Synthese (2015) 192:841–858
DOI 10.1007/s11229-014-0600-5

A simple argument for downward causation

Thomas Kroedel

Received: 15 May 2014 / Accepted: 20 October 2014 / Published online: 22 November 2014

Abstract Instances of many supervenient properties have physical effects. In particular, instances of mental properties have physical effects if non-reductive physicalism is true. This follows by a straightforward argument that assumes a counterfactual criterion for causation. The paper presents that argument and discusses several issues that arise from it. In particular, the paper addresses the worry that the argument shows too many supervenient property-instances to have physical effects. The argument is also compared to a similar argument that has been suggested by Lei Zhong and is shown to be superior to the latter.

Keywords Counterfactual conditionals · Downward causation · Mental causation · Non-reductive physicalism · Supervenience

1 Introduction

Many higher-level, supervenient events cause lower-level events. This follows by a straightforward argument from a counterfactual criterion for causation and further plausible assumptions. The argument applies to mental events as conceived of by non-reductive physicalism: it shows that, given non-reductive physicalism, some mental events have physical effects, which wouldn't have occurred if those mental events hadn't occurred. Some authors have invoked counterfactuals in order to show that non-reductive physicalism allows the mind to have physical effects (e.g., Baker 1993; Lepore and Loewer 1987), but in general they haven't attempted to show why the relevant counterfactuals are true (see Kim 1998, p. 71). The argument presented here

T. Kroedel (✉)
Institut für Philosophie, Humboldt-Universität zu Berlin,
Unter den Linden 6, 10099 Berlin, Germany
e-mail: thomas.kroedel@hu-berlin.de

gives a rigorous derivation of those counterfactuals and defends the claim that their truth suffices to draw causal conclusions. A recent argument by Zhong (2011, 2012) also attempts rigorously to show that supervenient mental events have physical effects by drawing on certain counterfactuals; I contend, however, that the argument presented here is superior to Zhong's in several respects.

The plan is as follows. Section 2 presents the argument for downward causation. Section 3 discusses some worries that have been raised about the counterfactual criterion for causation. Section 4 discusses novel objections according to which the argument should be read as an argument against the counterfactual criterion or against other assumptions of the argument rather than as an argument for downward causation; according to these objections, the argument shows too many supervenient property-instances to have physical effects. Section 5 compares the argument to Zhong's argument. Section 6 concludes with a brief discussion of the exclusion problem.

2 The argument for downward causation

The argument for the claim that some supervenient events cause lower-level events is simple, but it requires a bit of stage-setting.

First, a standard definition and a consequence thereof. Say that a set of properties **A** *supervenes* on a set of properties **B** if and only if, necessarily, if anything instantiates some property F in **A** at a given time, then there is a property G in **B** such that that thing instantiates G at that time, and, necessarily, everything that instantiates G at a given time also instantiates F at that time.[1] If a set of properties **A** supervenes on a set of properties **B** in the sense just defined, then the following is true: for each property F in **A** there is a subset $\mathbf{P}_F$ of **B** such that, necessarily, if something instantiates F at a given time, then it instantiates some member of $\mathbf{P}_F$ at that time; and, necessarily, if something instantiates a member of $\mathbf{P}_F$ at a given time, then it instantiates F at that time. Let us call the members of the set $\mathbf{P}_F$ the *realizers* of F. For simplicity, let us leave reference to times and to the things that instantiate the properties in question implicit. Then we can formulate the consequence of **A**'s supervenience on **B** as follows: for each property F in **A**, there is a set of realizers $\mathbf{P}_F$ (with $\mathbf{P}_F \subseteq \mathbf{B}$) such that

(i) necessarily, if F is instantiated, then a realizer of F is instantiated ($\Box[\mathrm{F} \supset \mathsf{U}\mathbf{P}_F]$); and
(ii) necessarily, if a realizer of F is instantiated, then F is instantiated ($\Box[\mathsf{U}\mathbf{P}_F \supset \mathrm{F}]$).

(In the symbolism, a roman capital letter stands for the proposition that the property referred to by the corresponding italicized capital letter is instantiated. For a set of properties **S**, '$\mathsf{U}\mathbf{S}$' stands for the proposition that some member of **S** is instantiated. '$\supset$' is the material conditional.) Turning (i) and (ii) into a strict biconditional, we can

[1] Here and throughout, 'necessarily' expresses metaphysical necessity. The notion of supervenience defined here is standardly called *strong* supervenience when contrasted with other stripes of supervenience (none of which is relevant for our purposes). More precisely, my formulation of the definition follows what McLaughlin calls "Modal-Operator Strong Supervenience" (1995, p. 95). For further discussion, see Kim (1984) and McLaughlin (1995).

express the consequence of **A**'s supervenience on **B** more concisely: for each property F in **A**, there is a set of realizers $\mathbf{P}_F$ (with $\mathbf{P}_F \subseteq \mathbf{B}$) such that

(iii) Necessarily, F is instantiated if and only if a realizer of F is instantiated ($\Box[\mathrm{F} \equiv \cup\mathbf{P}_F]$).

Second, some assumptions. I assume Lewis's (1973b) truth-conditions for counterfactual conditionals. According to these truth-conditions, a counterfactual 'If ϕ were the case, then ψ would be the case' ($\phi \mathrel{\Box\!\!\rightarrow} \psi$) is true if and only if either

(i) there is no possible world where ϕ is true (the case of *vacuous truth*); or
(ii) there is a possible world where ϕ and ψ are true that is closer (i.e., more similar overall) to the actual world than any worlds where ϕ is true while ψ is false.

For the time being, I shall not assume anything about the nature of the relation of closeness or overall similarity between worlds, but there will be more to say about it in Sect. 3.1.

I make two assumptions about the nature of the causal relata. I assume that they are events (or at least that there is one kind of causation that relates events), and that events are property-instances (see Kim 1976). The assumption that the relata of causation are events is very common (see Paul and Hall 2013, p. 6n). The assumption that events are property-instances is made mainly for convenience: it allows for an easy transition between claims about supervenience, which are stated in terms of properties, and claims about causation, which are stated in terms of events. Section 4.2 will revisit the assumption that events are property-instances.

Lastly, I assume that counterfactual dependence is sufficient for causation that is forward in time. In other words, if event e occurs later than event c, and e wouldn't have occurred if c hadn't occurred, then c causes e.[2] While it is commonly made, this is perhaps the most controversial of my assumptions. We shall return to it in Sects. 3 and 4.

We can now state the argument for downward causation. I present it for a special case, namely, the causation of physical events by mental events given non-reductive physicalism about the mind, but we shall see in Sect. 4 that the argument easily generalizes. According to non-reductive physicalism, mental properties are distinct from, and irreducible to, physical properties, but supervene on physical properties (see Kim 2005, pp. 33–34). For the purposes of the argument, only the supervenience claim is relevant. We saw that the instantiation of a supervening property is strictly equivalent to the instantiation of one of its realizers. Let M be a specific mental property. Given that mental properties supervene on physical properties, we have:

(1) Necessarily, M is instantiated if and only if a realizer of M is instantiated. ($\Box[\mathrm{M} \equiv \cup\mathbf{P}_M]$)

[2] Sometimes the counterfactual dependence of e on c is taken to involve not just that e wouldn't have occurred if c hadn't occurred, but also that e would have occurred if c had occurred. Given Lewis's truth-conditions, the second counterfactual is redundant given that (i) c and e actually occur and that (ii) the actual world is closer to itself than any other worlds are. The account of mental causation developed in List and Menzies (2009) rejects claim (ii).

Unless M is instantiated at the last moment of history, some physical properties are instantiated later than M. Plausibly, some of them wouldn't have been instantiated if M's actual realizer hadn't been instantiated. Even more plausibly, some of them wouldn't have been instantiated if none of M's realizers had been instantiated. Let P^* be such a physical property that is instantiated later than M and that wouldn't have been instantiated if none of M's realizers had been instantiated:

(2) If none of M's realizers had been instantiated, then P^* wouldn't have been instantiated. ($\sim \mathbf{UP}_M \mathrel{\Box\!\!\rightarrow} \sim P^*$)

Given Lewis's truth-conditions for counterfactuals, from (1) and (2) it follows logically[3] that the P^*-instance counterfactually depends on the M-instance:

(3) If M hadn't been instantiated, then P^* wouldn't have been instantiated. ($\sim M \mathrel{\Box\!\!\rightarrow} \sim P^*$)

We have assumed that counterfactual dependence is sufficient for causation that is forward in time. Applied to our case, this assumption yields:

(4) If P^* is instantiated later than M, and P^* wouldn't have been instantiated if M hadn't been instantiated, then the instance of M causes the instance of P^*.

We have assumed that

(5) P^* is instantiated later than M.

From (3)–(5) it follows logically that

(6) The instance of M causes the instance of P^*.

It follows, in other words, that there is downward causation of physical events by mental events.

3 Extant worries

The argument for downward causation assumes that counterfactual dependence is sufficient for causation that is forward in time: if event (property-instance) e occurs later than event (property-instance) c, and e wouldn't have occurred if c hadn't occurred, then c causes e. Since this sufficient condition is restricted to putative effects that occur after the putative cause, it obviously doesn't entail any controversial claims about effects that precede their causes or about effects that are simultaneous with their causes.[4] But even our relatively weak sufficient condition for causation is subject to a number of objections. Since these are well-known and replies are available, my discussion here will be brief.

[3] The inference from (1) and (2) to (3) is valid not only given Lewis's truth-conditions for counterfactuals, but in any logic for counterfactuals that creates non-hyperintensional contexts in Williamson's (2006, p. 312) sense. The inferences between counterfactuals that will be made in Sect. 4.2, however, are not valid in all logics that merely satisfy this minimal requirement.

[4] For further discussion of how counterfactual theories of causation might apply to such cases, see Fenton-Glynn and Kroedel (forthcoming).

3.1 Backtracking

The air pressure drops. Later, the barometer reading falls; later still, there is a storm. The drop in air pressure causes both the falling of the barometer reading and the storm, but the falling of the barometer reading doesn't cause the storm. Still, it might be held, the following counterfactual is true:

(7) If the barometer reading hadn't fallen, then there wouldn't have been a storm.

If (7) is indeed true, we have a counterexample to our sufficient condition for causation.

The appropriate response is to deny that (7) is true, at least in our context. More generally, we should combine the sufficient condition for causation in terms of counterfactuals with certain constraints on how these counterfactuals should be evaluated (see Lewis 2004, p 78). If such-and-such an event had happened (or hadn't happened), the past would have been as it actually was until just before that event. If the barometer reading hadn't fallen, the earlier drop in air pressure would still have occurred and thus the storm would still have occurred, rendering (7) false. Technically speaking, for counterfactuals whose antecedents talk about the non-occurrence of specific events that occur in the actual world, we should demand that there be a set of closest worlds where the antecedent is true but where history matches that of the actual world until just before the time at which the event occurs in the actual world. (Similarly, for counterfactuals whose antecedents talk about the occurrence of specific events that don't occur in the actual world, we should demand that there be a set of closest worlds where the antecedent is true but where history matches that of the actual world until just before the time at which the event occurs.)[5]

3.2 Strange relata

At midnight, the bottle shatters owing to external forces. At 11:59 p.m., the bottle had the property of shattering-in-one-minute (call this property *S*+). It is true that

(8) If the bottle hadn't had *S*+ at 11:59 p.m., it wouldn't have shattered at midnight.

But it doesn't seem that the instance of *S*+ causes the bottle's shattering.

The appropriate response is to restrict our sufficient condition for causation to certain kinds of properties. We should allow only instances of properties that are sufficiently (temporally) intrinsic, that is, roughly, properties that are about how things are with the object in question at the time of instantiation (see Lewis 2004, p. 78). Since property *S*+ isn't about how things are with the object at the time of instantiation, the

[5] This suggestion can be finessed in different ways. Lewis (1979) claims that the combination of his truth-conditions for counterfactuals with a similarity ordering of worlds that features big and small violations of (actual) laws of nature and match of particular fact with the actual world rules out 'backtracking' evaluations of counterfactuals whose antecedents talk about specific events. It is somewhat controversial whether Lewis's strategy succeeds (see Kment 2010, p. 84). Alternatives that could be built upon for our purposes that don't involve commitment to Lewis's similarity ordering include the accounts of Maudlin (2007) and Paul and Hall (2013, pp. 47–48).

restricted sufficient condition for causation no longer yields a case of counterfactual dependence without causation.[6]

3.3 Explanatory irrelevance

On the street I bump into a stranger, Albert, who subsequently misses his bus. On the next bus, Albert meets his future wife. They have a child, Berta, that dies 90 years later. It seems true that

(9) If I hadn't bumped into Albert, then Berta wouldn't have died.[7]

But it might seem that my bumping into Albert doesn't cause Berta's death.

The appropriate response is to accept that my bumping into Albert does cause Berta's death and to explain away appearances to the contrary as a pragmatic phenomenon. Counterfactual dependence is sufficient for a kind of causation that is "broad and non-discriminatory" (Lewis 1973a, p. 559). If event *e* counterfactually depends on an earlier event *c*, it follows that *c* is *a* cause of *e*. It doesn't follow that *c* is among those causes of *e* that are explanatorily relevant, and hence worth mentioning, in any given context. (A fortiori, it doesn't follow that *c* is *the* cause of *e*, if the definite article is supposed to single out the most explanatorily relevant one among *e*'s causes.) In most contexts, my bumping into Albert counts as irrelevant for a causal explanation of Berta's death. In those contexts, it would sounds strange to *say* that my bumping into Albert causes Berta's death, but it remains true that it is among the causes of Berta's death.[8]

4 Novel worries

The argument for downward causation shows that instances of mental properties, as conceived of by non-reductive physicalism, have physical effects. As we shall see in this section, it also shows that instances of a number of other supervenient properties have physical effects. On the face of it, it might seem that the argument yields the result that too many instances of supervenient properties have physical effects, which might lead one to reconsider the argument's assumptions, especially the sufficient condition for causation in terms of counterfactual dependence. It will turn out, however, that other responses are available, which leave the argument unscathed.

4.1 Too many supervenient causes

The argument for downward causation can be employed to show that virtually any property that is actually instantiated and that supervenes on physical properties has physical effects. For any such property *F*, it seems, we can find a physical property

[6] A related worry is about omissions. Lots of events counterfactually depend on omissions, but one might not want to accept that they are caused by those omissions. That omissions can't be causes is more controversial than that properties like *S*+ can't be causes. If one wants to rule out omissions as causes, one could pursue a similar strategy, however: restrict our sufficient condition for causation to instances of 'positive' properties and disallow instances of 'negative' properties. For further discussion, see Lewis (1986b, pp. 189–193).

[7] The example is a variation of an example from Lewis (1986b, p. 184).

[8] For a recent elaboration of this approach, see Swanson (2010).

*P** that is instantiated later than *F* and that wouldn't have been instantiated if none of *F*'s realizers had been instantiated; it follows by the argument that the instance of *F* causes the instance of *P**. Isn't this tantamount to causal overkill?

Before addressing this rhetorical question, a clarification. It does *not* follow from the argument for downward causation that the instances of any property that is necessitated by a property with certain physical effects inherit all those physical effects. Suppose that an instance of a physical property *P** counterfactually depends on, and hence is caused by, an earlier instance of property *F*, which in turn necessitates the instantiation of property *H*. These suppositions do *not* entail that the instance of *P** counterfactually depends on, and hence is caused by, the instance of *H*, for the inference from $\sim F \mathbin{\Box\!\!\rightarrow} \sim P^*$ and $\Box[F \supset H]$ (contrapositively, $\Box[\sim H \supset \sim F]$ to $\sim H \mathbin{\Box\!\!\rightarrow} \sim P^*$) is invalid (see Lewis 1973b, p. 32). So the worry can't be that, by the argument for downward causation, every higher-level property-instance takes on all the effects of any lower-level property-instance that necessitates it. But of course it is compatible with the argument for downward causation that this sometimes happens. And even if the effects of the high-level property-instance are different from those of the lower-level property-instance, the argument might seem to yield too many higher-level causes.

Higher-level causes are not in general objectionable, however. Assume, as many do, that moral and aesthetic properties supervene on physical properties. Then the argument for downward causation yields that the instances of moral and aesthetic properties have some physical effects, for the absence of all physical realizers of a moral or aesthetic property would have made a difference to the physical future. Sometimes the argument can even be employed to show that instances of moral and aesthetic properties have certain specific effects. By the supervenience of aesthetic properties on physical properties, beauty has certain physical realizers. If Helen of Troy hadn't instantiated any of those realizers while at Sparta, the arrowhead wouldn't have moved towards Achilles' heel some 9 years later. Hence the arrowhead's movement counterfactually depends on, and is caused by, the instance of beauty. Cases like this are interesting corollaries of the argument for downward causation rather than problems for it.

This isn't to say all cases of causation by supervenient property-instances that the argument for downward causation yields should be welcomed. Some supervenient properties might be generally ill-suited to enter into causal relations. We should restrict the sufficiency of counterfactual dependence for causation such that instances of those properties are no longer allowed. As was argued in Sect. 3.2, some restrictions of this sort need to be imposed anyway, so this manoeuvre wouldn't be ad hoc (or at least no more ad hoc than the original restriction). Besides, other accounts of causation might have to do the same if the instances of certain properties are generally ill-suited to cause (or be caused by) anything, so the argument for downward causation faces no special difficulty. For instance, everything trivially counterfactually depends on the instances of properties that everything necessarily instantiates (e.g., the property of being either round or not round). It would follow from the sufficiency of counterfactual dependence for forward-in-time causation that all later physical events are caused by those instances unless we disallow such properties. But it is also the case that everything trivially necessitates, and hence is nomologically sufficient for, instances of properties that everything necessarily instantiates. So at least simple accounts of causation in terms of nomological sufficiency need to impose the same restriction.

4.2 Overlapping realizers

I hold an aluminium ladder against a power line and subsequently get electrocuted.[9] Being made of aluminium, the ladder is an electrical conductor. Conductivity supervenes on physical properties and can be realized in different ways. If the ladder hadn't instantiated any realizer of conductivity, I wouldn't have been electrocuted. It follows from the argument for downward causation that that the instance of conductivity causes my electrocution. So far, so good. But being made of aluminium, the ladder is also opaque. Opacity too supervenes on physical properties and can be realized in different ways. The realizers of opacity are closely related to the realizers of conductivity. Almost all conductors are opaque. Some conductors are transparent, but they aren't used to make ladders. Thus, it seems that if the ladder hadn't instantiated any realizer of opacity, I wouldn't have been electrocuted either. It follows from the argument for downward causation that that the instance of opacity causes my electrocution. That, however, does not seem very plausible.[10]

Let us formulate the argument for this implausible conclusion along the lines of the argument for downward causation by using the following abbreviations:

C: being an electrical conductor
O: being opaque
E: being electrocuted

[Obviously, in the example the object that instantiates E (i.e., myself) is distinct from the object that instantiates C and O (the ladder). In the original argument, property P^* might or might not be instantiated by the same object as M.] By the supervenience of opacity, we have:

(1-O) Necessarily, opacity is instantiated if and only if a realizer of opacity is instantiated. ($\Box[\mathrm{O} \equiv \cup\mathbf{P}_O]$)

The close relation between the opacity-realizers and the conductivity-realizers seems to give us:

(2-O) If no opacity-realizer had been instantiated, then I wouldn't have been electrocuted. ($\sim\cup\mathbf{P}_O \,\Box\!\!\rightarrow \sim\mathrm{E}$)

From (1-O) and (2-O) it follows logically that

(3-O) If opacity hadn't been instantiated, then I wouldn't have been electrocuted. ($\sim\mathrm{O} \,\Box\!\!\rightarrow \sim\mathrm{E}$)

By the sufficiency of counterfactual dependence for (forward-in-time) causation, from (3-O) we get:

(4-O) The opacity-instance causes my electrocution.

[9] The example is due to Jackson and Pettit (1990), who use it in a different context.

[10] If artefacts such as ladders have their origin essentially, as Kripke (1980) holds, the ladder could not have been made of a different material from the one it is actually made of. If that is the case, the problem can be reformulated by taking the relevant events to be instances of the property of containing a ladder that is made of such-and-such a material by the spatial region that is occupied by the ladder.

If one finds the conclusion that the opacity-instance causes my electrocution too implausible, different responses are available.

The first response is to follow the strategy for dealing with backtracking counterfactuals that was discussed in Sect. 3.1 and to deny the counterfactual that expresses the counterfactual dependence of my electrocution on the opacity-instance, (3-O). Since (3-O) follows logically from (1-O) and (2-O), denying (3-O) requires denying either (1-O) or (2-O). (1-O) seems unassailable, so one has to deny (2-O). To see what denying (2-O) amounts to, consider the following argument *for* (2-O):

(5-O) If no opacity-realizer had been instantiated, then no conductivity-realizer would have been instantiated. ($\sim\mathbf{UP}_O \mathrel{\Box\!\!\rightarrow} \sim\mathbf{UP}_C$)

(6-O) If no opacity-realizer had been instantiated and no conductivity-realizer had been instantiated, then I wouldn't have been electrocuted. ($\sim\mathbf{UP}_O \,\&\, \sim\mathbf{UP}_C \,\Box \rightarrow \sim$E)

(2-O) If no opacity-realizer had been instantiated, then I wouldn't have been electrocuted. ($\sim\mathbf{UP}_O \mathrel{\Box\!\!\rightarrow} \sim$E)

The argument from (5-O) and (6-O) to (2-O) is an instance of the principle of restricted transitivity, which is valid (see Lewis 1973b, p. 35). Given the validity of the argument, denying (2-O) requires denying either (5-O) or (6-O). (6-O) looks very plausible. If all conductivity-realizers had been absent, I certainly wouldn't have been electrocuted. It would be strange if the additional absence of all opacity-realizers were to bring back my electrocution.[11]

So denying (3-O) ultimately requires denying (5-O). Denying (5-O) comes at a price, however. It is natural to think that if the ladder hadn't instantiated any opacity-realizer, then it would have been made of some middle-of-the-road transparent material (glass or transparent plastic, say), which wouldn't have been conductive. This natural thought must be given up if (5-O) is denied. Instead, worlds where the ladder is made of some exotic transparent conductive material have to be taken to be just as close to the actual world as worlds where the ladder is made of some middle-of-the-road transparent non-conductive material.

The second response is to drop the assumption that events are property-instances and to adopt a more coarse-grained conception of events. According to this conception, events are space–time regions that involve the instantiation of properties, but do so to different modal degrees (see Lewis 1986a). In our case, it seems promising, at least prima facie, to proceed as follows: Let o be the event of the ladder's being opaque. Event o should essentially involve the instantiation of opacity by the ladder. Otherwise we would have to say that the ladder's being opaque could have occurred while the ladder wasn't opaque, which seems strange.[12] Thus, we have:

(9) Necessarily, if o occurs, then opacity is instantiated. ($\Box[\mathrm{Oc}(o) \supset \mathrm{O}]$)

[11] Which is not to say that (6-O) follows logically from $\sim\mathbf{UP}_C \mathrel{\Box\!\!\rightarrow} \sim$E, for it doesn't (see Lewis 1973b, p. 31).

[12] At least in our case. In general, properties that feature in the description of an event don't have to be essential to that event, however. See Lewis (1986a, pp. 247–254) for discussion.

(In this addition to the symbolism, 'Oc(x)' stands for the proposition that event x occurs.) We should also demand that if o hadn't occurred, then the ladder wouldn't have been opaque:

(10) If o hadn't occurred, then opacity wouldn't have been instantiated. ($\sim\mathrm{Oc}(o) \mathrel{\Box\!\!\rightarrow} \sim\mathrm{O}$)

Lastly, we should demand that it's not the case that if o hadn't occurred, then the ladder wouldn't have been conductive:

(11) It's not the case that if o hadn't occurred, then conductivity wouldn't have been instantiated. ($\sim[\sim\mathrm{Oc}(o) \mathrel{\Box\!\!\rightarrow} \sim\mathrm{C}]$)

Claim (11) allows us to deny that event o causes my electrocution: if o hadn't occurred, I might still have been electrocuted because the ladder might still have been conductive.

The trouble with this response is that it is at least as problematic as the previous response, which sought to deny the claim that my electrocution counterfactually depends on the opacity-instance. By contraposition, claim (9) is equivalent to the claim that

(12) Necessarily, if opacity isn't instantiated, then o doesn't occur. ($\Box[\sim\mathrm{O} \supset \sim\mathrm{Oc}(o)]$)

Strict conditionals logically imply the corresponding counterfactual conditionals,[13] so from (12) we get:

(13) If opacity hadn't been instantiated, then o wouldn't have occurred. ($\sim\mathrm{O} \mathrel{\Box\!\!\rightarrow} \sim\mathrm{Oc}(o)$)

Claims (10), (11), and (13) logically imply:[14]

(14) It's not the case that if opacity hadn't been instantiated, then conductivity wouldn't have been instantiated. ($\sim[\sim\mathrm{O} \mathrel{\Box\!\!\rightarrow} \sim\mathrm{C}]$)

By our earlier assumption (1-O), the instantiation of opacity is strictly equivalent to the instantiation of a realizer of opacity. Similarly, the instantiation of conductivity is strictly equivalent to the instantiation of a realizer of conductivity:

(1-C) Necessarily, conductivity is instantiated if and only if a realizer of conductivity is instantiated. ($\Box[\mathrm{C} \equiv \cup\mathbf{P}_C]$)

Given (1-O) and (1-C), (14) is equivalent to:

(15) It's not the case that if no opacity-realizer had been instantiated, then no conductivity-realizer would have been instantiated. ($\sim[\sim\cup\mathbf{P}_O \mathrel{\Box\!\!\rightarrow} \sim\cup\mathbf{P}_C]$)

[13] Friends of false counterpossibles, that is, false counterfactuals with impossible antecedents, will disagree. For a given (allegedly) false counterpossible, they cannot accept that it is logically implied by the corresponding strict conditional, which is trivially true owing to the impossible antecedent. But friends of false counterpossibles can still accept the weaker claim that strict conditionals with possible antecedents logically imply the corresponding counterfactuals. This weaker claim is all that is needed for the present derivation.

[14] The inference has the form of an inference from $\phi \mathrel{\Box\!\!\rightarrow} \chi$, $\chi \mathrel{\Box\!\!\rightarrow} \phi$, and $\sim[\chi \mathrel{\Box\!\!\rightarrow} \psi]$ to $\sim[\phi \mathrel{\Box\!\!\rightarrow} \psi]$, which is valid if and only if the inference from $\phi \mathrel{\Box\!\!\rightarrow} \chi$, $\chi \mathrel{\Box\!\!\rightarrow} \phi$, and $\phi \mathrel{\Box\!\!\rightarrow} \psi$ to $\chi \mathrel{\Box\!\!\rightarrow} \psi$ is, which in turn is valid according to Lewis (1973b, p. 33).

Claim (15) is the negation of claim (5-O). We saw earlier that denying (5-O) is problematic because it requires giving up the natural thought that the ladder would have been made of some middle-of-the-road transparent material if it hadn't instantiated any opacity-realizer. Thus, the response that adopts more coarse-grained events as the causal relata is at least as costly as the first response.

The third response is to refine the sufficient condition for causation by taking into account counterfactuals with more complex antecedents. If opacity hadn't been instantiated, then I wouldn't have been electrocuted. But if opacity hadn't been instantiated *while conductivity had still been instantiated*, then I would still have been electrocuted. On the other hand, if conductivity hadn't been instantiated *while opacity had still been instantiated*, then I wouldn't have been electrocuted. More generally, the idea is that one event causes another if the first even makes a difference to the occurrence of the second even if we hold the occurrence of certain other events fixed.

We can, however, almost always find causal intermediaries between the putative cause and the putative effect that 'screen off' the latter from the former. If those are the relevant events whose occurrence is to be held fixed, we can almost never apply our envisaged new sufficient condition for causation. So what are the relevant events that are supposed to be held fixed?

So-called causal modelling accounts of causation provide the resources for answering this question; I can only sketch the answer here, however.[15] Start with the notion of a *causal model*, which consists of a set of *variables* and a set of *structural equations*. The variables represent the instantiation of certain properties, and the structural equations encode counterfactuals about those instantiations. Specifically, for each variable, there is a structural equation that encodes counterfactuals about what the value of that variable would have been given the values of other variables. The structural equations must not contain any redundant elements: if the value of a certain variable makes no difference to the value of the target variable over and above other variables, then that variable must not appear in the structural equation for the target variable. The set of structural equations determines a *causal graph*, in which arrows are drawn from one variable to another if and only if the first variable appears in the structural equation of the second variable. A *path* is a sequence of variables that can be traced by following the direction of arrowheads in the causal graph.

For the ladder example, let us choose the following variables: Let *Con*, *Op*, and *El* be binary variables (i.e., variables that can assume either value 0 or value 1) that represent, respectively, the instantiation of conductivity, opacity, and being electrocuted. For these variables, value 1 represents that the property in question is instantiated; value 0 represents that it isn't. Let *Phys* be a variable that can assume values 0, 1, 2, or 3. Value 0 represents that the ladder instantiates a realizer of both conductivity and opacity (e.g., aluminium). Value 1 represents that the ladder instantiates a realizer of conductivity that isn't also a realizer of opacity. Value 2 represents that the ladder instantiates a realizer of opacity that isn't also a realizer of conductivity (e.g., wood). Value 3 represents that the ladder instantiates neither a realizer of conductivity nor a realizer of opacity (e.g., transparent plastic). For our purposes we needn't write down

[15] I shall loosely follow Hitchcock's (2001, 2007) causal modelling framework for token causation.

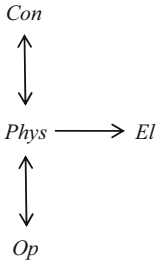

Fig. 1 The graph of a causal model for the electrocution case

the structural equations in details; what matters is the causal graph. The graph will look like this: There is an arrow to *El* from *Phys*, but from no other variable, since the value of *Phys* makes a difference to the value of *El* (for instance, the counterfactual '$Phys = 2 \,\Box\!\!\rightarrow El = 0$' is true), but no other variable makes a difference to *El* over and above the difference made by *Phys*.[16] There is an arrow from *Phys* to *Con*. There is also an arrow from *Con* to *Phys*, for we may assume that '$Con = 0 \,\Box\!\!\rightarrow Phys = 2$' is true (if the ladder hadn't been conductive, then it would have been made of wood rather than transparent plastic).[17] There is an arrow from *Phys* to *Op*. There is also an arrow from *Op* to *Phys*, for we may assume that '$Op = 0 \,\Box\!\!\rightarrow Phys = 3$' is true: we saw in the discussion of the first response that it is a natural thought that if the ladder hadn't been opaque, then it would have been made of transparent plastic rather than some transparent conductor. In sum, we get the graph depicted in Fig. 1.

In our model, both the value of *Con* and the value of *Op* make a difference to the value of *El*, because both '$Con = 0 \,\Box\!\!\rightarrow El = 0$' and '$Op = 0 \,\Box\!\!\rightarrow El = 0$' are true. '$Con = 0 \,\&\, Op = 1 \,\Box\!\!\rightarrow El = 0$' is also true, while '$Con = 1 \,\&\, Op = 0 \,\Box\!\!\rightarrow El = 0$' is false. So far, these are merely technical reformulations of the counterfactuals we assessed before introducing causal modelling: both the conductivity-instance and the opacity-instance make a difference with respect to my electrocution, but the conductivity-instance makes a difference even when we hold the opacity-instance fixed, while the opacity-instance doesn't make a difference when we hold the conductivity-instance fixed. Our causal graph suggests a rationale for taking these counterfactuals to imply that the conductivity-instance causes the electrocution while not taking them to imply that the opacity-instance causes the electrocution: *Op* is off the path from *Con* to *El*, and *Con* is off the path from *Op* to *El*.[18] What matters is whether a variable makes a difference to another if we hold the off-path variables fixed at their actual values; if it does, we may infer a causal relation.

[16] Difference-making is supposed to captured by counterfactuals that are non-vacuously true. The fact that, say, '$Phys = 0 \,\&\, Con = 0 \,\Box\!\!\rightarrow El = 0$' is vacuously true while '$Phys = 0 \,\Box\!\!\rightarrow El = 0$' is false isn't a reason for including *Con* in the equation for *El*.

[17] If that assumption seems implausible, we could still justify the inclusion of *Con* in the equation for *Phys*, and hence the arrow from *Con* to *Phys* in the causal graph, from the truth of the counterfactual '$Con = 0 \,\Box\!\!\rightarrow Phys = 2 \vee Phys = 3$'.

[18] Strictly speaking, there are paths from *Con* to *El* that include *Op*, e.g., the sequence $Con \rightarrow Phys \rightarrow Op \rightarrow Phys \rightarrow El$. Thus, we should understand a path as a non-cyclic sequence here, that is, a sequence that doesn't include any variable more than once.

Fig. 2 The graph of a simple causal model for the electrocution case

More generally, we can formulate our new sufficient condition for causation as follows: Let X and Y be binary variables in an appropriate causal model, where '$Y = 1$' represents a property-instance that occurs later than the property-instance represented by '$X = 1$'. If '$X = 0$ & FIX $\Box\!\!\rightarrow Y = 0$' is true in that model, where FIX stands for the claim that all variables that are not on any (non-cyclic) path from X to Y are held fixed at their actual values, then the property-instance represented by '$X = 1$' causes the property-instance represented by '$Y = 1$'.[19]

The new sufficient condition for causation no longer yields the verdict that the opacity-instance causes my electrocution. It has two disadvantages, however. First, the sufficient condition involves the notion of an *appropriate* causal model. Indeed, in order to solve our problem, it has to do so. If the condition merely said that difference-making while holding off-paths variables fixed in *some* causal model sufficed for causation, we couldn't avoid the implausible conclusion that the opacity-instance causes the electrocution. For instance, take a simple model, including merely the variables *Op* and *El*, whose causal graph is depicted in Fig. 2. In the simple model, there are no variables that are off the path from *Op* to *El*. Hence it is trivially true that the value of *Op* makes a difference with respect to the value of *El* if all off-path variables are held fixed at their actual values.[20] If the simple model is deemed appropriate, it follows that the property-instance represented by '$Op = 1$' causes the property-instance represented by '$El = 1$' after all. If the simple model is deemed inappropriate, some principled reason should be given for why it is inappropriate. This might not be impossible, but it complicates the task of giving a new sufficient condition for causation in terms of causal models.[21] Second, our original causal model is unorthodox in several respects, which might in turn be taken to indicate that it is inappropriate: the causal graph of our model is cyclic owing to the presence of double-arrows; the variable *Phys* is multi-valued while the remaining variables are binary; moreover, the values of *Phys* necessitate the values of other variables (viz., *Con* and *Op*).[22]

If the responses discussed so far all seem unsatisfactory, we have two more options, which are more radical. The fourth response is to deny that counterfactual dependence

[19] While the antecedent is merely supposed to be sufficient, not necessary, for causation, its falsity might sometimes be taken to be a defeasible reason for denying, or at least doubting, the existence of a causal relation. This might explain why some doubt that an overdetermining event causes the overdetermined event: an overdetermining event makes no difference to the occurrence of the overdetermined event if we hold the occurrence of the other overdetermining event, which is off-path, fixed.

[20] Similarly for a model that includes *Con* in addition to *Op* and *El*, but doesn't contain *Phys*. In that model, the value of *Op* makes a difference to the value of *El*, but this difference is nothing over and above the difference made by the value of *Con*. So *Con* is on the path from *Op* to *El*, and again it is trivially true that the value of *Op* makes a difference with respect to the value of *El* if all off-path variables are held fixed at their actual values.

[21] See Halpern and Hitchcock (2010) for discussion of the notion of an appropriate causal model.

[22] For a defence of these unorthodoxies for the purposes of modelling mental causation, see Kroedel (ms).

is sufficient for causation without attempting to replace it with a modified sufficient condition (such as the sufficient condition in terms of causal modelling). The fifth response is to maintain the original sufficient condition and accept that the opacity-instance causes the electrocution. Denying that counterfactual dependence is sufficient for causation is simple. But so is the idea that what makes a difference is a cause. It seems premature to give that idea up unless all alternatives turn out to be untenable. The other radical option, viz., accepting that the opacity-instance causes the electrocution, might initially seem like excessive bullet-biting. But a closer looks reveals it to be not so unattractive. If we choose that option, we can hold on to our original simple and elegant sufficient condition for causation. We will have to accept the result that the opacity-instance causes the electrocution, but perhaps this result isn't so implausible after all. It is by virtue of the intimate relation between the realizers of conductivity and the realizers of opacity that the electrocution counterfactually depends on the opacity-instance. We might still hesitate to call the opacity-instance a cause of the electrocution, but perhaps we hesitate because the opacity-instance is a cause that has little explanatory relevance in our context, not because it isn't a cause at all.[23] (Recall the example of my bumping into Albert as an explanatorily irrelevant cause of Berta's death, from Sect. 3.3.) But this problem, if it is a problem, doesn't affect the paradigm application of the argument for downward causation, which shows instances of mental properties as conceived of by non-reductive physicalism to have physical effects, for such mental causes do typically count as explanatorily relevant.

5 Comparison with Zhong's argument

Let us now compare my argument to Zhong's. Zhong argues as follows.[24] Assume non-reductive physicalism. Assume further that an instance of a mental property M causes an instance of a mental property M^* that is realized by a physical property P^*. By the realization of M^* by P^*, that P^* is instantiated entails that M^* is instantiated ($\Box[P^* \supset M^*]$). Contrapositively, that M^* isn't instantiated entails that P^* isn't instantiated ($\Box[\sim M^* \supset \sim P^*]$). Thus, the P^*-instance counterfactually depends on whatever the M^*-instance counterfactually depends on, since $\sim X \Box\!\!\rightarrow \sim P^*$ follows logically

[23] It might seem promising to apply Swanson's (2010) account of the context-sensitivity of causal talk to our case. Unfortunately, there are some prima facie difficulties with this application. Swanson appeals to the principle that when ascribing causal responsibility for a given effect to a causal path, one should use good representatives of that path (see *ibid*., p. 225). One can't use this principle to show that the conductivity-instance is a better representative of a path that contains both the conductivity-instance and the opacity-instance than the opacity-instance is, for both by Swanson's definition and by the causal modelling definition the two property-instances are on different paths. (They are also in different causal clusters in Swanson's (*ibid*., p. 237) sense, because they have different effects; for instance, the opacity-instance causes the ladder to cast a shadow, but the conductivity-instance doesn't.) Perhaps it could be shown that the opacity-instance is a poor representative of a path that contains it but doesn't contain the conductivity-instance. But showing that wouldn't be straightforward either, since one of Swanson's principal criteria for an event's being a *good* representative, the effect's counterfactually depending on the representative, does apply to the opacity-instance and the electrocution.

[24] See Zhong (2011, pp. 141–143; 2012, pp. 80–81). The 2012 version of the argument, which Zhong prefers (see 2012, p. 81n) and which I follow here, differs in the details of clause (ii) below. These details are inessential for the purposes of our discussion, however.

from $\sim X \Box\!\!\rightarrow \sim M^*$ and $\Box[\sim M^* \supset \sim P^*]$.[25] Now if the *M*-instance causes the *M**-instance, then either

(i) the *M**-instance counterfactually depends on the *M*-instance; or
(ii) there is an intermediary, namely an instance of a mental property M' which is caused by the *M*-instance and on which the *M**-instance counterfactually depends.

In case (i), it follows that the *P**-instance counterfactually depends on the *M*-instance; hence the *M*-instance causes the *P**-instance. In case (ii), it follows that the *P**-instance counterfactually depends on the M'-instance; hence the M'-instance causes the *P**-instance; hence, by the transitivity of causation, the *M*-instance causes the *P**-instance. In sum, if some instances of mental properties cause instances of other mental properties, then they also cause instances of the realizers of these mental properties.

Zhong's conclusion is weaker than mine. He concludes that a mental property-instance causes the instance of the realizer of another mental property *if* the first mental property-instance causes the second mental property-instance. I conclude that some mental property-instances cause physical property-instances *tout court*. That some mental property-instances cause other mental property-instances isn't very controversial, however,[26] so the fact that Zhong's conclusion is a conditional one while mine isn't doesn't make for a substantial difference between our arguments. Zhong's argument is also more specific than mine, however. My argument can easily be generalized to supervenient properties besides mental properties as conceived of by non-reductive physicalism. All that is required for this generalization is that the absence of all realizers of the property in question would have made a difference to the physical future. In order to generalize Zhong's argument to other supervenient properties, we would first have to identify a future instance of another supervenient property that is caused by the instance of the target supervenient property. It might not be straightforward to find such a future instance. We saw in the previous section that the ease with which my argument can be generalized is a mixed blessing. It might therefore be taken to be an advantage of Zhong's argument that it doesn't generalize so easily. This advantage, however, is outweighed by a number of disadvantages.

Zhong's assumptions about causation are stronger than mine. He assumes that counterfactual dependence is sufficient for causation and that counterfactual dependence—or counterfactual dependence via a caused intermediary—is necessary for causation. I merely assume that counterfactual dependence is sufficient for causation. (Strictly speaking, I assume even less: that counterfactual dependence between property-instances is sufficient for causation that is forward in time. But Zhong could do so as well without jeopardizing the validity of his argument, so we are on a par here.) In spite of the worries we discussed in previous sections, the sufficiency of counterfactual dependence for causation is very plausible. The necessity of counterfactual dependence—or counterfactual dependence via a caused intermediary—isn't very

[25] This valid inference should not be confused with the similar but invalid inference that was discussed in Sect. 4.1.

[26] Proponents of the so-called autonomy approach such as Gibbons (2006) accept mental-to-mental causation while denying mental-to-physical causation.

plausible; witness cases of symmetrical overdetermination and late pre-emption (see Lewis 1986b, pp. 193–212 for discussion). Zhong assumes that causation is transitive, which is controversial; witness so-called switching cases (see McDermott 1995, p. 532). I do not assume transitivity. Perhaps Zhong's argument would still be valid if the controversial assumptions were appropriately weakened. He could assume, for instance, that counterfactual dependence—or counterfactual dependence via a caused intermediary—is necessary for causation *in the absence of redundancy* and that causation is transitive *in standard cases*. The weakened assumptions would be less controversial. But controversy might arise over whether they can be applied in particular cases. If we can give them up completely, so much the better.

Zhong's argument is open to an objection to which my argument is immune. Christensen and Kallestrup (2012) object to Zhong's argument as follows. The necessitation of M^* by its realizer P^*, which is required to establish that the P^*-instance counterfactually depends on the M-instance or the M'-instance, holds only if P^* is a "total realizer" of M^* (*ibid*., p. 515). That is, P^* has to be a conjunctive property that includes various "background properties" such as "properties pertaining to pertinent causal laws of nature" besides its "core realizer" properties, which are more narrowly circumscribed (*ibid*., p. 514). The background properties, however, are not themselves "causal properties" that could feature as causes or effects (*ibid*., p. 515). Given that P^* includes those background properties, the claim that the instance of P^* is an effect becomes problematic. Moreover, the background properties are shared between P^* and the actual realizer of M. Thus, M and P^* are no longer sufficiently distinct to be causally related.

Whatever the success of Christensen and Kallestrup's objection against Zhong's argument, it does not touch mine.[27] Granted, I would have to restrict the sufficient condition for causation to causal properties if the objection were sound, for otherwise background properties would yield counterexamples. (As we saw in previous sections, we need to impose some restriction along these lines anyway.) Granted, the realizers of M that feature in (1) and (2) would have to be read as total realizers if the objection were sound, for otherwise the instantiation of M would no longer be strictly equivalent to the instantiation of one of its realizers, as (1) claims. But these concessions wouldn't threaten the causal relation between our M-instance and our P^*-instance. Zhong's argument is open to the Christensen–Kallestrup objection because there is a realizer (viz., the realizer of M^*) whose instance is claimed to be an effect. In my argument no realizers need to have instances that are causes or effects. The set of M's realizers [as represented by proposition $\cup\mathbf{P}_M$ in (1) and (2)] is merely a logical intermediary, not a causal one. And our property P^* can be as causal as one likes, since it need not realize anything.

What is the role of the actual realizer of M (call it P)? One might object that I can't avoid treating at least the instance of P as a causal intermediary, since it follows from M's being necessitated by P that P wouldn't have been instantiated if M hadn't been instantiated, wherefore the M-instance causes the P-instance. I have to concede only the first half of this reasoning, however. It does follow that the P-instance counterfactually

[27] Zhong addresses the objection by Christensen and Kallestrup in Zhong (forthcoming).

depends on the *M*-instance. But the sufficient condition for causation I have used remains silent on whether or not the *M*-instance causes the *P*-instance, because the two are simultaneous. Further, given that we restrict the sufficient condition for causation to causal properties, it would remain silent on whether or not the *P*-instance causes the *P**-instance should it turn out that our *P**-instance counterfactually depends on the *P*-instance while *P* isn't a causal property. So it neither follows that the *P*-instance is an effect of the *M*-instance nor that the *P*-instance is a cause of the *P**-instance. Nonetheless it still follows that the *M*-instance is a cause of the *P**-instance.

6 Conclusion

I have argued that instances of supervenient properties can easily be shown to have physical effects. In particular, non-reductive physicalism, together with a counterfactual criterion for causation, straightforwardly yields the causation of physical property-instances by mental property-instances. The argument can be employed in different ways, depending on how serious one takes the so-called exclusion problem to be.[28] One could read the argument as a *reductio* of non-reductive physicalism: Presumably the physical effect has a physical cause that is simultaneous with its mental cause. Thus, the physical effect is overdetermined. But the physical effects of mental causes aren't overdetermined. Contradiction! Non-reductive physicalism has to go. Alternatively, my argument can be read in favour of non-reductive physicalism: It brings the good news that non-reductive physicalism allows the mental to have physical effects. Granted, a physical effect with a mental cause presumably has a physical cause simultaneous with its mental cause. Depending on what we mean by overdetermination, we might or might not decide to call the physical effect overdetermined. But even if we call it overdetermined, there is nothing objectionable about the situation. Far from being a coincidence, the fact that the physical effect has two simultaneous causes is neatly explained by the dependence relation between them that non-reductive physicalism posits. I prefer the second use of the argument, but shall not argue for it here.[29]

Acknowledgments For helpful comments and suggestions, I would like to thank Sebastian Bender, Catharine Diehl, Luke Fenton-Glynn, Jan Gertken, Franz Huber, Beate Krickel, Rory Madden, Beau Madison Mount, Christian Nimtz, Tobias Rosefeldt, Thomas Sattig, Peter Schulte, Moritz Schulz, Wolfgang Schwarz, Barbara Vetter, Timothy Williamson, audiences in Aarhus, Berlin, Bielefeld, and Dubrovnik, and two anonymous referees.

References

Baker, L. R. (1993). Metaphysics and mental causation. In J. Heil & A. Mele (Eds.), *Mental causation*. Oxford: Clarendon Press.

Christensen, J., & Kallestrup, J. (2012). Counterfactuals and downward causation: A reply to Zhong. *Analysis*, *72*, 513–517.

Fenton-Glynn, L., & Kroedel, T. (forthcoming). Relativity, quantum entanglement, counterfactuals, and causation. *British Journal for the Philosophy of Science*.

[28] Compare, respectively, Kim (1998) and Loewer (2007).

[29] See Kroedel (forthcoming).

Gibbons, J. (2006). Mental causation without downward causation. *Philosophical Review*, *115*, 79–103.
Halpern, J. Y., & Hitchcock, C. (2010). Actual causation and the art of modeling. In R. Dechter, H. Geffner, & J. Y. Halpern (Eds.), *Heuristics, probability and causality: A tribute to Judea Pearl*. London: College Publications.
Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy*, *98*, 273–299.
Hitchcock, C. (2007). Prevention, preemption, and the principle of sufficient reason. *Philosophical Review*, *116*, 495–532.
Jackson, F., & Pettit, P. (1990). Causation in the philosophy of mind. *Philosophy and Phenomenological Research Supplement*, *50*, 195–214.
Kim, J. (1976). Events as property exemplifications. In M. Brand & D. Walton (Eds.), *Action theory*. Dordrecht: Reidel.
Kim, J. (1984). Concepts of supervenience. *Philosophy and Phenomenological Research*, *45*, 153–176.
Kim, J. (1998). *Mind in a physical world*. Cambridge, MA: MIT Press.
Kim, J. (2005). *Physicalism, or something near enough*. Princeton, NJ: Princeton University Press.
Kment, B. (2010). Causation: Determination and difference-making. *Noûs*, *44*, 80–111.
Kripke, S. A. (1980). *Naming and necessity*. Oxford: Blackwell.
Kroedel, T. (forthcoming). Dualist mental causation and the exclusion problem. *Noûs*.
Kroedel, T. (ms). Modeling mental causation.
Lepore, E., & Loewer, B. (1987). Mind matters. *Journal of Philosophy*, *84*, 630–641.
Lewis, D. (1973a). Causation. *Journal of Philosophy*, *70*, 556–567.
Lewis, D. (1973b). *Counterfactuals*. Oxford: Blackwell.
Lewis, D. (1979). Counterfactual dependence and time's arrow. *Noûs*, *13*, 455–476.
Lewis, D. (1986a). Events. In *Philosophical papers* (Vol. II). New York: Oxford University Press.
Lewis, D. (1986b). Postscripts to 'Causation'. In *Philosophical papers* (Vol. II). New York: Oxford University Press.
Lewis, D. (2004). Causation as influence. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals*. Cambridge, MA: MIT Press.
List, C., & Menzies, P. (2009). Nonreductive physicalism and the limits of the exclusion principle. *Journal of Philosophy*, *106*, 475–502.
Loewer, B. (2007). Mental causation, or something near enough. In B. P. McLaughlin & J. Cohen (Eds.), *Contemporary debates in philosophy of mind*. Oxford: Blackwell.
Maudlin, T. (2007). A modest proposal concerning laws, counterfactuals, and explanations. In *The metaphysics within physics*. Oxford: Clarendon Press.
McDermott, M. (1995). Redundant causation. *British Journal for the Philosophy of Science*, *46*, 523–544.
McLaughlin, B. P. (1995). Varieties of supervenience. In E. E. Savellos & U. Yalcin (Eds.), *Supervenience: New essays*. Cambridge, MA: Cambridge University Press.
Paul, L. A., & Hall, N. (2013). *Causation: A user's guide*. Oxford: Oxford University Press.
Swanson, E. (2010). Lessons from the context sensitivity of causal talk. *Journal of Philosophy*, *107*, 221–242.
Williamson, T. (2006). Indicative versus subjunctive conditionals, congruential versus non-hyperintensional contexts. *Philosophical Issues*, *16*, 310–333.
Zhong, L. (2011). Can counterfactuals solve the exclusion problem? *Philosophy and Phenomenological Research*, *83*, 129–147.
Zhong, L. (2012). Counterfactuals, regularity and the autonomy approach. *Analysis*, *72*, 75–85.
Zhong, L. (forthcoming). Why the counterfactualist should still worry about downward causation. *Erkenntnis*.