CHAPTER 4

The Exclusion Problem

4.1 Introduction

The physical world does not manifest any outside influence. A physical event occurs. If we trace its causes, we are likely to find plenty of physical causes. Indeed, it may well be that it is impossible for the physical event not to occur while the laws of physics and the past are as they actually are.[1] How does the physical world leave any room for mental causes that are distinct from physical causes?

There are two issues here. The first issue is that the physical world might not allow any causal influence of the mental whatsoever if the mental is distinct from the physical. Recall Leibniz's argument from the conservation of momentum and kinetic energy from Section 0.2: given these conservation laws, Leibniz held, the mental cannot have physical effects at all. We saw that Leibniz's argument was unsatisfactory as it stands, but the point can be made more generally. Given that the actual physical laws and the actual past necessitate the occurrence of a certain future physical event, it might seem that this event cannot have any additional mental causes.

The first issue can be resolved, at least in principle, by claiming that mental causes of physical effects are never alone in causing these effects, but always act in tandem with physical causes of the same effect. This suggestion, however, gives rise to the second issue. If the physical effects of mental causes always have additional physical causes, it seems to follow that they are overdetermined. Thus, cases of mental causation seem to be similar to firing squads, where the deaths of the victims are overdetermined by the

[1] This does not follow from our earlier assumption of determinism alone, which I shall continue to make in this chapter for simplicity. Determinism alone makes it impossible for the actual laws *as a whole* to hold and the actual past to obtain while the future differs. The laws as a whole might contain psychophysical laws as well as physical laws. The claim from the main text does follow, however, if we add our assumption that any actual psychophysical laws are synchronic.

https://doi.org/10.1017/9781108762717.006 Published online by Cambridge University Press

firings of the squad members. It seems implausible, however, that there is this kind of overdetermination whenever there is mental causation.

Both issues would deserve the label 'exclusion problem'.[2] It has been more common to apply the label to the second issue, however, which has also been more prominent in the mental causation debate.[3] In what follows, I shall follow this tradition and use 'exclusion problem' to refer to the second of the issues. It is the exclusion problem in the sense of the second issue that is the main focus of this chapter.

Here is a more rigorous way of presenting the exclusion problem. We can introduce the problem as a set of five claims:[4]

(DISTINCTNESS)	All mental events are distinct from physical events.
(EFFICACY)	Some mental events have physical effects.
(COMPLETENESS)	Every physical event that has a cause at all has a physical cause.
(EXCLUSION)	No effect has more than one cause at a given time, unless it is overdetermined.
(NON-OVERDETERMINATION)	The effects of mental events are not systematically overdetermined.

Each of these five claims seems plausible, yet they are inconsistent or at least in tension with one another.[5] (That the claims fall short of genuine inconsistency is indicated by the sentences starting with 'presumably' in the following reasoning.) Given (EFFICACY), a certain physical event has a mental cause. By (DISTINCTNESS), this mental cause is distinct from the physical causes (if any) of the physical effect. By (COMPLETENESS), the physical effect has a physical cause (for, by (EFFICACY), it has a cause in the first place, namely a mental cause). Presumably, it has a physical cause that is

[2] Sometimes the exclusion problem is introduced through the metaphor of causal work – indeed, I have done so myself in Section 0.1. Saying that the physical does all the causal work and leaves nothing for the mental to do remains neutral between the two issues presented here: the causal work metaphor can be read as saying that there is no job opportunity of any kind left for the mental (first issue) or as saying that the mental can do the job of causing physical events only by partaking in some kind of job-sharing with the physical causes of those events (second issue).

[3] In the sense of the second issue, the exclusion problem is due to Malcolm (1968). It has been refined and much discussed by Kim (e.g., 1989, 1998, 2005, 2007). For a recent overview, see Bennett 2007.

[4] I am loosely following Bennett (2008: 281) here. The main difference is that she talks about sufficient causes in her formulation of the exclusion problem while I talk about causes *tout court*. We will discuss sufficient causes in Section 4.5.

[5] For a presentation of the claims of the exclusion problem where they are strictly inconsistent, see Hitchcock 2012b. Such a presentation comes at the price of additional complexity, however.

simultaneous with the mental cause. By (EXCLUSION), the physical effect is overdetermined by its simultaneous physical and mental causes. Presumably, the present case is far from uncommon. Thus, given (EFFICACY), (DISTINCTNESS), (COMPLETENESS), and (EXCLUSION), there is widespread and systematic overdetermination of physical effects by mental and physical causes. By (NON-OVERDETERMINATION), however, there is no such overdetermination; contradiction.

I have just presented the exclusion problem as a set of five principles that are inconsistent or at least in tension with one another. Other presentations are possible along different and independent dimensions. First, one can use slightly different formulations of the principles. Second, one can present the problem not as an inconsistent (or near-inconsistent) set, but as an argument. When the principles are presented as an argument, the negation of one of the principles features as the conclusion of the argument while the remaining principles feature as premises. Typically, the negation of (DISTINCTNESS) or the negation of (EFFICACY) is chosen as the conclusion of the argument, and the whole thing is called the 'exclusion argument' rather than the 'exclusion problem'. Third, one can present the exclusion problem by using fewer principles than I have. Typically, (EXCLUSION) gets omitted or fused with (NON-OVERDETERMINATION) into a single principle.[6]

I have chosen the formulation of the problem as an inconsistent (or near-inconsistent) set of five principles, because on this formulation the logical relation between our principles is straightforward, and, unlike in formulations as arguments, there is no default about which principle to reject. The points from this chapter could also be applied to different formulations, however. For example, we shall encounter objections against (EXCLUSION). Take a presentation of the exclusion problem as an argument for the negation of (DISTINCTNESS) or the negation of (EFFICACY) whose premises comprise the remaining principles except (EXCLUSION). In the context of that presentation, we can read our objections to (EXCLUSION) as objections to a tacit assumption of the argument that is required to make the argument valid.

[6] To give some examples of the various possibilities of formulating the exclusion problem: Carey (2011: 251–252) presents the problem as inconsistent set, but with four instead of five principles. Yablo (1992: 247–248) presents the exclusion problem as an argument against the efficacy of the mental with three premises. Lowe (2000: 571–572) and Gibb (2014: 328) present the exclusion problem as an argument against the distinctness of the mental and the physical with three premises that do not include (EXCLUSION). All of these authors use slightly different formulations of the principles. Bennett (2008: 282) also discusses different shapes the exclusion problem can take.

Let us consider the exclusion problem as exemplified by the five (near-) inconsistent claims (DISTINCTNESS), (EFFICACY), (COMPLETENESS), (EXCLUSION), and (NON-OVERDETERMINATION) then. Given the tension between the five claims, everyone, regardless of their views about the nature of mind and their views about causation, has to give up at least one of them.[7] Non-reductive physicalists and dualists hold that all mental properties are distinct from all physical properties. Given the strong Kimian account of events, a difference in constitutive properties entails the distinctness between events, so it follows that all mental events are distinct from all physical events. Thus, non-reductive physicalists and dualists cannot reject (DISTINCTNESS). Rejecting (EFFICACY) means accepting epiphenomenalism, which is far from attractive in its own right. Besides, the arguments from Chapters 2 and 3 have shown that the existence of mental causation follows straightforwardly from non-reductive physicalism and can be accommodated by dualists by making certain assumptions about the status of the psychophysical laws as well. Thus, non-reductive physicalists and dualists who make these assumptions cannot reject (EFFICACY). Rejecting (COMPLETENESS) seems implausible, not just from a general scientific point of view, but also in the case at hand.[8] It is very plausible that the instance of the actual physical realizer or base of our mental property-instance is a cause of the physical effect of the mental property-instance,[9] 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 non-reductive physicalists and dualists then.

The plan for the remainder of this chapter is as follows. The next two sections argue that both the option of rejecting (EXCLUSION) and the option of rejecting (NON-OVERDETERMINATION) are viable for non-reductive physicalists and dualists. Section 4.2 argues that non-reductive physicalists and dualists can reject (EXCLUSION) by making a case for the falsity of certain counterfactuals that are necessary conditions for the overdetermination of physical effects by mental and physical causes. Section 4.3 argues that, even if the argument against (EXCLUSION) should

[7] See Bennett 2008: 281. Kim (e.g., 2005) argues that, in the light of the exclusion problem, non-reductive physicalists cannot but deny the efficacy of the mental, which, according to him, amounts to a *reductio* of non-reductive physicalism.

[8] For a discussion of the history of the claim that the physical world is causally complete, see Papineau 2001. For dualist critiques of physical completeness claims, see Lowe 2000, 2003, 2008, BonJour 2010, and Gibb 2015b.

[9] Ignoring, once more, the worries that instances of realizers cannot be causes or effects in principle; see Section 4.4 for further discussion.

fail and non-reductive physicalists and dualist are committed to the overdetermination of the physical effects of mental causes, they can reject (NON-OVERDETERMINATION) instead. If those physical effects are overdetermined, the argument goes, the cases are very dissimilar to prototypical cases of overdetermination such as firing squads. Section 4.4 takes up the issue of whether the instances of realizers can both necessitate instances of mental events and cause physical events. It shows that the argument for the causal inertia of realizers is at best inconclusive and that any such inertia does not spread to the instances of the realized mental properties. Section 4.5 discusses a formulation of the exclusion problem in terms of sufficient causes. It argues that the problem is more severe if thus formulated, but that the severity does not carry over to the solution of the original exclusion problem, because our principle about causation in terms of counterfactual dependence does not commit us to any potentially problematic claims about sufficient causes.

4.2 Denying Exclusion

According to (EXCLUSION), no effect has more than one cause at a given time, unless it is overdetermined. In other words, if an effect has more than one cause at a given time, then it is overdetermined. What is overdetermination? It seems to be an essential feature of cases of overdetermination such as the firing squad that the overdetermined event would still have occurred had either overdetermining event occurred without the other. In a firing squad of two, for instance, the victim would still have died if the first squad member had fired while the second had not; likewise the victim would still have died if the second squad member had fired while the first squad member had not. Thus, in order for events *c* and *d* to overdetermine event *e*, the following counterfactuals have to be true:[10]

(O_1) If *c* had occurred without *d*, then *e* would still have occurred.
(O_2) If *d* had occurred without *c*, then *e* would still have occurred.[11]

[10] If we are dealing with more than two overdetermining events, there are two ways of generalizing (O_1) and (O_2). We can demand that the overdetermined event would still have occurred if any of the overdetermining events had not occurred while *all* the other overdetermining events had still occurred. Alternatively, one can demand that the overdetermined event would still have occurred if any of the overdetermining events had not occurred while *some* of the other overdetermining events had still occurred. For further discussion of these generalizations, see Kroedel 2008: 129 n. 14.

[11] Bennett (2003, 2008) and Mills (1996) endorse the stronger requirement that (O_1) and (O_2) be not merely true, but non-vacuously true. I will not explore the strategy of pleading vacuity, because it is not available for dualists. For alleged counterexamples to the truth of (O_1) and (O_2) as a necessary condition for overdetermination, see Aimar 2011: 474–476, Bennett 2008: 289 n. 13, and Won 2014:

Claims (O_1) and (O_2) are put forward merely as necessary conditions for overdetermination, not as necessary and sufficient conditions.

It may seem that our sufficient condition for causation in terms of counterfactual dependence allows us to give short shrift to (EXCLUSION). Take the case where there is a car crash, and it would not have happened if the driver had not been drunk. If the road had not been icy, the crash would not have happened either. By our sufficient condition for causation, both the driver's being drunk and the road's being icy caused the car crash.[12] But assume that the crash was sensitive both to the driver's being drunk and to the road's being icy in that the crash would not have occurred if either of the driver's being drunk and the road's being icy had occurred without the other. Then (O_1) and (O_2) are false. Given that the truth of (O_1) and (O_2) is necessary for overdetermination, the driver's being drunk and the road's being icy do not overdetermine the car crash. We have a counterexample to (EXCLUSION).

Here is another counterexample to (EXCLUSION) where it is even clearer that (O_1) and (O_2) are false.[13] A defendant faces trial by a jury of two. Both jurors vote to convict, and the defendant goes to jail. At the court, convictions have to be unanimous. Thus, if the first juror had not voted to convict, then the defendant would not have gone to jail. Similarly, if the second juror had not voted to convict, then the defendant would not have gone to jail. By our sufficient condition for causation, the first juror's voting to convict causes the defendant's imprisonment, and so does the second juror's voting to convict. But if either juror had voted to convict while the other had not, the defendant would not have gone to jail either. So again (O_1) and (O_2) are false while we have two simultaneous causes.[14]

These counterexamples against (EXCLUSION) do not dissolve the exclusion problem, however. While (EXCLUSION) was formulated very generally, without reference to mental or physical causes, it was of course intended to be applied to the case of simultaneous mental and physical

212–214. For criticism of the claim that the vacuous truth of (O_1) or (O_2) removes overdetermination, see Bernstein 2016.

[12] The example is from Lewis 1986a: 214. As in previous examples, we might have to take a suitable temporal part of, say, the road's being icy in order for our sufficient condition for causation to be applicable.

[13] The example is a simplified version of an example from Kroedel 2008: 127–128.

[14] Friends of (EXCLUSION) might respond by redefining overdetermination such that overdetermination simply is causation by several causes, or perhaps by several simultaneous causes, and no longer requires the truth of (O_1) and (O_2). But this response merely shifts the vulnerability to the counterexamples to (NON-OVERDETERMINATION), which, on the suggested definition of overdetermination, can in turn be given short shrift.

causes of the same effect. We can make this qualification explicit and formulate the claim as follows:

(EXCLUSION*) No effect has a mental cause and a distinct physical cause that occur at the same time, unless it is overdetermined.

The car crash and jury examples are not counterexamples to (EXCLUSION*), and it is *prima facie* unclear whether we can find analogous cases that involve mental and physical causes. Like (EXCLUSION), (EXCLUSION*) is inconsistent, or at least in tension, with the other four claims from the original presentation of the exclusion problem, so the problem persists.

Non-reductive physicalists and super-nomological dualists can argue against (EXCLUSION*) by showing that, even in the case of mental and physical causes, at least one of the counterfactuals (O_1) and (O_2) is false. Applying (O_1) and (O_2) to our case of mental property *M*, its actual realizer *P*, and physical property *P** that is instantiated later than *M*, we get:

(O_1*) If *M* had been instantiated without *P*, then *P** would still have been instantiated. ($M \;\&\; \sim P \;\Box\!\!\rightarrow P^*$)

(O_2*) If *P* had been instantiated without *M*, then *P** would still have been instantiated. ($\sim M \;\&\; P \;\Box\!\!\rightarrow P^*$)

Let us consider (O_2*) first. Non-reductive physicalists endorse the strong supervenience of mental properties on physical properties. By this strong supervenience, the instantiation of a realizer necessitates the instantiation of the realized property. In our case, the instantiation of *P* necessitates the instantiation of *M*. Thus, *P* cannot be instantiated without *M*; hence, the antecedent of (O_2*) is impossible; hence, (O_2*) is vacuously true given non-reductive physicalism, so non-reductive physicalists cannot reject (EXCLUSION*) by rejecting (O_2*).

Like all dualists, super-nomological dualists deny the strong supervenience of mental properties on physical properties. They hold that there are worlds where *P* is instantiated without *M*, although this requires a violation of the actual psychophysical laws. Given the modified miracles approach to overall similarity that was presented in Section 2.5 on behalf of super-nomological dualism, it is of the first importance to avoid psychophysical miracles and of the second importance to avoid 'ordinary' miracles. Thus, worlds that do not involve any 'ordinary' miracles in addition to the psychophysical miracle that is requires for *P* to be instantiated without *M* are closer to the actual world than any worlds that do. Assuming that the *P**-instance follows lawfully from the previous physical

state, which includes the *P*-instance, *P** is still instantiated in the closest worlds where *P* is instantiated without *M*. In other words, the closest worlds where the antecedent of (O_2*) is true are just like the actual world except that the instantiation of *M* is removed by a psychophysical miracle. Hence (O_2*) is non-vacuously true given super-nomological dualism,[15] so super-nomological dualists cannot reject (EXCLUSION*) by rejecting (O_2*) either.

Thus, rejecting (EXCLUSION*) by rejecting (O_2*) does not look promising for non-reductive physicalists and super-nomological dualists. The case of (O_1*) is different, however. Both non-reductive physicalists and super-nomological dualists can make a case against (O_1*). They can argue that, given their respective views, *P** might not have been instantiated if *M* had been instantiated without *P*. This result contradicts (O_1*).

Here are the details of the argument against (O_1*). Suppose that we are dealing with our paradigmatic case of putative mental causation, where *M* is the property of having a headache, *P* is the property of having firing c-fibres, and *P** is the property of having one's hand moving towards an aspirin. The argument against (O_1*) begins with a counterfactual that has a slightly different antecedent than (O_1*). What would or might have been the case if *M* had been not only instantiated without *P*, but accompanied by a physical realizer or base other than *P*?[16] We already saw in the previous chapter (but ignored the issue for simplicity) that implementing some of the alternative realizers of headaches is likely to be so disruptive that the instantiation of those realizers would no longer make my hand move towards the aspirin. It seems plausible that even the closest possibility of realizing or having a base for headaches other than through firing c-fibres requires some large-scale tampering with my nervous system. Suppose that the closest such possibility is that x-fibres, which are not actually present in humans, be implanted in my brain. It seems plausible that the easiest way of implanting them might not leave all the outgoing connections intact, such that having firing x-fibres no longer makes my hand move towards the aspirin. This seems equally plausible for the non-reductive physicalist case and for the super-nomological dualist case. Thus, we get:

(1) If *M* had been instantiated with a different physical realizer/base instead of *P*, then *P** might not have been instantiated. ($M \;\&\; \sim P \;\&\; \cup\mathbf{P} \Diamond\!\!\rightarrow \sim P^*$)

[15] Bennett (2008: 291–292) puts forward a similar argument for the truth of (O_2*) given standard dualism.

[16] Given non-reductive physicalism (though not given dualism), it is of course impossible for *M* to be instantiated in the absence of *P* without being accompanied by a realizer other than *P*.

Further, if I had had a headache without having had firing c-fibres, some other physical realizer or base of headaches would have been instantiated instead:

(2) If *M* had been instantiated without *P*, then some other physical realizer/base of *M* would have been instantiated instead. (M & ~P $\Box\!\!\rightarrow \cup\mathbf{P}$[17])

Claim (2) is true according to non-reductive physicalism, according to which the instantiation of headaches is strictly equivalent to the instantiation of a realizer of headaches.[18] It is also true according to the modified miracles approach that is endorsed by super-nomological dualists. If we face the choice between antecedent-worlds of (2) where some other physical base of headaches is instantiated and antecedent-worlds of (2) where none is, the former worlds come out closer to the actual world according to the modified similarity account, because they do not involve a psychophysical miracle.

Now (1) and (2) logically imply:

(3) If *M* had been instantiated without *P*, then *P** might not have been instantiated. (M & ~P $\Diamond\!\!\rightarrow$ ~P*)[19]

However, by the definition of the 'might' conditional, (3) is true if and only if (O_1*) is false. So (O_1*) is false.

It might be objected that this result is an artefact of our strong Kimian conception of events, according to which the occurrence of the physical event that underlies the mental event is strictly equivalent to the instantiation of its constitutive property, namely the property of having firing c-fibres, at the relevant time by the relevant subject. According to a weak Kimian conception or a Lewisian conception of events, the event that we

[17] In order to facilitate later derivations, the formalization does not include ~P as a conjunct of the consequent. This does not render the formalization unfaithful, for M & ~P $\Box\!\!\rightarrow \cup\mathbf{P}$ is logically equivalent to M & ~P $\Box\!\!\rightarrow \cup\mathbf{P}$ & ~P.

[18] For further discussion of (2) in the context of non-reductive physicalism, see Loewer 2001b: 319–320, Bennett 2003: 481–482, and references therein.

[19] The inference from (1) and (2) to (3) has the form of an inference from χ & ϕ $\Diamond\!\!\rightarrow \psi$ 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 ~[χ & $\phi \Box\!\!\rightarrow$ ~ψ] and $\chi \Box\!\!\rightarrow \phi$ to ~[$\chi \Box\!\!\rightarrow$ ~ψ] is, which is valid if and only if the inference from $\chi \Box\!\!\rightarrow$ ~ψ and $\chi \Box\!\!\rightarrow \phi$ to χ & $\phi \Box\!\!\rightarrow$ ~ψ is, which in turn is valid if and only if the inference from $\chi \Box\!\!\rightarrow \psi$ and $\chi \Box\!\!\rightarrow \phi$ to χ & $\phi \Box\!\!\rightarrow \psi$ is. The premises and conclusion of the last 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 χ & $\phi \Box\!\!\rightarrow \psi$ according to Lewis 1973c: 433.

https://doi.org/10.1017/9781108762717.006 Published online by Cambridge University Press

refer to as 'the c-fibre firing' (call it p) can have a different modal profile. In particular, it might be that p would already fail to occur if I had lacked a feature that is more specific than merely having firing c-fibres. It might be, for instance, that p is essentially not merely a c-fibre firing, but a c-fibre firing at a rate between 99 and 101 Hz; then my c-fibres' firing at a rate of 102 Hz instead of the actual rate of, say, 100 Hz would have been enough for p not to occur. If this is the case, p is rather *fragile*; that is, p could not easily have occurred in a different manner (in other words, if a p-like event had occurred in a manner different from p's, it would not have been p, but a distinct event).[20] Similarly for my hand's moving towards the aspirin (call this event p^*). Event p^*, too, might or might not be very fragile. Let us assume that, in fact, my hand moves slowly towards the aspirin with my thumb facing sideways. If p^* is not very fragile, it would still have occurred if my hand had moved fast and with the thumb facing upwards; not so if p^* is very fragile. Now, it seems that the argument against (O_1*) works only if p is assumed to be not very fragile. If p is very fragile – especially while p^* is not very fragile – then premise (1) no longer seems plausible, for in this case my c-fibres fire only slightly differently in the closest antecedent-worlds of (1), and my hand still moves towards the aspirin.

Friends of counterfactual accounts of causation are not well advised to conceive of events as very fragile, for this might yield too many cases of counterfactual dependence and, consequently, causation (see Lewis 1986d: 196–199). (The strong Kimian conception of events that I have advocated yields fragility along the temporal dimension, which, as we saw in Section 1.3 is undesirable, but worth the overall utility of the strong Kimian account.) But even if we accept for the sake of argument that events may be very fragile – particularly that event p may be very fragile – we can still argue against (O_1*). More precisely, we can argue against an analogue of (O_1*) that talks about the (non-)occurrence of token events p and p^* instead of the instantiation of properties P and P^*:

(O_1**) If M had been instantiated while p had not occurred, then p^* would still have occurred. $(M \,\&\, \sim\mathrm{Oc}(p) \mathrel{\square\!\!\rightarrow} \mathrm{Oc}(p^*))$[21]

[20] The terminology is due to Lewis (1986d). Lewis takes events to be very fragile if they could not easily have differed in time and manner.

[21] We could formulate a principle analogous to (O_1**) that talks about the occurrence of a (weak Kimian/Lewisian) event m instead of the instantiation of property M. That would introduce some unnecessary complications, however, for it would make the transition between claims about the mental event and claims about the underlying physical events that correspond to M's realizers/bases much more cumbersome.

The strategy behind the argument against (O_1**) that I will present is that, irrespective of what assumptions we make about the fragility of p and of p*, (O_1**) is false. We can read the above argument against (O_1*) as an argument against (O_1**) that assumes event p to be not very fragile: if p is not very fragile, its non-occurrence and replacement with an alternative realizer or base event might have resulted in the failure of p* to occur. This 'might' claim seems plausible if p* is itself not very fragile; *a fortiori*, it seems plausible if p* is very fragile, for in this case it takes even less for p* not to occur. Thus, the above argument can be read as an argument against (O_1**) that covers two out of four sub-cases, namely the sub-case where p is not very fragile while p* is very fragile and the sub-case where neither p nor p* are very fragile.

To cover the remaining two sub-cases, suppose now that p is very fragile. In this case, if p had not occurred, a realizer/base of M might still have been instantiated:

(4) If p had not occurred, a realizer/base of M might have been instantiated. ($\sim Oc(p) \Diamond\!\!\rightarrow \cup\mathbf{P}$)

We could even turn (4) into a 'would' counterfactual. If a very fragile p had not occurred, then, it seems, my c-fibres would have fired slightly differently, in which case the different c-fibre-firing property that would have been instantiated would still have been among the realizers/bases of M. For our purposes, however, the weaker 'might' conditional will suffice.

Claim (4) does not talk about the later physical event and thus is true independent of whether or not p* is very fragile. But the truth of certain counterfactuals about the relation between p, instances of realizers/bases of M, and p* depends on the fragility (or lack thereof) of p* as well. Let us therefore treat the two sub-cases about the fragility of p* separately. Suppose first that p* is *not* very fragile, while continuing to suppose that p is very fragile. Let us assume that actually my hand moves slowly towards the aspirin with my thumb facing sideways; then the assumption that p* is not very fragile seems to allow that p* would still have occurred if my hand had moved quickly towards the aspirin with my thumb facing upwards. It seems that if p* is not very fragile while p is very fragile, then p* would still have occurred if some physical realizer/base of M had been instantiated in the absence of p:

(5) If a realizer/base of M had been instantiated in the absence of p, then p* would still have occurred. ($\sim Oc(p)$ & $\cup\mathbf{P} \,\Box\!\!\rightarrow Oc(p^*)$)

https://doi.org/10.1017/9781108762717.006 Published online by Cambridge University Press

For it seems that if my c-fibres had fired only slightly differently, I would still have reached for an aspirin, although presumably my hand would have moved somewhat differently (a bit faster than it actually did, say); this would still have sufficed for p^* to occur if p^* is not very fragile. Claims (4) and (5) logically imply:

(6) If p had not occurred, then p^* might still have occurred. $(\sim Oc(p) \Diamond\!\!\rightarrow Oc(p^*))$

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.[22] We saw that it is plausible that the (putative) physical effects of mental property-instances counterfactually depend on the instances of the realizers/bases of those mental property-instances (see Section 2.6). If we apply this result to our case, we get:

(7) If p had not occurred, then p^* would not have occurred. $(\sim Oc(p) \Box\!\!\rightarrow \sim Oc(p^*))$

Claim (7) is inconsistent with (6), however, for by the definition of the 'might' conditional, (7) is equivalent to the negation of (6). So if we assume that p is very fragile (which yields (4)) while p^* is not very fragile (which yields (5)), it follows (via (6)) that p^* does not counterfactually depend on p. Contrapositively, if we want to uphold the claim that p^* counterfactually depends on p, we have to reject either the assumption that p is very fragile or the assumption that p^* is not very fragile. I take it that the plausibility of the claim that p^* counterfactually depends on p outweighs that of either assumption. We can therefore conclude that the sub-case where p is very fragile while p^* is not does not obtain. This leaves us with the sub-case where p and p^* are both very fragile.

Suppose, then, that p and p^* are both very fragile. In this case, it seems, p^* might have failed to occur if a physical realizer/base of M had occurred in the absence of p. For instance, my hand might have moved faster towards the aspirin if my c-fibres had fired at a rate of 102 Hz instead of firing at the actual rate of 100 Hz; in this case, p^* would not have occurred if

[22] 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.

p^* is very fragile. So we can reject (5), and there is no obstacle to the joint truth of (7) and (4).

To complete the argument against (O_1**) for the sub-case in which both p and p^* are very fragile, let us leave the counterfactual relation between M's physical realizers/bases and p^* for a moment and consider the relation between p, the physical realizers/bases, and the M-instance itself. What would have been the case if a realizer/base of M had been instantiated in the absence of p? It takes an 'ordinary' miracle to bring about the instantiation of a realizer/base of M while preventing the occurrence of p. According to non-reductive physicalism, it is impossible for M not to be instantiated if a realizer of M is instantiated. According to super-nomological dualism, this is possible, but it requires a psychophysical miracle, which has to be avoided at all costs.[23] Thus, M is still instantiated in the closest worlds where a realizer/base of M is instantiated in the absence of p, and the following is true:

(8) If some physical realizer/base of M had been instantiated while p had not occurred, then M would still have been instantiated. ($\sim Oc(p)$ & $\cup \mathbf{P}\ \Box\!\!\rightarrow M$)

Claims (7), (4), and (8) are inconsistent with (O_1**). To see this, note first that, by the inference rule we used in the derivation of (6) above, (4) and (8) logically imply

(9) If p had not occurred, then M might still have been instantiated. ($\sim Oc(p)\ \Diamond\!\!\rightarrow M$)

By another application of the same rule, (9) and (O_1**) logically imply (6); as we saw, (6) contradicts (7). In sum, (7) (4), (8), and (O_1) are jointly inconsistent. In other words (7), (4), and (8) logically imply that (O_1**) is false.[24]

To summarize the argument against (O_1**): there are four possible cases depending on whether or not event p is very fragile and on whether or not event p^* is very fragile. Given that p^* counterfactually depends on p, we can rule out the case where p is very fragile while p^* is not very fragile. In all other cases, (O_1**) is false. Therefore, (O_1**) is false. If the overdetermination of p^* by p and the M-instance requires the truth of (O_1**), both non-reductive

[23] Moreover, on either view, not having M instantiated detracts from match of particular fact with the actual world, where M is instantiated.

[24] Since (7) by itself is consistent with (O_1**) (more on this in the next section), using (7) as a premise in an argument against (O_1**) does not beg the question.

https://doi.org/10.1017/9781108762717.006 Published online by Cambridge University Press

physicalists and super-nomological dualists can reject (EXCLUSION*) and deny that mental causation entails overdetermination.

I have presented a number of arguments that assumed certain events to be very fragile. Friends of our counterfactual principle about causation are ill advised to adopt a fragile conception of events lest they should be committed to too much counterfactual dependence, however (see Lewis 1986d). Thus, in what follows I shall revert to the strong Kimian conception of events, according to which the occurrence of an event is strictly equivalent to the instantiation of its constitutive property by the constitutive object at the constitutive time. If we assume the constitute property not to be too specific, the strong Kimian conception will not yield events that are fragile owing to their constitutive property. (According to the strong Kimian conception, events are still rather fragile owing to their constitutive time, but for simplicity I will continue to ignore this issue.)

Even on a strong Kimian conception of not-so-fragile events, one might have worries about some of the arguments against (EXCLUSION*) that were presented in this section. The argument for (1) works only if realizers/bases of headaches other than c-fibre firings are rather difficult to implement. More generally, in order for the strategy of this section to work even if events are conceived of as not very fragile, it needs to be the case that replacing the actual realizer or base of my headache with an alternative realizer or base would have been sufficiently disruptive to no longer bring it about that my hand moves towards the aspirin. Counterfactuals about what would have been the case if the actual realizer or base had been thus replaced are not the most straightforward claims to evaluate. Their truth is also hostage to empirical fortune, because it is partly an empirical question what the alternative bases or realizers of headaches are and how easy it is to implement them. It would be good for non-reductive physicalists and dualists to have a contingency plan in case the arguments against (EXCLUSION*) turn out to be unworkable.

4.3 Denying Non-Overdetermination

The previous section argued that, modulo some empirical uncertainties, non-reductive physicalists and super-nomological dualists can make a good case against the claim that physical effects must be overdetermined if they have both a physical cause and a simultaneous mental cause. If you are not convinced or too worried about the empirical uncertainties, never mind. In this section I will argue that, even if non-reductive physicalists and super-nomological dualists *are* committed to the overdetermination of those

physical effects, they can deny that this kind of overdetermination is objectionable. In other words, even if they accept (EXCLUSION*), they can still deny (NON-OVERDETERMINATION).

Strictly speaking, a failure of the arguments from the previous section need not amount to conceding that the *P**-instance is overdetermined by the *M*-instance and the *P*-instance, since the conditions from the $(O_1)/(O_2)$ family were presented only as necessary conditions for overdetermination, not as necessary and sufficient conditions. Indeed, (O_1) and (O_2) are true not only in cases of overdetermination, but also in cases of pre-emption. Recall the example of Billy and Suzy, who each throw a rock at a bottle at the same time. Billy's rock arrives there first and shatters the bottle (see Section 1.4). If Billy had not thrown and Suzy had, then the bottle would still have shattered; likewise if Suzy had not thrown and Billy had.

The difference between the cases seems to be that Billy's throw has a better claim to causing the bottle's shattering than Suzy's throw has – indeed, Suzy's throw seems to have no such claim at all – while the firings of the two squad members have an equally good claim to causing the victim's death. It may seem promising to conjoin the condition that the putative overdeterminers have an equally good or bad claim to being a cause to the claim that (O_1) and (O_2) be true in order to formulate a necessary and sufficient condition for overdetermination (see Lewis 1986d: 193–200).

I will not, however, assume this or any other specific characterization of overdetermination. After all, 'overdetermination' is a technical term, so, to a certain extent, theorists are free to stipulate how they understand it.[25] Non-reductive physicalists and super-nomological dualists might insist on a comparatively demanding notion of overdetermination that requires much besides the truth of (O_1^*) and (O_2^*) and thus insist on the falsity of (NON-OVERDETERMINATION), but by itself this would not win the day. As I will explain in more detail below, the rationale behind the (EXCLUSION*)-cum-(NON-OVERDETERMINATION) part of the exclusion problem is that cases where physical effects 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 the physical effects of mental causes are overdetermined is

[25] For instance, Bennett holds that overdetermination requires that both overdetermining events be 'causally sufficient' for the overdetermined event (2008: 288). For a discussion of various alternative proposals, see Carey 2011.

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 non-reductive physicalists and super-nomological dualists by insisting on any specific characterization of overdetermination, I shall concede not merely the truth of the claims from the $(O_1)/(O_2)$ family for the sake of argument, but also the truth of any other conditions required by a reasonable characterization of overdetermination.[27]

Assume, then, that the *M*-instance and the *P*-instance overdetermine the *P**-instance 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. In particular, it would be a strange coincidence if a physical effect had a mental cause in addition to its physical cause; that it has a physical cause in the first place does not seem surprising.[28]

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

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

[27] If these conditions include the condition that both overdetermining events be causes of the overdetermined event, one can ignore worries about their individual efficacy. In any event, such worries do not arise in our case (setting aside the general worries about whether realizer-instances can in principle be causes or effects), owing to the counterfactual dependence of the *P**-instance on both the *M*-instance and the *P*-instance. For further discussion of the efficacy of individual overdetermining events, see Schaffer 2003.

If the conditions on overdetermination include Lewis's requirement that both overdeterminers have an equally good (or bad) claim to be a cause of the overdetermined event, one might think that the exclusion problem dissolves because mental events have a worse claim to be causes of physical events than physical events have. This reasoning, however, only solves the exclusion problem at the expense of a highly unattractive assumption, namely the assumption that mental events have a worse claim to having physical effects.

[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) and in the sense of not standing in a relation of synchronic dependence. 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.

(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, because our *x* is not a prototypical *F*. 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.

On the account of mental causation presented here, there is a synchronic dependence relation between the physical cause and the mental cause of the physical effect. According to super-nomological dualism, the two causes are related by psychophysical laws, which could not have failed so

easily as ordinary laws of nature. According to non-reductive physicalism, the two causes are related by metaphysical necessity. In prototypical cases of overdetermination like the firing squad, no such synchronic dependence relation holds between the two overdeterminers (which might, of course, still have a common cause).

The synchronic dependence relation also explains, or at least contributes to explaining, why the physical effect in question has a mental cause in addition to the physical cause; thus, it is no coincidence that the effect is overdetermined. This explanation has two parts. One part is an explanation of why the mental event occurs. The other is an explanation of why that mental event causes the physical effect.[29] Take the instances of our properties *M, P*, and *P**. Given that the instantiation of *P* implies the instantiation of *M* at least with super-nomological necessity, we can straightforwardly explain why *M* is instantiated from *P*'s being instantiated. That the *M*-instance causes the *P**-instance does not follow quite so straightforwardly from the relation between *M* and *P*, even if we take into account that *P* causes *P** by counterfactual dependence. Given this counterfactual dependence, we have the following claim:

(10) If *P* had not been instantiated, then *P** would not have been instantiated. ($\sim P \mathrel{\Box\!\!\rightarrow} \sim P^*$)

According to non-reductive physicalism, it is metaphysically necessary that *M* is instantiated if *P* is instantiated. Contrapositively, we have the following:

(11) Necessarily, if *M* is not instantiated, then *P* is not instantiated. ($\Box[\sim M \supset \sim P]$)

According to super-nomological dualism, (11) is false, but the synchronic relation between the *P*-instance and the *M*-instance still yields the following:

(12) If *M* had not been instantiated, then *P* would not have been instantiated. ($\sim M \mathrel{\Box\!\!\rightarrow} \sim P$)

Neither the conjunction of (10) and (11) nor the conjunction of (10) and (12) entails that the *P**-instance counterfactually depends on the *M*-instance (see Section 2.2). Still, we can regard claims (10), (11), and (12) as contributing to an explanation of why the *M*-instance causes the *P**-instance. The claims are close cognates of the premises we have used

[29] Sharpe (2015) also holds that the worry should be addressed by giving these two explanations.

https://doi.org/10.1017/9781108762717.006 Published online by Cambridge University Press

to argue for this causal claim in Chapter 2. We can also combine them with further premises to get a watertight argument for this claim.[30] Together with the explanation of *M*'s instantiation as such, this should sufficiently attenuate the coincidence worry about the overdetermination of the *P**-instance by the *P*-instance and the *M*-instance.[31]

The most important dissimilarity between cases of mental causation and cases of overdetermination is that, according to the account I have presented, the physical effect counterfactually depends on its mental cause and also counterfactually depends on the instance of the actual physical realizer/base of the mental cause. Thus, if *M* had not been instantiated, then *P** would not have been instantiated; nor would *P** have been instantiated if *P* had not been instantiated. This is compatible with the assumed truth of (O_1*) and (O_2*). Dualists have to assume that (O_1*) and (O_2*) are non-vacuously true if true. In this case, all that is required is, first, that the closest worlds where *M* is instantiated without *P* (where *P** is instantiated by (O_1*)) be further from actuality than the closest worlds where *P* is not instantiated (where the *P**-instance does not occur, by its counterfactual dependence on the *P*-instance), and, second, that the closest worlds where *P* is instantiated without *M* (where *P** is instantiated by (O_2*)) be likewise further from actuality than the closest worlds where *M* does not occur (where the *P**-instance does not occur, by its counterfactual dependence on the *M*-instance).[32] Non-reductive physicalists will take claim (O_2*) to be vacuously true. Its vacuous truth is even more straightforward to square with the counterfactual dependence of the *P**-instance on the *M*-instance, for in this case there are no worlds where *P* is instantiated in the absence of *M* to consider. Figure 4.1 illustrates the point for the dualist case. In the figure, ~P is true inside the bottom left parabola, and ~M is true in the area that largely coincides with the ~P-area, but diverges from it where

[30] We can, for instance, recover premise (18) ($\sim\cup\mathbf{P}_M \mathrel{\Box\!\!\rightarrow} \sim P^*$) from the argument for the causation of the *P**-instance by the *M*-instance from Section 2.5 as follows. Assume that (i) no alternative realizer/base of *M* would have been instantiated if *P* had not been instantiated ($\sim P \mathrel{\Box\!\!\rightarrow} \sim\cup\mathbf{P}_M$). It is trivially true that (ii) necessarily, if no realizer of *M* had been instantiated, then *P* would not have been instantiated ($\Box[\sim\cup\mathbf{P}_M \supset \sim P]$). From (i), (ii), and (10) we get (18). Premise (17) ($\sim M \mathrel{\Box\!\!\rightarrow} \sim\cup\mathbf{P}_M$) follows from (12) and the assumption that if neither *M* nor *P* had been instantiated, then no realizer of *M* would have been instantiated ($\sim M \,\&\, \sim P \mathrel{\Box\!\!\rightarrow} \sim\cup\mathbf{P}_M$).

[31] For further discussion of the coincidence worry, see Carey 2011. See also Sider 2003.

[32] As is witnessed by so-called reverse Sobel sequences, there can be a tension between a counterfactual $\phi \,\&\, \psi \mathrel{\Box\!\!\rightarrow} \chi$ and a counterfactual $\phi \mathrel{\Box\!\!\rightarrow} \sim\chi$ if they are asserted in this order (see von Fintel 2001 and Gillies 2007). Following Moss (2012), I think this tension is best explained as a pragmatic phenomenon, and at any rate it does not seem to arise in our case: it seems fine to say, for instance, that *P** would have been instantiated if *P* had been instantiated in the absence of *M*, while also saying that *P** would not have been instantiated if *M* had not been instantiated.

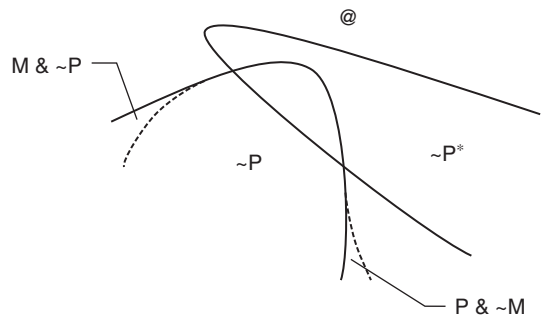

Figure 4.1. Overdetermination with counterfactual dependence

indicated by the dashed lines. The figure shows a situation in which (O_1*) and (O_2*) are non-vacuously true while the *P**-instance also counterfactually depends on both the *M*-instance and the *P*-instance.

While mental causation involves both counterfactual dependence and overdetermination if we assume (EXCLUSION*), in prototypical cases of overdetermination such as the firing squad, 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 physical 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, such that (NON-OVERDETERMINATION) can be rejected.[33]

I have argued that the exclusion problem can be solved even if we grant that the physical effects of mental causes are overdetermined, because the resulting cases of overdetermination are very dissimilar to prototypical cases of overdetermination such as the firing squad. It might be objected that such a solution misses the point of the exclusion problem. The point,

[33] It might seem that cases of mental causation by counterfactual dependence are dissimilar to prototypical cases of overdetermination like firing squad cases in a further respect. Firing squad cases, it seems, involve separate transfers of energy from each of the shooters to the victim, but cases of mental causation by counterfactual dependence need not involve such transfers; in particular, they need not involve a separate transfer of energy from the mental cause to the physical effect. The claim that firing squad cases involve separate transfers of energy is less straightforward than it seems, however, for various stages in the process involve double prevention, for instance, the shooters' muscle contractions as they pull the trigger and the guns' operations (see Schaffer 2000a, 2004a, and Sections 1.6 and 2.6).

the objection goes, is that it is unintelligible how physical effects can have distinct mental causes if they already have physical causes.[34] To put the point more positively, the challenge posed by the exclusion problem is to explain how distinct mental events can make a causal contribution to the physical world in light of the fact that physical effects already have physical causes.[35] According to the objection, this challenge is not met if it is merely pointed out that cases of mental causation are dissimilar to prototypical cases of overdetermination.

A version of this objection is due to Sara Bernstein.[36] Bernstein argues that the problem that arises from the overdetermination of physical effects by mental and physical causes is not that the resulting cases are like prototypical cases of overdetermination. Rather, she holds, the problem is that we cannot give a precise explanation of the individual contribution of the mental cause. In particular, according to her objection, non-reductive physicalists have a hard time answering the following two questions: first, the question of where the extra causal power of the mental 'comes from', that is, the source of the extra causal contribution of the mental; second, the question of where the extra causation by the mental event 'goes', that is, how exactly it contributes to the outcome.[37] The corresponding questions are much more straightforward to answer about prototypical cases of overdetermination, Bernstein holds. Given a firing squad of two, for example,[38] it can neatly be explained where the extra causation 'comes from': it is there because there is a second shooter. It can also be neatly explained where the extra causation 'goes': the victim's heart is hit with twice the force, owing to the presence of the second bullet. Thus, according to Bernstein, dissimilarity to prototypical cases of overdetermination is a problem rather than an advantage for theories of mental causation, because it makes it harder for them to explain where mental causation 'comes from' and where it 'goes'.

My account of mental causation has no difficulty in explaining the causal contribution of the mental, however. According to the account, mental events have physical effects because certain physical events

[34] See Kim (1998: 53), Morris (2015), and Bernstein (2016). For critical discussion, see Árnadóttir and Crane 2013.

[35] Thus, the present challenge has similarities both with the first issue and with the second issue from Section 4.1.

[36] See Bernstein 2016. Bernstein puts her objection forward in the context of transference theories of causation, but it can equally well be made without this presupposition.

[37] See Bernstein 2016: 30–31. The scare quotes are hers.

[38] I'm substituting our example for Bernstein's here; she uses the case of two rocks that shatter a window at the same time.

counterfactually depend on mental events. This counterfactual dependence is in turn explained by the fact that a given physical event would not have occurred if no realizer or base of a given mental property-instance had been instantiated, together with the intimate modal relation between the mental property-instance and its realizers or bases. Thus, we can give a perfectly satisfying explanation of where the mental causes 'come from'. To be sure, the source of the mental cause is not an extra object, unlike the source of the extra causation in the firing squad case, where the extra causation is due to the presence of the second shooter. But it would be gratuitous to demand that any explanation of where the contribution of an overdetermining cause 'comes from' mirror the explanation in the firing squad case in this respect. For one thing, there are satisfying explanations of where the contribution of an overdetermining cause 'comes from' that do not involve a distinct object but merely a distinct feature of a single object. A bee's attraction to a flower might be overdetermined by the smell and the colour of the flower. Explaining the contribution of, say, the smell does not require invoking an extra object similar to the second shooter in the firing squad case. For another, explaining the contribution of the mental cause by explaining the counterfactual dependence of the physical effect on it seems perfectly satisfying without invoking a distinct object (which, in our context, could only be a Cartesian soul or something of that kind).

Where does the causation by the mental event 'go'? The account of mental causation I have defended suggests an answer similar to the answer to the first question: the mental event brings about the physical event by counterfactual dependence. This counterfactual dependence can in turn be explained by certain facts about the counterfactual relation between the realizers or bases on the one hand and the physical event on the other, plus the intimate modal relation between the mental event and the realizers or bases. Again, this explanation is somewhat disanalogous to the explanation of where the extra causation 'goes' in the firing squad case. There, the extra cause modifies the effect, which involves the victim's heart being hit with twice the force compared to a case where only one shooter fires.

As was the case with the explanation of where the causation by the mental event 'comes from', it is not a problem that the explanation of where it 'goes' does not perfectly mirror the corresponding explanation for the firing squad case. Again, the explanation seems satisfying in spite of this difference. And again, there are unproblematic cases of overdetermination that are like mental causation in that an overdetermining cause does not modify the overdetermined event either. Take, for instance, an overdetermination case that involves idealized neurons. Neurons *c* and *d* both fire. Each has an

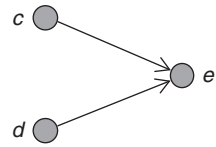

Figure 4.2. A neuron case of overdetermination

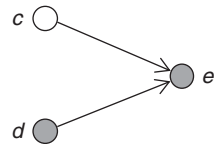

Figure 4.3. If *c* had not fired . . .

excitatory connection to neuron *e*, which also fires (see Figure 4.2). The firings of *c* and *d* overdetermine the firing of *e*. We can make perfect sense of where the causation of, say, *c* 'goes'. We can do so, at least in part, because of the counterfactuals that are true: if only *c* had fired, then *e* would still have fired; if only *d* had fired, then *e* would still have fired; if neither *c* nor *d* had fired, *e* would not have fired; etc. Still, in the idealized neuron scenario, if *c* had not fired, then *e* would not merely still have fired, but would have fired in exactly the same way that it actually did (see Figure 4.3). We can make perfect sense of where the causation by *c* goes, but *c* does not modify the overdetermined event in the way that one shooter's firing modifies the way in which the victim dies.

Admittedly, mental causation on my account differs both from the firing squad and from the neuron example in that the overdetermining mental and physical causes are difficult to disentangle. Given non-reductive physicalism, it is metaphysically impossible for the physical cause to occur without the mental cause. Given super-nomological dualism, such a situation is possible, but it is more remote from actuality than a violation of the ordinary laws of nature. One might worry that the intimate modal relation between the mental cause and its realizer or base threatens the causal contribution of the mental *qua* mental and makes it parasitic on the contribution of its realizer.[39] But this worry can be allayed. As we saw in Section 2.4, the mental cause is not parasitic on the physical cause in the sense that the physical cause is a causal intermediary. The set of

[39] Bernstein (2016: 31–32) expresses a similar worry.

realizers or bases of the mental cause is a logical intermediary when the counterfactual dependence of the physical effect on the mental cause is derived, but this is not detrimental to the genuine efficacy of the mental cause. Further, as we saw in Section 2.2, the counterfactual dependence of the physical effect on the mental cause also shows that the mental cause is efficacious *qua* mental, because, given the strong Kimian conception of events, the counterfactual dependence is due to a general feature of the mental event, namely its constitutive mental property.

In summary, the account of mental causation that I have defended can meet further objections even when we grant that the physical effects of mental causes are overdetermined. It can explain where the contribution of the mental cause 'comes from' and where it 'goes', and it can uphold that the mental cause is efficacious *qua* mental.

The upshot of this section is that non-reductive physicalists and super-nomological dualists may grant that mental causation involves overdetermination without running into trouble. They can deny that cases of mental causation are in any interesting sense like prototypical cases of overdetermination, such as firing squad cases. Unlike the firing squad, cases of mental causation involve an intimate synchronic relation between the simultaneous causes, which also partially explains why the mental event is a cause of the later physical event. Unlike the firing squad, in cases of mental causation the effect counterfactually depends both on the mental cause and on the physical cause. Non-reductive physicalists and super-nomological dualists can also meet objections that locate the source of the problem not in the similarity of mental causation to firing squad cases, but in the difficulty in explaining the causal contribution of the mental cause or of the mental cause *qua* mental. In sum, there would be nothing objectionable about physical effects' being overdetermined by mental and physical causes.

I should perhaps repeat that there might be no need to establish these claims in the first place. If the argument from the previous section succeeds, non-reductive physicalists and dualists have no need to admit that physical effects of mental causes are overdetermined by their mental and physical causes. But should they turn out to be thus overdetermined after all, no harm would befall non-reductive physicalists and super-nomological dualists.

4.4 The Efficacy of Realizers

In the previous sections, I have made some assumptions about the actual realizer P of our mental property M in the non-reductive physicalist case. I have assumed that P necessitates M and that the P-instance causes the P^*-instance

https://doi.org/10.1017/9781108762717.006 Published online by Cambridge University Press

because the *P**-instance counterfactually depends on the *M*-instance. Indeed, I have tacitly assumed that the *P*-instance is the single physical cause of the *P**-instance that is simultaneous and, as it were, coordinated with the *M*-instance. If there are alternative physical causes of this kind, the arguments I have presented need to be augmented, for, as it stands, they leave it open that our physical effect, the *P**-instance, is overdetermined by the *M*-instance and a physical cause other than the *P*-instance. They also leave it open that *that* instance of overdetermination is particularly objectionable.

Are the assumptions from the previous sections justified? We took property *P* to be the property of having firing c-fibres and property *M* to be the property of having a headache. Consider the first assumption. Does having firing c-fibres necessitate having a headache? On the face of it, the answers seems to be 'Yes', at least by the lights of non-reductive physicalism. But what about a case where the c-fibres are disconnected from the rest of the nervous system (see Bennett 2003: 485, 2008: 291)? And what about a case where the laws of physics are radically different, such that (say) the electrical impulses that actually travel down the c-fibres in an orderly fashion randomly pop up and disappear? It might seem that, in either case, there would be no headache.[40] Let us grant that there would indeed be no headache in either case. Then we can no longer say that having firing c-fibres by itself necessitates having a headache. But, it seems, having a headache still is necessitated by a conjunctive property that has having firing c-fibres as a conjunct, the other conjuncts being that the laws of physics are such-and-such and that suitable background conditions obtain (such as the c-fibres' being appropriately connected to the rest of the nervous system). Call this conjunctive property a *total realizer,* P_{total}, of headaches. Call the c-fibre firing itself, that is, the total realizer minus the other conjuncts, a *core realizer,* P_{core}, of headaches.[41] (Non-reductive physicalists might claim that it was the total realizer that they had in mind all along when they talked about the c-fibre firing. Nonetheless, there is also a more narrow sense of 'c-fibre firing', so the distinction between total and core realizers is useful irrespective of the antecedent meaning of 'c-fibre firing'.)

[40] It is not clear whether realizer functionalists such as Lewis (1966) would agree that no pain is instantiated when the disconnected c-fibres fire.

[41] The terminology – strictly speaking, the terminology of properties that are 'core realizations' vs properties that are 'total realizations' – is due to Shoemaker (1981: 97). Bennett (2003, 2008) makes a similar distinction but does not use the same terminology. She holds that the additional properties that are required for what we have called the core realizer to necessitate the mental property are exactly those that are required for the core realizer to bring about the physical effect. Her suggestion is criticized by Aimar (2011) and Keaton and Polger (2014).

If there is a total realizer and a distinct (but included) core realizer, we must reassess the arguments against (EXCLUSION*) and (NON-OVERDETERMINATION) from the previous two sections. Whether the instance of the actual total realizer P_{total} of the mental property M still qualifies as a cause of the later physical property-instance P^* will be discussed in a moment. The instance of the actual core realizer P_{core} of M certainly has a good claim to being a cause of the P^*-instance. Not least because it seems that the P^*-instance counterfactually depends on the P_{core}-instance: if the c-fibre firing – that is, the c-fibre firing in the narrow sense, not including laws or background conditions – had not occurred, then my hand would not have moved towards the aspirin.

The argument against (EXCLUSION*) – at any rate, the argument against (O_1*) from (1) and (2) – is still plausible if we read 'realizer' as 'core realizer' throughout. The argument against (NON-OVERDETERMINATION) also still works in this case. While P_{core} no longer necessitates M, there is still an intimate synchronic relation between the P_{core}-instance and the M-instance that is absent in prototypical cases of overdetermination, for P_{core} is a constitutive part of a related property (namely P_{total}) that necessitates M. And if the P^*-instance is overdetermined by the M-instance and the P_{core}-instance, there is still counterfactual dependence between the individual overdetermining events and the overdetermined event, which also makes the case very dissimilar to prototypical cases of overdetermination such as the firing squad.

Thus, having a causally efficacious core realizer by itself does not revive the exclusion problem. What if the instance of the total realizer as well as the instance of the core realizer is causally efficacious? The previous two sections tacitly assumed the actual realizer to be the actual total realizer of the mental property, because we assumed the realizer to necessitate the mental property. If the arguments from those sections are sound, the P^*-instance is not overdetermined by the P_{total} -instance and the M-instance, or at least not overdetermined in any objectionable way. We just saw that, *mutatis mutandis*, the same holds for P_{core} and M. What about the pair P_{total} and P_{core}? Perhaps the instances of P_{total} and P_{core} overdetermine the P^*-instance. But, as is the case with M and P_{total}, there is an intimate synchronic relationship between P_{total} and P_{core}, for P_{total} has P_{core} as a conjunct and thus necessitates it. If a necessary connection between the overdetermining property-instances is enough to dispel worries about the case's involving a particularly objectionable kind of overdetermination

with respect to M and P_{total}, it should also dispel any such worries in the case of P_{total} and P_{core}.[42]

While it seems plausible that the instance of the actual core realizer is a cause of the later physical event, there are reasons for doubting that the instance of the actual total realizer is in fact capable of causing the later physical event. The possibility of the total realizer's thus being causally inert has advantages and disadvantages for non-reductive physicalists. On the one hand, it attenuates problems that arise from the possible overdetermination of the physical effect: the fewer potential overdeterminers, the better. On the other hand, there is a danger that the mental property inherits the causal inertia of the total realizer. In what follows, I shall argue that the reasons for thinking that the instance of the total realizer is causally inert are at best inconclusive and that, even if it is thus inert, it does not follow that the instance of the mental property is likewise inert.

One might doubt that the instance of the actual total realizer causes the later physical event if one doubts that the later physical event counterfactually depends on the instance of the total realizer. The total realizer P_{total} is a conjunctive property some of whose conjuncts are about the laws of physics. Assume that one of the conjuncts of P_{total} is the property of being such that the laws of electricity are such-and-such. That conjunct fails to be instantiated at a world w which is like the actual world in particular fact until some time in the distant future, when a small violation of the laws of electricity occurs; hence, in w, P_{total} is not instantiated at the time at which P_{total} is instantiated in the actual world. Moreover, by the miracles approach, w comes out closer to the actual world than any other worlds where P_{total} is not instantiated (provided no big miracle occurs in those worlds), for w has more perfect match of particular fact with the actual world than those worlds. But P^* is still instantiated at w, because w does not differ from the actual world in particular fact until well after P^* is instantiated. Thus, 'If P_{total} had not been instantiated, then P^* would not have been instantiated' is false, and the P^*-instance does not counterfactually depend on the P_{total}-instance.

[42] Bennett (2003) is worried about the possible overdetermination of the physical effect by the instances of what I have called P_{total} and P_{core} too, but her own account of overdetermination should in fact dispel the worries even more straightforwardly. Bennett holds that what is required for overdetermination is not merely the truth of the claims from the $(O_1)/(O_2)$ family, but their non-vacuous truth. Owing to the necessitation of P_{core} by P_{total}, the counterfactual 'If P_{total} had been instantiated without P_{core}, then P^* would still have been instantiated' comes out vacuously true, so, by Bennett's standards, the P^*-instance is not overdetermined by the P_{total}-instance and the P_{core}-instance in the first place. See also Sider 2003.

It does not follow, however, that the P_{total}-instance does not cause the P^*-instance, for we did not assume counterfactual dependence to be necessary for causation. Moreover, the argument against the counterfactual dependence of the P^*-instance on the P_{total}-instance can be questioned. The argument assumes that the conjunct of P_{total} that is about laws of physics is about those laws' holding everywhere and at any time. But if the purpose of the nomic conjunct is to ensure that the realizer necessitates the mental property, it suffices for it to require that the laws of physics are such-and-such more locally, that is, in and perhaps around the space–time region where the core realizer is instantiated. The local failure of the laws of physics would still suffice to prevent the total realizer from being instantiated, but perhaps it likewise suffices to prevent the later physical event from occurring. If the laws of electricity had been different while and where my c-fibres fired, perhaps my hand would not have moved towards the aspirin after all.

Setting the issue of counterfactual dependence aside, one might deny that the P_{total}-instance causes the P^*-instance if one thought that instances of P_{total} are *per se* incapable of causing anything. Thus, one might think that P_{total}, just like the property of shattering-in-one-minute (see Section 1.5), fails to be a causal property. Why should one think so? One of the conjuncts of P_{total} is the property that the laws of physics are such-and-such. It might seem that instances of this kind of property can never cause or be caused by anything.[43] It might seem, further, that being a non-causal property is closed under conjunction:

(CLOSURE-&) If the property of being F is non-causal, then the property of being F and G is non-causal for any G.[44]

Given that the nomic conjunct of P_{total} is non-causal, it follows from (CLOSURE-&) that P_{total} itself is non-causal.

If (CLOSURE-&) is true, the causal inertia of the nomic conjunct of P_{total} spreads to P_{total} itself. Douglas Keaton thinks that it spreads even further. He holds that being a non-causal property is closed not merely under conjunction, but under disjunction as well (Keaton 2012: 253):

[43] McLaughlin (2009) holds this with respect to Shoemaker's 'causal laws' (2007: 6). Christensen and Kallestrup (2012) build on McLaughlin's argument; they focus on the issue of whether realizers can be effects (see Section 2.4).

[44] Keaton (2012: 251) endorses the principle that a conjunctive property is causal just in case each of its conjunct properties is causal. This principle entails (but is not entailed by) (CLOSURE-&). McLaughlin (2009) seems to assume something like (CLOSURE-&) too.

(CLOSURE-V) If the property of being *F* is non-causal, then the property of being *F* or *G* is non-causal for any *G*.

Given (CLOSURE-V) and given that P_{total} is non-causal, the disjunctive property that contains P_{total} and all other possible total realizers of *M* as disjuncts is non-causal too. By the strong supervenience of mental properties on physical properties, the instantiation of that disjunctive property is strictly equivalent to the instantiation of *M*. One might or might not be inclined to identify properties whose instantiations are strictly equivalent,[45] but at any rate it seems that a property that is thus strictly equivalent to a non-causal property fails to be causal itself:

(CLOSURE-≡) If the property of being *F* is non-causal and necessarily *F* is instantiated just in case property *G* is, then *G* is non-causal.

Given (CLOSURE-≡) and given that the disjunctive property that contains P_{total} and all other possible total realizers of *M* as disjuncts is non-causal, *M* comes out non-causal too.

In sum, assuming that the nomic conjunct of P_{total} is non-causal, one can use the principles (CLOSURE-&), (CLOSURE-V), and (CLOSURE-≡) to argue that *M* is non-causal too.

It is easy to show that there must be something wrong with that argument. All we need to show this is the existence of a property *F* such that both *F* and its complement, that is, the property of being not-*F*, are non-causal. Such a property, it seems, is not hard to find.[46] A property that everything necessarily instantiates, such as the property of being self-identical, is a good candidate, for it seems that neither instances of this property nor instances of the complementary property of failing to be self-identical (which, of course, do not exist, because the complementary property cannot ever be instantiated by anything) can cause anything. Alternatively, we can take a property that is randomly distributed over all

[45] See Kim 1992 and Keaton 2012: 252–253 for discussion.

[46] Such a property would be very easy to find if being a non-causal property were closed under negation. This additional closure principle would yield a lot of causation by omission, however, which some might find objectionable. Here is how the commitment to causation by omission would come about. Assume that the property of being not-*F* is not causal. By the envisaged closure of being non-causal under negation, it follows that the property of being not-not-*F* is not causal. By (CLOSURE-≡), it follows that the property of being *F* is not causal (because the properties of being *F* and of being not-not-*F* are necessarily co-instantiated). In sum, if the property of being not-*F* is not causal, then the property of being *F* is not causal. Contrapositively, if the property of being *F* is causal, so is the property of being not-*F*.

actual and possible individuals. Such a property and its complement have a good claim to being non-causal. Suppose, then, that both the property of being *F* and the property of being not-*F* are non-causal. Let *G* be any causal property. By (CLOSURE-&), the property of being both *G* and *F* is non-causal; likewise for the property of being both *G* and not-*F*. By (CLOSURE-∨), the property of being either both *G* and *F* or both *G* and not-*F* is non-causal too. Necessarily, something instantiates the property of being either both *G* and *F* or both *G* and not-*F* just in case it instantiates the property of being *G*. So by (CLOSURE-≡), the property of being *G* is non-causal (substitute 'the property of being either both *G* and *F* or both *G* and not-*F*' for 'the property of being *F*' in (CLOSURE-≡)). But we assumed *G* to be causal: contradiction.

Thus, the argument that purports to show that *M* is non-causal fails. Assuming that it is the same feature that is supposedly responsible for the nomic conjunct's being non-causal and for P_{total}'s being non-causal, we can locate the failure of the argument more precisely if we know more about why exactly the nomic conjunct of P_{total} is supposed to be non-causal. Perhaps the source of the nomic conjunct's being non-causal is that the property of being such that the laws of physics are such-and-such is extrinsic. It seems plausible that extrinsicality is closed under conjunction and strict equivalence.[47] It also seems plausible that extrinsicality is closed under complementation: if the property of being *F* is extrinsic, so is the property of being not-*F* (see Lewis 1983: 199). These features of extrinsicality allow us to construct a counterexample to the closure of extrinsicality under disjunction. Take the properties of being accompanied, of being unaccompanied, and of being square. The property of being accompanied is extrinsic. By the closure of extrinsicality under complementation, so is the property of being unaccompanied. By the closure of extrinsicality under conjunction, the property of being square and accompanied is extrinsic, as is the property of being square and unaccompanied. If extrinsicality were closed under disjunction, the property of being either both square and accompanied or both square and unaccompanied would be extrinsic, but by the closure of extrinsicality under strict equivalence, if this property were extrinsic, so would the equivalent property of being square, which is not extrinsic.[48] So extrinsicality cannot be closed under disjunction. If extrinsicality is the source of the nomic

[47] At least the closure of extrinsicality under strict equivalence seems plausible on the orthodox view that intrinsicality and extrinsicality are non-hyperintensional. Hyperintensional accounts of intrinsicality and extrinsicality such as Bader's (2013) may reject that kind of closure.

[48] The example is from Lewis (1983: 200), whose assumption of the closure of extrinsicality under conjunction is tacit; he also tacitly assumes that properties that are strictly equivalent are identical.

conjunct's being non-causal, the argument for the total realizer's being non-causal fails because (CLOSURE-V) fails. (Notice that, if extrinsicality is behind being non-causal in that argument, the counter-argument is strengthened, for given the closure of extrinsicality under complementation, it is straightforward to find a property such that both the property and its complement are extrinsic and, thus, non-causal.)

We need not, however, concede in the first place that total realizers are non-causal properties because of the supposed extrinsicality of the nomic properties they contain as conjuncts. For one thing, the nomic properties need not be extrinsic at all. If the laws of nature can vary independently of whether the bearer of our nomic property is alone in the universe or not, the nomic properties might come out intrinsic (see Langton and Lewis 1998: 339). Even if they come out extrinsic, they might still be causal. Extrinsicality is a matter of degree. The property of being a sibling, say, is more extrinsic than the property of being a brother (see Lewis 1983: 197). In Section 1.5, we merely required that a property be *sufficiently* intrinsic (and temporally intrinsic) in order to be causal. Even if they are on the extrinsic side of the spectrum, nomic properties might meet this requirement. (Analogously, someone might be rather short, yet sufficiently tall to do a certain task.)

The total realizers contain properties about background conditions as conjuncts besides the core realizers and the nomic conjuncts. The properties about background conditions are likely to qualify as extrinsic too and thus are another potential source of the total realizers' being non-causal. Like the nomic conjuncts, however, the mere extrinsicality of the properties about background conditions does not imply that they are too extrinsic to be causal, for they might still be *sufficiently* intrinsic.

So far we have considered the distinction of core vs total realizers in the context of non-reductive physicalism. Is an analogous distinction needed in the case of super-nomological dualism? According to super-nomological dualism, the relation between the bases of mental properties and the mental properties themselves is a matter of psychophysical laws. In other words, the instantiation of a physical base together with the psychophysical laws necessitates the instantiation of the mental property. It seems that, like the necessitation of the mental property by its realizers according to non-reductive physicalism, this necessitation holds only if the physical base is a conjunctive property that includes properties pertaining to background conditions and properties pertaining to 'ordinary' laws of nature as conjuncts, besides a physical base property in the narrow sense. Indeed, it seems that, for a given mental property, this base property in the narrow sense is identical to what non-reductive physicalists take to be the core

realizer of the mental property, and it seems that the conjunctive property is identical to what non-reductive physicalists take to be the total realizer of the mental property. Thus, super-nomological dualists should make the same kind of distinction as non-reductive physicalists and apply it to the very same physical properties. Super-nomological dualists could call the non-reductive physicalists' total realizers *total bases* of mental properties and the non-reductive physicalists' core realizers *core bases*.

The issues ramify for super-nomological dualists just as they did for non-reductive physicalists. The solutions are analogous too. The instance of the core base of the mental property causes the later physical effect by counterfactual dependence. So does the instance of the mental property. Thus, if the later physical effect is overdetermined, the case is very dissimilar to prototypical cases of overdetermination. It is also dissimilar to prototypical cases of overdetermination by virtue of the intimate synchronic relation between the instance of the core base of the mental property and the instance of the mental property itself. While the instantiation of the core base no longer implies the instantiation of the mental property with super-nomological necessity, the core base is a constitutive part of another property (namely the total base) that does imply the instantiation of the mental property. The question of whether the instance of the total base is capable of causing anything receives the same answer as the question of whether the total realizer is capable of causing anything, for we are dealing with one and the same property. The question of whether the mental property somehow inherits the (putative) causal inertia of the total base is even more likely to be answered in the negative, however. For super-nomological dualists cannot follow the step of the argument for the causal inertia of realizers for the non-reductive physicalist case that assumed the instantiation of the mental property to be strictly equivalent to the instantiation of the disjunction of its total realizers.[49]

The upshot of this section is that if non-reductive physicalists make a distinction between total realizers, which necessitate the mental properties they realize, and core realizers, which do not, this distinction does not affect the solution to the exclusion problem. In particular, if the instance of a core realizer overdetermines a later physical event together with the instance of a mental property, the case we get is still very dissimilar to prototypical cases of overdetermination such as firing squad cases. The

[49] A proponent of the argument for the super-nomological dualist case could add a nomic conjunct about the actual psychophysical laws' being such-and-such to every disjunct. That would yield the strict equivalence between the mental property and the disjunction of its bases-cum-psychophysical-laws. This manoeuvre might seem to turn the disjuncts into rather arbitrary properties, however. At any rate, the counter-argument for the non-reductive physicalist case stands undefeated.

total realizer might or might not turn out to be a non-causal property, but if it does, this does not affect the efficacy of the mental property. Similarly for super-nomological dualists who distinguish between core bases and total bases of mental properties.

It is worth repeating a point from Section 2.4. The argument for mental causation from Chapter 2 that showed the P^*-instance to be caused by the M-instance is completely independent of whether or not total realizers/bases are causal properties. The argument used the disjunction of M's (total) realizers merely as a logical intermediary, not as a causal one. Thus, it does not require instances of total realizers to be capable of causing anything. And the property that is involved in the effect, P^*, need not realize anything in the first place, so the causal relation between the M-instance and the P^*-instance can be established independently of whether total realizers/bases turn out to be causal.

4.5 Sufficient Causes

In Section 4.1, I formulated the exclusion problem in terms of causation without qualifying the kind of causation that is supposed to be in play. Thus, (EFFICACY) says that some mental events cause physical events; (COMPLETENESS) says that every physical effect has a physical cause; and (EXCLUSION) says that no effect has more than one cause at a given time, unless it is overdetermined. Often the exclusion problem is formulated not in terms of causation *simpliciter*, but in terms of *sufficient* causation. Thus reformulated, the principle corresponding to (EFFICACY) says that some mental events are *sufficient* causes of physical events; the principle corresponding to (COMPLETENESS) says that every physical effect has a *sufficient* physical cause; and the principle corresponding to (EXCLUSION) says that no effect has more than one *sufficient* cause at a given time, unless it is overdetermined.[50] Friends of formulations of the exclusion problem in terms of sufficient causes would also, I take it, be inclined to replace (EXCLUSION) with a version of the more specific principle (EXCLUSION*), which says that no effect has a mental and a distinct

[50] Formulations of the exclusion problem are found, for instance, in Kim 2005, Bennett 2007, Moore 2012, Carey 2013, and Morris 2015. Sometimes only some of the principles are formulated in terms of sufficient causation and others in terms of causation *simpliciter*. For instance, Bennett formulates the principles corresponding to (COMPLETENESS) and (EXCLUSION) in terms of sufficient causes, but formulates the principle corresponding to (EFFICACY) in terms of causation *simpliciter*. For the claims in Bennett's exclusion problem to be inconsistent, however, the principle corresponding to (EFFICACY) must be read as talking about sufficient causation too. Perhaps hybrid formulations of the problem are possible that still make the principles inconsistent, but I shall confine myself to formulations that talk about sufficient causes throughout.

physical cause at the same time, unless it is overdetermined. Reformulated in terms of sufficient causes, the principle says that no effect has a *sufficient* mental cause and a distinct *sufficient* physical cause that occur at the same time, unless it is overdetermined. In sum, we get the following new version of the exclusion problem in terms of sufficient causes (the principles (DISTINCTNESS) and (NON-OVERDETERMINATION) do not talk about causation, so they retain their original formulations):

(DISTINCTNESS)	All mental events are distinct from physical events.
(EFFICACY-S)	Some mental events are sufficient causes of physical effects.
(COMPLETENESS-S)	Every physical event that has a cause at all has a sufficient physical cause.
(EXCLUSION*-S)	No effect has a sufficient mental cause and a distinct sufficient physical cause that occur at the same time, unless it is overdetermined.
(NON-OVERDETERMINATION)	The effects of mental events are not systematically overdetermined.

Like the principles from the original presentation of the exclusion problem, (DISTINCTNESS), (EFFICACY-S), (COMPLETENESS-S), (EXCLUSION*-S), and (NON-OVERDETERMINATION) are inconsistent, or at least in tension with one another.

This section investigates how the exclusion problem fares if it is formulated by these five principles. We shall see that the nature of the problem depends on what exactly we understand by sufficient causation. Generally, the exclusion problem will turn out to be harder to solve than the problem formulated in terms of causation *simpliciter*. The difficulties do not spread to the original exclusion problem and our solution, however, for we shall see that, on our solution, the original exclusion problem does not entail the version of the problem in terms of sufficient causes.

There are two salient ways of understanding sufficient causation. First, we can understand a sufficient cause as one that transfers a physical quantity on its effect. Second, we can understand a sufficient cause as one that is modally sufficient, in a sense to be spelled out, for its effects. In what follows, I will discuss these different ways of understanding sufficient causes in turn. The discussion of sufficient causation as transference will allow us to build on results from previous chapters.

Let us first consider sufficient causes as characterized in terms of transference. On the face of it, it might seem tempting to flesh out such a characterization by saying that event *c* is a sufficient cause of event *e* if and only if *c* transfers a physical quantity to *e*. But there is an immediate complication. Recall our observation from Section 1.6, according to which there can be transference without causation, as in the case of a rock that is first heated over a fire and then thrown at a bottle, which shatters. The fire transfers a physical quantity (namely heat) on the bottle, but does not qualify as a cause of the shattering; *a fortiori*, the fire does not qualify as a *sufficient* cause of the shattering. So transference cannot be a sufficient condition for being a sufficient cause and hence transference cannot be a necessary and sufficient condition, as our initial characterization has it. We can, however, still formulate at least a partial characterization of sufficient causes in terms of transference by formulating a necessary condition for sufficient causation in terms of transference, that is, by saying that event *c* is a sufficient cause of event *e* only if *c* transfers a physical quantity on *e*. For our purposes, this partial characterization will suffice.

(In what follows, I shall confine myself to discussing sufficient causation as partially characterized in terms of the transference of a physical quantity. *Mutatis mutandis*, the arguments would also apply if we partially characterized sufficient causation in terms of the transference of powers instead.)

Suppose that (DISTINCTNESS), (EFFICACY-S), and (COMPLETENESS-S) are all true. Thus, there is a mental event that transfers a physical quantity on a later physical event. There is also a physical event distinct from, but simultaneous with, the mental event that also transfers a physical quantity on the later physical event; presumably, the first physical event is the realizer or base event of the mental event. Would such a situation be acceptable? It seems hard to make sense of the idea that the mental event transfers an additional amount of the physical quantity on the later physical event, or an additional kind of quantity. Perhaps non-reductive physicalists could try to explain such a double transference by claiming that mental causes and the simultaneous physical causes transfer the same (token) dose of the quantity in question (see Bennett 2008: 294). The mass of a statue and the mass of the lump of clay that constitutes the statue do not add up; rather, the statue and the lump seem to have the same (token) mass. Similarly, non-reductive physicalists might claim, when a physical cause and a distinct mental cause transfer a quantity to a physical effect, this does not require that the mental cause transfer a distinct (token) dose of that quantity. Whether we think of this response as a denial of the (EXCLUSION*-S) principle or the (NON-OVERDETERMINATION) principle does not matter for the dialectic; what does matter is that the resulting

situation where the mental cause and the physical cause transfer the same (token) dose of the physical quantity to the physical effect is claimed not to be metaphysically problematic.

I will leave it open how plausible this claim is, but things certainly get worse for mental causation with sufficient causation as transference. Whatever its merits given non-reductive physicalism, the 'two transfers, one dose' response is not open to dualists, super-nomological or otherwise. By dualists' lights the relation between the mental cause and the physical cause is less intimate than the relation between a mental property-instance and its realizer-instance is according to non-reductive physicalism, because the instance of the base of the mental property no longer necessitates the instance of the mental property. Given that the connection between the mental property-instance and its physical base is contingent, it is hard to see how dualists could still claim that the mental cause and the physical cause transfer one and the same (token) dose of a physical quantity on the effect. (Compare: if, *per impossibile*, the existence of the lump of clay no longer necessitated the existence of the statue, could one still claim that their masses did not add up?)

Further, given the characterization of sufficient causation as transfer, non-reductive physicalists and dualists have a hard time accommodating certain cases which, it seems, have as good a claim as any to be cases of mental causation. In particular, they have a hard time accommodating the causation of bodily movements by mental events via muscle contraction. We saw in Section 2.6 that muscle contractions work by double prevention: calcium release at the neuromuscular junction causes the muscle to contract by preventing the obstruction of the binding sites of myosin and actin, which, unless prevented, prevents the muscle contracting. No transfer takes place between the calcium release and the muscle contraction. Thus, the calcium release is not a sufficient cause of the contraction. Moreover, there can be no chain of events that are connected by sufficient causation that contains the link between the calcium release and the contraction as a (non-redundant) link. But mental events can only be sufficient causes of bodily movements through such a chain. Therefore, mental events cannot be sufficient causes of bodily movements. Strictly speaking, this result is not a denial of (EFFICACY-S), because mental events might still be sufficient causes of physical events further upstream on the causal chain, but it comes close enough to epiphenomenalism to be unacceptable.[51]

[51] For further discussion, see Russo 2016.

https://doi.org/10.1017/9781108762717.006 Published online by Cambridge University Press

A parallel argument shows that the instances of realizers or bases of mental properties, such as my c-fibre firing, cannot be sufficient causes of bodily movements either, for they, too, could only achieve this via a chain of sufficient causes that contains the link between the calcium release and the contraction as a (non-redundant) link. On the one hand, the result that realizers or bases are not sufficient causes of bodily movements attenuates the difficulties of explaining how physical effects such as those movements can have both a sufficient physical cause and a sufficient mental cause: as far as these difficulties are concerned, the fewer sufficient causes, the better. But this is not much of a consolation, for by itself the result that realizers or bases cannot be sufficient causes of bodily movements is very implausible. So is the parallel result that mental events cannot be sufficient causes of bodily movements. Overall, the situation looks bleak if sufficient causation requires transfer.

Fortunately, the trouble does not carry over to the counterfactual account of mental causation that I have defended. For we saw that double-prevention cases such as muscle contractions not only show that there is no transfer in what seem to be genuine cases of causation, but also show that there are cases of counterfactual dependence (and hence of causation) without transfer. Thus, mental causation on my account does not entail sufficient causation in a sense that requires the transfer of a physical quantity. Nor does my account entail that the physical causes of bodily movements are sufficient causes of these movements in a sense that requires transfer. Since neither the mental cause of a physical effect nor the physical cause of this effect needs to transfer anything on the effect, there is also no problem of explaining how *both* the mental cause and the physical cause could transfer a physical quantity on the effect.

Let us now consider sufficient causes that are characterized not in terms of the transfer of a physical quantity, but in terms of modal sufficiency. The idea behind this characterization is that the occurrence of sufficient causes implies the occurrence of their effects, at least in a suitable range of circumstances. In what range of circumstances? Certainly not in all possible circumstances, for otherwise there is never causation between distinct events that are sufficiently (temporally) intrinsic, for it is always possible for such distinct events to occur separately. How about restricting the relevant circumstances to nomologically possible ones? In other words, how about characterizing sufficient causes as events whose occurrence implies the occurrence of their effects with nomological necessity?[52]

[52] If some events imply other, *simultaneous* events with nomological necessity, this characterization would yield spurious cases of simultaneous causation. Even worse, we will get cases of backward

On such a characterization, only very big events would qualify as sufficient causes. For unless the causes are made very big, it is always nomologically possible for something to interfere and prevent the effect. Take, for instance, my throwing a dart at a balloon, which causes the balloon to burst. It might seem that my throwing is a sufficient cause of the balloon's bursting. But it does not, in conjunction with our laws of nature, entail that the balloon bursts. Something might interfere. And the possible interference is not limited to the actions of bystanders. Suppose that it takes a bit more than a second for my dart to reach the balloon. Then a strong laser beam sent at the time of my throw from one light-second away could have destroyed the dart and prevented the balloon from bursting. Actually, no such thing happens, but to make sure that no such thing happens in any worlds where our laws of nature hold and the sufficient cause of the bursting occurs, we need to make the sufficient cause big enough to include all the space–time points that are a potential source of interference. Given that such interference propagates at or below the speed of light, we would need to make the sufficient cause big enough to include a cross-section of the effect's past light cone that includes my throw (see Loewer 2007: 253–254). Such a cross-section is big indeed. It needs to include everything that is going on at the time of my throw as far as a light-second away from it – that is, everything that is going on within a radius of 300,000 kilometres.

It seems implausible that only very big events can be sufficient causes. We therefore need a characterization of them that does not require sufficient causes to imply their effects with nomological necessity, but still somehow deals with potential interfering factors. It seems promising to use counterfactuals for that purpose, by characterizing a sufficient cause as an event such that, had it occurred, the effect would have occurred. The advantage of using counterfactuals is that they take care of the potential interfering factors. If none of them are actually present, then, it seems, none of them are present in the closest worlds where the cause occurs. Notice that, while the present suggestion also uses counterfactuals, it is very different from our sufficient condition for causation in terms of counterfactual dependence. That sufficient condition uses counterfactuals of the form 'If this event had *not* occurred, then that event would *not* have occurred', while sufficient causation, according to the suggestion, is spelled

causation whenever there is backward nomological necessitation. Therefore it seems sensible to restrict the characterization to pairs of events where one occurs later than the other, as we did in previous chapters.

out by using counterfactuals of the form 'If this event *had* occurred, then that event *would have* occurred.'

The present suggestion is in need of refinement. Given that the putative sufficient cause and its effect both occur, it is trivially true that the effect would have occurred had the cause occurred, given Lewis's truth-conditions for counterfactuals. At least this is trivially true if we assume that the actual world is closer to itself than any other world is (see Lewis 1973b: 26–31). Given this assumption, the actual world, where the antecedent of our counterfactual is true, is the closest antecedent-world. In the actual world, the consequent of our counterfactual is also true, so the counterfactual is true. In order to avoid this triviality, we should demand not that the effect would have occurred if the cause had occurred, but that the effect would also have occurred if the cause had occurred in different circumstances.

What are the relevant different circumstances? It is tempting to say that the relevant different circumstances are sufficiently similar to the actual circumstances and that they should not involve interfering factors, but that is somewhat vague, and there is a danger of circularity if 'interfering factors' are in turn defined in terms of sufficient causes. A more promising suggestion is that the relevant circumstances are those where there are no other factors in play whatsoever, interfering or otherwise. In other words, the suggestion is that if the sufficient cause had been the only event occurring at the time at which it actually occurred, then the effect would have occurred as well.[53]

Again, this suggestion needs refinement, for few causes would have brought about their effects all by themselves. For instance, if I had thrown the dart, but there had been no gravitational field, then I might have missed the balloon. If I had thrown the dart, but there had been no balloon, then the balloon's bursting would not have occurred. We could build the various factors that are required for the effect's occurrence into the cause, but then causes again become too big and cumbersome. A more elegant alternative is to characterize sufficient causes in terms of membership of a set of simultaneous events that is sufficient to bring about the effect in the sense that, if all the members of the set had occurred and no other contemporaneous events, then the effect would still have occurred. (In the following, I will use the notion of a set of events' being *sufficient* for an event in this sense.) The

[53] This suggestion, together with the further refinement discussed below, is due to Paul and Hall 2013: 14–16.

members of the set besides the sufficient cause can represent background conditions, such as the presence of a gravitational field and of the balloon.

Membership in a set of events that is sufficient for the effect is not quite what we need yet. Often such sets can be enlarged with further events while retaining the collective sufficiency for the effect. For instance, if my throw, the gravitational field, etc. are sufficient for the balloon to burst, presumably the president's drinking tea, my throw, the gravitational field, etc. are also sufficient for the balloon to burst. But it seems odd to say that the president's drinking tea is a sufficient cause of the balloon's bursting.[54] In the example, the difference between the enlarged set and the original set is that the enlarged set minus the president's drinking tea is still sufficient for the balloon to burst, while the original set minus my throw is no longer sufficient for the balloon to burst. This suggests that what matters for being a sufficient cause is not membership in a set of events that are collectively sufficient for the effect, but membership in a set that is just big enough to be sufficient for the effect.[55] Let us define a set of events that occur at the same time as *minimally sufficient* for an event *e* if and only if, first, *e* would have occurred if all the members of the set had occurred, but no other contemporaneous events, and, second, it is not the case that *e* would have occurred if only some, but not all of the members of the set had occurred, but no other contemporaneous events (see Paul and Hall 2013: 16). In other words, a set of events is minimally sufficient for *e* if and only if it is sufficient for *e*, but no proper subset of it is sufficient for *e*. We can now state the modal characterization of sufficient causes: a *sufficient cause* of an event *e* is an event that is a member of a set of actually occurring events that is minimally sufficient for *e* (Paul and Hall 2013: 16). This characterization states necessary and sufficient conditions for sufficient causes. For our purposes, it will again be enough to work with a necessary condition. Thus, in what follows I will merely make use of the claim that *if* an event *c* is a sufficient cause of *e*, then *c* is a member of a set of actually occurring events that is minimally sufficient for *e*.[56]

[54] By the failure of counterfactuals to obey the rule of strengthening the antecedent (see Section 1.4), the sufficiency of the larger set does not follow logically from the sufficiency of the original set, but nonetheless there are many cases where both sets are sufficient.

[55] The suggestion is similar in spirit to Mackie's (1965) idea that causes are INUS conditions, that is, conditions that are insufficient but necessary parts of unnecessary but sufficient conditions.

[56] Taking membership in a minimally sufficient set of events to be a sufficient condition for being a sufficient cause would have the implausible result that the background conditions from the set are sufficient causes. By merely assuming the corresponding necessary condition for being a sufficient cause, we are not committing ourselves to this result.

What does the exclusion problem look like if sufficient causation is understood like this? Let us first consider claim (EFFICACY-S), which says that some mental events are sufficient causes of physical effects. We can easily find a set of simultaneous events that contains my headache and that is sufficient for my hand to move towards the aspirin. We can find a set of events, that is, that contains my headache and that is such that my hand would have moved had all members in the set occurred, but no other events at the time in question. This set contains my headache, my c-fibre firing and events that represent various background conditions, such as the presence and integrity of the rest of my body. The question is whether this set is also *minimally* sufficient for my hand's movement. In particular, the question is whether the events in the set minus the headache would still have brought about my hand's movement if no other events had occurred at the time in question.

The answer depends on whether non-reductive physicalism or dualism is true. Consider dualism first. Given dualism, it is metaphysically possible for the c-fibre firing to occur without the headache. If the c-fibre firing had occurred without the headache, but together with the actual background conditions, my hand would still have moved. Thus, the set {c-fibre firing, background conditions} is sufficient for my hand's movement, so the set {c-fibre firing, headache, background conditions}, while also sufficient for my hand's movement, is not minimally sufficient for it. Hence my headache is not a sufficient cause of my hand's movement. More generally, given dualism, mental events cannot be sufficient causes of physical events: (EFFICACY-S) is false, and epiphenomenalism about sufficient causes is true.

Now consider non-reductive physicalism. For simplicity, I will assume my headache to be an instance of a total realizer of headaches, that is, an event whose occurrence necessitates the occurrence of the headache. Given non-reductive physicalism, it is still the case that the set {c-fibre firing, headache, background conditions} is sufficient for my hand's movement. Is there still a danger that the set is not minimally sufficient because its proper subset, {c-fibre firing, background conditions}, is already sufficient?

One might think that there is, for the following reason. It is impossible for the c-fibre firing to occur without the headache. *A fortiori* it is impossible for the c-fibre firing to occur and the background conditions to obtain without the headache. *A fortiori* it is impossible for the c-fibre firing to occur and the background conditions to obtain without the occurrence of *any other* contemporaneous events. Thus, when we consider the counterfactual 'If the c-fibre firing had occurred and the background conditions had obtained, but no other contemporaneous events had occurred, then

my hand would still have moved', we are considering a counterfactual with an impossible antecedent. This counterfactual is (vacuously) true, one might continue to reason; therefore the set {c-fibre firing, background conditions} is sufficient for my hand's movement; therefore the set {c-fibre firing, headache, background conditions} is not minimally sufficient; therefore the headache is not a sufficient cause of my hand's movement.

This way of reasoning leaves few sufficient causes, because it leaves few minimally sufficient sets of events. Consider the following parallel argument: I throw a dart at a balloon, but this time I throw the dart particularly vigorously. The balloon bursts. It seems that my throwing the dart vigorously has a good claim to being a sufficient cause of the balloon's bursting, but so does my throwing the dart *simpliciter* (it does not take a particularly vigorous thrown to burst the balloon).[57] But, the reasoning goes, my throwing the dart *simpliciter* cannot be a sufficient cause of the balloon's bursting. For suitable background conditions, the set {my throwing, my throwing vigorously, background conditions} is sufficient for the balloon to burst. It is impossible for me to throw the dart vigorously without throwing it. *A fortiori* it is impossible for me to throw the dart vigorously and for the background conditions to obtain without the occurrence of *any other* contemporaneous events, such as my throwing the dart *simpliciter*. Thus, it is vacuously true that if I had thrown the dart vigorously, the background conditions had obtained, but *no other* contemporaneous events had occurred, then the balloon would have burst. Therefore, the set {my throwing vigorously, background conditions} is sufficient for the balloon's bursting; therefore, the original set, {my throwing, my throwing vigorously, background conditions}, is not minimally sufficient; therefore, my throwing *simpliciter* is not a sufficient cause of the balloon's bursting.

One might conclude that neither my throwing *simpliciter* nor my headache are sufficient causes. Generalizing from the headache example, non-reductive physicalists then have to deny (EXCLUSION-S) and accept epiphenomenalism about sufficient mental causes, just as dualists had to. But non-reductive physicalists do not have to give in so easily. A natural response to the above arguments, which supposedly show that neither my headache nor my throwing the dart *simpliciter* are sufficient causes, is to

[57] I am not engaging in the use of any 'proportionality' constraint on causation here (see Yablo 1992); my argument merely requires that my throwing the dart *simpliciter* should count as a sufficient cause of the balloon's bursting. For recent discussions of proportionality in the context of mental causation, see Weslake 2013, Harbecke 2014, and McDonnell 2017.

insist on a reading of sufficiency and minimal sufficiency that does not have this consequence.

Here is the reading we should advocate on behalf of non-reductive physicalists who want to claim that there are sufficient mental causes. We should, first, demand that the phrase 'no other contemporaneous events' in the definition of sufficiency and minimal sufficiency not be read as 'no other contemporaneous *whatsoever*', which is the reading used in the above arguments. Rather, we should demand that 'no other contemporaneous events' be read as 'no other contemporaneous events *other than those necessitated by the events in the set in question*'. Unless we use the latter reading, virtually any set of events counts as sufficient for virtually any event owing to the vacuous truth of the relevant counterfactual, which would leave us with few *minimally* sufficient sets of events.

Second, we should not infer lack of minimal sufficiency from the sufficiency of sets that are subsets in name only. For instance, even if we read 'no other contemporaneous events' in the way just suggested, the set {c-fibre firing, background conditions} is still already sufficient for my hand's movements, for if all members of the set had occurred, but no other contemporaneous events other than those necessitated by the events in the set, then my hand would still have moved. But if all members of the set had occurred, but no other contemporaneous events other than those necessitated by the events in the set, then my headache would still have occurred. So the sufficiency of the set {c-fibre firing, background conditions} does not show that the headache is dispensable in bringing about the movement of my hand. If one takes the sufficiency of the set {c-fibre firing, background conditions} to show that the occurrence of *some, but not all* events in the set {headache, c-fibre firing, background conditions} suffices for the movement of my hand, one is again relying on the vacuous truth of the relevant counterfactual, because a situation where the c-fibre firing occurs and the background conditions obtain but not all members of {headache, c-fibre firing, background conditions} occur is impossible.

In order to accommodate these insights, I suggest adopting the following modified definition of minimal sufficiency: a set of events that occur at the same time is *minimally sufficient* for an event *e* if and only if, first, *e* would have occurred if all the members of the set had occurred, but no other contemporaneous events *other than those necessitated by the events in the set in question* and, second, it is not the case *or merely vacuously true* that *e* would have occurred if only some, but not all of the members of the set had occurred, but no other contemporaneous events *other than those necessitated by the events in the subset in question*.

Given the new definition of minimal sufficiency, it is possible for the headache to be a member of a set that is minimally sufficient for the hand's movement, because it is possible for {headache, c-fibre firing, background conditions} to be such a set. At least this is possible in principle and not forestalled by the necessary connection between the c-fibre firing and the headache. The headache can satisfy our necessary condition for being a sufficient cause by being a member of this set.[58] Similarly, the c-fibre firing can satisfy the necessary condition for being a sufficient cause of the hand's movement by virtue of being a member of the set {c-fibre firing, background conditions}, which is also minimally sufficient for the hand's movement. Given the new definition of minimal sufficiency, the sufficiency of the set {c-fibre firing, background conditions} does *not* undermine the minimal sufficiency of the set {headache, c-fibre firing, background conditions}.

Assume that the headache and the c-fibre firing are indeed both sufficient causes of the hand's movement. Would this yield overdetermination? If so, would the overdetermination be problematic? On the face of it, it seems that one could still deny that overdetermination follows. One could deny, that is, that (EXCLUSION*-S) is true.[59] For nothing that has been argued in the meantime diminishes the plausibility of the argument against claims (O_1*) and (O_1**) from Section 4.2. These claims, recall, stated a necessary condition for the hand's movement's being overdetermined by the headache and the c-fibre firing, namely the condition that the hand would still have moved if the headache had occurred without the c-fibre firing.

Perhaps the situation is different and a denial of (O_1*) and (O_1**) is more problematic if we are assuming that the hand's movement has two distinct sufficient causes and not merely two distinct causes, however. Karen Bennett holds that denying (O_1*) or (O_1**) threatens the causal sufficiency of the mental cause for the physical effect (see Bennett 2003: 481, 2008: 289). Her argument, adapted to our example, is as follows: if (O_1*) or (O_1**) is false, it is not the case that my hand would still have moved if the headache had

[58] Likewise for a number of other events that are constituted by supervenient properties. But this yields no problematic overgeneration of sufficient causes, because we are merely dealing with a necessary condition for sufficient causation.

[59] Lowe (1999, 2000, 2008) discusses a case of the following kind: a physical event p_1 is a sufficient cause of a simultaneous mental event m, which in turn is a sufficient cause of a later physical event p_2, such that, by the transitivity of sufficient causation, p_1 is also a sufficient cause of p_2. This case would constitute a counterexample to (EXCLUSION*-S), because p_1 and m are two simultaneous sufficient causes of p_2, but do not seem to overdetermine p_2. I am sceptical about this way of arguing against (EXCLUSION*-S), however, because of worries about the transitivity of causation and about simultaneous causation that were discussed in earlier chapters.

occurred without the c-fibre-firing. Thus, the headache needs the help of the c-fibre firing to bring about my hand's movement. But if the headache needs the help of the c-fibre firing in order to bring about the hand's movement, it is not itself sufficient for it. So a denial of (O_1*) or (O_1**) is incompatible with the headache's being a sufficient cause of my hand's movement.

Bennett's argument succeeds only if sufficient causes have to bring about their effects all by themselves. As we saw earlier in this section, however, this notion of sufficient causes is unworkable, because only very big events bring about other events all by themselves. If, as we are currently assuming, a sufficient cause has to be a member of a set of events that is minimally sufficient for the effect, then Bennett's argument no longer applies, for nothing prevents the c-fibre firing from also being a member of the relevant set of events that is minimally sufficient for the hand's movement, like the set {headache, c-fibre firing, background conditions}.[60] In fact, if the c-fibre firing is a member of a minimally sufficient set that also includes the headache, it is the truth of (O_1*) and (O_1**) that threatens the causal sufficiency of the headache for the hand's movement, not their falsity. For the truth of these counterfactuals at least comes close to saying that the set {headache, c-fibre firing, background conditions} is not minimally sufficient because the subset {headache, background conditions} already is sufficient.[61] Perhaps the lesson to be learned is merely that we have started with the wrong set and that we should take {headache, background conditions} as minimally sufficient for the hand's movement. But in any event Bennett's argument does not prevent us from endorsing an argument against (EXCLUSION*-S) that is based on the denial of (O_1*) or (O_1**).[62]

[60] Mills (1996: 106) endorses a different necessary condition for an event's being 'causally sufficient' for another event. Applied to our case, Mills's condition says that if the headache had not occurred, *then if it had*, my hand would have moved. This condition is consistent with the falsity of (O_1*) and (O_1**), however, since the headache-worlds that are closest to the no-headache-worlds that are closest to the actual world need not coincide with the no-c-fibre-firing-but-headache-worlds that are closest to the actual world. The condition invoked by Mills is also discussed in Yablo 1992.

[61] Since the occurrence of the headache merely necessitates the instantiation of some realizer or other of headaches, but does not necessitate the instantiation of a specific realizer, the set {headache, background conditions} is not a subset-in-name-only of {headache, c-fibre firing, background conditions} (unlike the set {c-fibre firing, background conditions}).

[62] Bennett's argument does not threaten the causation *simpliciter* of the hand's movement by virtue of the counterfactual dependence of the hand's movement on both the headache and the c-fibre firing, for this counterfactual dependence is clearly not undermined by the falsity of (O_1*) or (O_1**). Nor does her argument have to rule out the causal sufficiency of the headache if one takes it to be a necessary condition for event *c*'s being a sufficient cause of event *e* that the material conditional 'If *c* occurs, then *e* occurs' is true in a range of worlds that are sufficiently similar to the actual world. For if the c-fibre firing could not easily have failed to occur (perhaps because it is not very fragile), then the closest worlds where the headache occurs in the absence of the c-fibre firing (in some of which my hand does not move if (O_1*)/(O_1**) is false) are rather dissimilar to the actual world. It is at least

Should such an argument against (EXCLUSION*-S) fail, one could follow the fall-back strategy we used in Section 4.3 and deny (NON-OVERDETERMINATION) instead. It is inessential to this strategy how the overdetermination of physical effects by mental and physical causes (which is assumed for the sake of the argument) comes about. Thus, the fall-back strategy can also be followed if the overdetermination is assumed be due to the fact that both the mental cause and the physical cause are sufficient causes of the physical effect in the sense of each being a member of a set of events that is minimally sufficient for the effect. For irrespective of how the overdetermination is assumed to come about, the case of mental causation is very dissimilar to prototypical cases of overdetermination like the firing squad. In particular, irrespective of the kind of causal relation that is assumed, the metaphysical connection between the two causes is much more intimate owing to the metaphysically necessary (or, in the case of super-nomological dualism, at least nomologically necessary) connection between the headache and the c-fibre firing than it is in cases like the firing squad. Moreover, nothing we have learned in this section tells against the counterfactual dependence of the physical effect on both the mental and the physical cause. This counterfactual dependence remains a major aspect in which cases of mental causation are dissimilar to prototypical cases of overdetermination.

Thus, if we formulate a necessary condition for sufficient causation in terms of minimal sufficiency and if both the headache and the c-fibre firing can be sufficient causes of the hand's movement, then the exclusion problem can be solved by denying either (EXCLUSION*-S) or (NON-OVERDETERMINATION). Unfortunately, it turns out that neither the headache nor the c-fibre firing is a sufficient cause of the hand's movement, so this route to solving the exclusion problem is blocked. The reason does not lie in the modal connection between the two causes, which we discussed in the context of the original formulation of minimal sufficiency above, but, as with sufficient causation as transfer, lies with the inability of our account of sufficient causes to deal with cases of double prevention. To see why double prevention makes trouble for sufficient causes as members of minimally sufficient sets of events, consider the example of double prevention involving idealized neurons from Section 1.6 (see Figure 1.1). In the

an open question whether these worlds are within the range of worlds where the occurrence of the headache needs to materially imply my hand's movement in order for the former to be a sufficient cause of the latter.

example, the firing of *c* causes the firing of *e*. But there is no set of events that is minimally sufficient for the firing of *e* and that includes *c*. The sets {the firing of *c*, the firing of *b*, the firing of *a*} and {the firing of *c*, the firing of *a*} are sufficient for the firing of *e*, but they are not minimally sufficient, because the set {the firing of *a*} is sufficient too: if the only event to occur at the relevant time had been the firing of *a*, then *e* would still have fired.[63] Thus, the firing of *c* is not a sufficient cause of the firing of *e*.

As was the case for sufficient causation as transfer, the fact that muscle contractions operate by double prevention threatens the status of mental and physical causes of bodily movements if sufficient causation is spelled out in terms of minimal sufficiency. Since the calcium release at the neuromuscular junction causes the muscle contraction by double prevention, the calcium release is not a member of a set of events that is minimally sufficient for the contraction and hence is not a sufficient cause of the contraction. Consequently, as was the case with sufficient causation as transfer, there can be no chain of events connected by sufficient causation that contains the link between the calcium release and the contraction as a (non-redundant) link. But, again, it seems that both the headache and the c-fibre firing could only be sufficient causes of a muscle contraction by being a link in such chain. Perhaps this last claim is a bit less obvious than it was in the case of sufficient causation as transfer, but it certainly seems plausible enough to shift the burden of proof to advocates of causal sufficiency. Thus, we are again left with a situation that is tantamount to epiphenomenalism and that also leaves fewer physical causes of bodily movements than seem to exist.

Fortunately, the problems of sufficient causation as minimal sufficiency, like the problems of sufficient causation as transfer, do not carry over to the account of mental causation by counterfactual dependence that I have defended. For the sufficiency of counterfactual dependence for causation yields the correct verdict that cases of double prevention, in muscle contraction or elsewhere, are cases of causation.

I have argued that the exclusion problem, when it is formulated in terms of sufficient causes, is no problem for the counterfactual account of causation, because causation by counterfactual dependence entails neither that a transfer takes place nor that the cause is a member of a set of events that is minimally sufficient for the effect. So far, I have focused on brain

[63] See Paul and Hall 2013: 190–194. Strictly speaking, all the sets would have to include background conditions such as the presence of neuron *e*. These background conditions are omitted for simplicity.

events such as my c-fibre firing as potential sources of trouble for mental causes. One might worry that physical events other than brain events exclude mental causes or that certain physical events exclude mental causes without being a sufficient cause in either of the two senses we have discussed. Before concluding this section, I shall briefly address such worries.

Let us first revisit the causation of bodily movements by double prevention. Owing to the mechanism of muscle contraction, events in the brain do not transfer anything on muscles, but merely release energy that has been transferred from a different source.[64] What about the events that do transfer the energy that is released when muscles contract? Might they exclude the existence of further causes? That seems unlikely. Energy is transferred to the myosin filaments by adenosine triphosphate (ATP) molecules that are present in the muscle fibre. The ATP 'cocks' the myosin filaments, providing them with energy that is released when the myosin moves the actin filaments forward during muscle contraction.[65] The earlier presence of the ATP molecules in the muscle fibre certainly does not rule out any further causes of the later muscle contraction that are simultaneous with the earlier presence of ATP.[66] On the contrary, the earlier presence of ATP clearly allows further causes of the muscle contraction as much as the tenseness of the myosin filaments does. More specifically, the earlier presence of ATP allows additional causes of the muscle contraction by counterfactual dependence. According to the counterfactual account I have presented, mental causes can operate by counterfactual dependence, so they are not threatened by the transfer source in muscle contractions.

A second potential threat comes from the large event whose occurrence implies the occurrence of the effect with nomological necessity. We discarded the suggestion that only events that imply the effect with nomological necessity should count as sufficient causes. We did so for good reason, because otherwise events that are smaller than a cross-section of an effect's past light cone could not be sufficient causes of the effect. One might worry, however, that irrespective of whether we characterize sufficient causation in terms of nomological necessitation, the big events that do nomologically necessitate other events do not leave room for any other

[64] At least the brain events do not transfer anything relevant, although in fact the calcium might transfer, say, a little bit of momentum to the tropomyosin. As in Section 1.6, I am idealizing slightly.

[65] See Guyton and Hall 2006: 76–79. ATP is in turn supplied by several processes, among them glycolysis (Guyton and Hall 2006: 76–79).

[66] By 'the ATP molecules' I mean those ATP molecules that are actually involved in the later 'cocking' of the myosin filaments.

causes of the latter events. If a big event and the laws of nature necessitate another event, can the necessitated event still have other causes? It can. We have implicitly assumed so all along. For we have assumed that determinism is true. If determinism is true, every time-slice of the universe is necessitated by the laws of nature and any other time-slice of the universe. *A fortiori*, every event is necessitated by the laws of nature and any (non-contemporaneous) time-slice of the universe. A time-slice of the universe is basically a very big event. But that all events are necessitated by the laws of nature and these very big events does not prevent the necessitated events from having ordinary causes, by counterfactual dependence or otherwise. In particular, an event may be necessitated by the laws of nature together with a time-slice of the universe while also having an ordinary cause that is simultaneous with this time-slice.[67] If there can be causation of effects by ordinary events under determinism, then there can be causation of effects by ordinary events even if these effects are necessitated by the laws of nature and events that are bigger than the ordinary causes.

It might be objected that, in ordinary cases, the ordinary cause is a part of the bigger event that nomologically necessitates the effect. For instance, my throwing the dart is a part of a cross-section of the past light cone of the balloon's bursting that nomologically necessitates that the balloon bursts. Putative mental causes, the objection continues, are not parts of those bigger nomologically necessitating events, however, and are therefore threatened in their efficacy by the bigger events.

As it stands, the objection misses the point, because nothing I have said rules out that the bigger events have mental events as parts.[68] But presumably what is meant is this: if a certain event is necessitated by another *physical* event (perhaps as big as a physical time-slice of the universe) and the laws of *physics*, then it cannot have a distinct mental cause. This new formulation of the objection brings us back to the first problem from the beginning of this chapter: if something is determined by the physical, how can it have additional mental causes? The answer should not come as a surprise: physical events can have mental causes by counterfactual dependence. Nothing prevents mental events from making a difference to physical events, even if these physical events are necessitated by physical laws and events. Indeed, for the reasons given in Chapter 2, this difference-making relation is almost unavoidable for non-reductive physicalists, and dualists can easily have it too by adopting

[67] For simplicity I'm talking here as though all events are instantaneous, which of course they are not. But the points can easily be generalized by taking into account all the time-slices that a given event overlaps.

[68] Although if the psychophysical laws are synchronic, they are idle in this necessitation; see note 1.

super-nomological dualism. What we cannot have, of course, is that mental events make a difference to a given physical event (except perhaps vacuously) if we hold fixed the physical past and the physical laws. But, as we saw in Chapter 3, this is not the right way to think about the situation. For we can also prevent *physical* events from making a difference to other physical events by holding further physical events fixed (suitable intermediate events, for instance). If we assess the difference-making claims correctly, mental events can be *bona fide* causes of physical events.

The upshot of this section is that the exclusion problem is generally more severe if it is formulated in terms of sufficient causation rather than in terms of causation *simpliciter*. We can spell out sufficient causation in terms of transfer of a physical quantity or in terms of minimal sufficiency. Either way, mental causation is in jeopardy; somewhat surprisingly, the causation of bodily movements by events in the brain is also in jeopardy. But the problems do not carry over to the account of mental causation by counterfactual dependence. Thus, the exclusion problem for other notions of causation yields indirect support to the account of mental causation that I have defended.[69]

4.6 Conclusion

At the heart of the exclusion problem are two claims: the claim that physical effects of physical causes can have distinct mental causes only if the mental and physical causes overdetermine the physical effects; and the claim that there is no widespread overdetermination of physical effects. We have seen that non-reductive physicalists and super-nomological dualists can respond in two ways. They can deny that mental causation entails overdetermination, or they can deny that the kind of overdetermination that would ensue would be problematic. We have also seen that non-reductive physicalists can deny that the nature of physical realizers has any detrimental consequences for the efficacy of the mental properties they realize. Similarly, super-nomological dualists can deny that the nature of physical bases has any such detrimental consequences for the efficacy of mental properties. When formulated in terms of sufficient causes rather than in terms of causation *simpliciter*, the exclusion problem is generally more severe, but none of this severity spills over to the account of mental causation by counterfactual dependence.

[69] For further discussion of the relation between different theories of causation and accounts of mental causation, see Bennett 2008: 293; Lycan 2009: 557–558; Hitchcock 2012b.