₁), however, since the M-worlds that are closest to the \sim M-worlds that are closest to the actual world need not coincide with the \sim P₁-&-M-worlds that are closest to the actual world. The condition invoked by Mills is also discussed in Yablo 1992.

²⁵ For instance, Lewis (1986a: 199–200) holds that overdetermination requires that the overdetermining events have an equally good (or bad) claim to be a cause of the overdetermined event; Bennett (2008: 288) holds that overdetermination requires that both overdetermining events be causally sufficient for the overdetermined event. For a discussion of various alternative proposals, see Carey 2011.

²⁶ Bennett (2007: 327; 2008: 281 n. 3) agrees.

²⁷ Since these conditions are likely to include the condition that both overdetermining events be causes of the overdetermined event, I shall ignore worries about their individual efficacy, but see Schaffer 2003 for discussion.

²⁸ If the two firings in firing squad cases have a common cause, such as someone's command, they might still be considered independent in the sense of involving two distinct causal processes (see Bennett 2008: 287). It is sometimes claimed (e.g., by Zhong (2011: 132 n. 4)) that in the context

of mental causation appeals to overdetermination would be “ad hoc”. Presumably, such appeals are taken to be ad hoc by virtue of overdetermination’s having certain objectionable features such as the ones described.

²⁹ For similar arguments in the context of non-reductive physicalism, see Carey 2011. See also Sider 2003.

³⁰ See Hall 2004 on a distinction of this kind.

³¹ See Bennett 2008: 293. The relation between different theories of causation and dualist accounts of mental causation is also discussed in Lycan 2009: 557–558.

³² Kment (2010: 84, 107 n. 10) credits this objection to Peter Lipton. Woodward (2003: 133–145) raises a similar issue.

³³ Lewis writes that “[a] big miracle consists of many little miracles together, *preferably* not all alike” (1986b: 56, my emphasis). The context of his discussion strongly suggests that the “preferably” qualification can be dropped, however, for the following sentence states that “[w]hat makes the big miracle more of a miracle is [...] that it is divisible into many *and varied* parts, any one of which is on a par with the little miracle” (*ibid.*, my emphasis; see also 1979: 471).

³⁴ By contrast, Lewis’s paradigmatic big miracles, the so-called reconvergence miracles, are spread out in the sense of being scattered (see 1979: 471).

³⁵ For helpful comments and suggestions, I would like to thank Luke Glynn, Nick Haverkamp, Erasmus Mayr, Moritz Schulz, an anonymous referee for *Noûs*, and audiences and workshop participants in Aachen, Berlin, Dubrovnik, Erlangen, and Konstanz.

References

- Aimar, S. 2011. Counterfactuals, Overdetermination and Mental Causation. *Proceedings of the Aristotelian Society*. 111:469–477.
- Baker, L.R. 1993. Metaphysics and Mental Causation. In: *Mental Causation*. J. Heil and A. Mele, editors. Clarendon Press, Oxford.
- Bennett, K. 2003. Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract It. *Noûs*. 37:471–497.
- Bennett, K. 2007. Mental Causation. *Philosophy Compass*. 2:316–337.
- Bennett, K. 2008. Exclusion Again. In: *Being Reduced*. J. Hohwy and J. Kallestrup, editors. Oxford University Press, Oxford.
- Carey, B. 2011. Overdetermination and the Exclusion Problem. *Australasian Journal of Philosophy*. 89:251–262.
- Chalmers, D.J. 1996. *The Conscious Mind*. Oxford University Press, New York.
- Dunn, J. 2011. Fried Eggs, Thermodynamics, and the Special Sciences. *British Journal for the Philosophy of Science*. 62:71–98.
- Glynn, L., and T. Kroedel. (forthcoming). Relativity, Quantum Entanglement, Counterfactuals, and Causation. *British Journal for the Philosophy of Science*.
- Hall, N. 2004. Two Concepts of Causation. In: *Causation and Counterfactuals*. J. Collins, N. Hall, and L.A. Paul, editors. MIT Press, Cambridge, MA.
- Hitchcock, C. 2001. The Intransitivity of Causation Revealed in Equations and Graphs. *Journal of Philosophy*. 98:273–299.
- Horgan, T. 1989. Mental Quausation. *Philosophical Perspectives*. 3:47–76.
- Kallestrup, J. 2006. The Causal Exclusion Argument. *Philosophical Studies*. 131:459–485.
- Kim, J. 1976. Events as Property Exemplifications. In: *Action Theory*. M. Brand and D. Walton, editors. Reidel, Dordrecht.
- Kim, J. 1998. *Mind in a Physical World*. MIT Press, Cambridge, MA.
- Kim, J. 2005. *Physicalism, or Something Near Enough*. Princeton University Press, Princeton.
- Kim, J. 2007. Causation and Mental Causation. In: *Contemporary Debates in Philosophy of Mind*. B.P. McLaughlin and J.D. Cohen, editors. Blackwell, Oxford.
- Kistler, M. 2006. *Causation and Laws of Nature*. Routledge, London.
- Kment, B. 2006. Counterfactuals and Explanation. *Mind*. 115:261–310.

- Kment, B. 2010. Causation: Determination and Difference-Making. *Noûs*. 44:80–111.
- Kroedel, T. 2008. Mental Causation as Multiple Causation. *Philosophical Studies*. 139:125–143.
- Lepore, E., and B. Loewer. 1987. Mind Matters. *Journal of Philosophy*. 84:630–641.
- Lewis, D. 1973a. Causation. *Journal of Philosophy*. 70:556–567.
- Lewis, D. 1973b. *Counterfactuals*. Blackwell, Oxford.
- Lewis, D. 1973c. Counterfactuals and Comparative Possibility. *Journal of Philosophical Logic*. 2:418–446.
- Lewis, D. 1979. Counterfactual Dependence and Time's Arrow. *Noûs*. 13:455–476.
- Lewis, D. 1986a. Postscripts to 'Causation'. In: *Philosophical Papers, Volume II*. Oxford University Press, New York.
- Lewis, D. 1986b. Postscripts to 'Counterfactual Dependence and Time's Arrow'. In: *Philosophical Papers, Volume II*. Oxford University Press, New York.
- Lewis, D. 1994. Reduction of Mind. In: *A Companion to the Philosophy of Mind*. S. Guttenplan, editor. Blackwell, Oxford.
- Lewis, D. 2004. Causation as Influence. In: *Causation and Counterfactuals*. J. Collins, N. Hall, and L.A. Paul, editors. MIT Press, Cambridge, MA.
- List, C., and P. Menzies. 2009. Nonreductive Physicalism and the Limits of the Exclusion Principle. *Journal of Philosophy*. 106:475–502.
- Loewer, B. 2001a. From Physics to Physicalism. In: *Physicalism and Its Discontents*. C. Gillett and B. Loewer, editors. Cambridge University Press, Cambridge.
- Loewer, B. 2001b. Review of J. Kim, *Mind in a Physical World*. *Journal of Philosophy*. 98: 315–324.
- Loewer, B. 2007. Mental Causation, or Something Near Enough. In: *Contemporary Debates in Philosophy of Mind*. B.P. McLaughlin and J.D. Cohen, editors. Blackwell, Oxford.
- Lowe, E.J. 2003. Physical Causal Closure and the Invisibility of Mental Causation. In: *Physicalism and Mental Causation*. S. Walter and H.-D. Heckmann, editors. Imprint Academic, Exeter.
- Lycan, W.G. 2009. Giving Dualism Its Due. *Australasian Journal of Philosophy*. 87:551–563.
- Malcolm, N. 1968. The Conceivability of Mechanism. *Philosophical Review*. 77:45–72.
- Mills, E. 1996. Interactionism and Overdetermination. *American Philosophical Quarterly*. 33:105–117.
- Moore, D. 2012. Causal Exclusion and Dependent Overdetermination. *Erkenntnis*. 76:319–335.
- Schaffer, J. 2003. Overdetermining Causes. *Philosophical Studies*. 114:23–45.
- Sider, T. 2003. What's So Bad About Overdetermination? *Philosophy and Phenomenological Research*. 67:719–726.
- Stoljar, D. 2008. Distinctions in Distinction. In: *Being Reduced*. J. Hohwy and J. Kallestrup, editors. Oxford University Press, Oxford.
- Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, New York.
- Yablo, S. 1992. Mental Causation. *Philosophical Review*. 101:245–280.
- Zhong, L. 2011. Can Counterfactuals Solve the Exclusion Problem? *Philosophy and Phenomenological Research*. 83:129–147.