Exclusion Again[1]

Karen Bennett
Princeton University
Final version, March 2007
Forthcoming in Jakob Hohwy and Jesper Kallestrup, eds., *Being Reduced* (Oxford)

1. Introduction: Two Uses of The Exclusion Problem

I am not moved by the exclusion problem. I am not alone in this; many others agree that the argument is not convincing. However, I think that there is more to be learned about what exactly is wrong with it. I therefore propose to say a few more words about why the exclusion problem is not really a problem after all—not, at least, for the nonreductive physicalist. The genuine *dualist* is still in trouble. Indeed, one of my main points will be that the nonreductive physicalist is in a rather different position vis à vis the exclusion problem than the dualist is. Properly understanding nonreductive physicalism—and clearly recognizing that it is, after all, a form of *physicalism*—goes a long way towards solving the exclusion problem. My goal in this paper is to further defend my preferred solution (2003), and to argue that it is not available to the dualist.

Let us begin by reminding ourselves how the exclusion problem is supposed to go. The basic idea is that if everything that happens can be accounted for in purely physical terms, then the mental seems to be left with nothing to do. More precisely, the problem is that the following five claims are incompatible with each other:

Distinctness: Mental properties (and perhaps events) are distinct from physical properties (or events).[2]
Completeness: Every physical occurrence has a sufficient physical cause.[3]
Efficacy: Mental events sometimes cause physical ones, and sometimes do so in virtue of their mental properties.

[1] I intitally drafted a version of this paper for a conference on mental causation put on by the NAMICONA group at Århus University. Many thanks to all the conference participants for very helpful discussion, particularly Helen Beebee, Frank Jackson, and Barry Loewer. Thanks also to the Corridor reading group, constituted on that occasion by Cheryl Chen, Adam Elga, Liz Harman, Sean Kelly, Jim Pryor, and Ted Sider. I presented a more recent version at the AAP meetings and at a workshop on mental causation at Macquarie University. Thanks especially to David Chalmers, Kit Fine, Peter Menzies, Adam Pautz, Denis Robinson, and Daniel Stoljar. Finally, thanks to Jesper Kallestrup and Jakob Hohwy for comments on the final version.
[2] Different versions of the exclusion problem arise depending upon whether it is type identity, token identity, or both that is denied. I shall be as neutral as possible on this question.
[3] Three quick points. First, not much is affected by weakening Completeness to the claim that every physical occurrence that *has a cause* has a sufficient physical cause. Second, not much is affected by weakening Completeness in a different direction, to the claim that every physical occurrence merely has its probability fixed by entirely physical antecedents. Third, notice that none of these versions says that everything that happens has *only* physical causes. That claim is stronger, and is not a good way to start out the exclusion argument (Kim flirts with using it in 2003, 162-164, but rightly decides not to).

Nonoverdetermination: The effects of mental causes are not systematically overdetermined; they are not on a par with the deaths of firing squad victims.[4]

Exclusion: no effect has more than one sufficient cause unless it is overdetermined.

This way of presenting the problem is neutral about which claim should be rejected. It is not always presented that way—simply as a package of inconsistent claims, some one of which has to give, and any one of which will, in principle, do.

One virtue of presenting it this way is that doing so lays bare what solving the exclusion problem does and does not require. The question is simply how to avoid commitment to an apparently inconsistent set of propositions. That means that solving the exclusion problem requires, *and only requires*, doing one of two things—either arguing that one of the five incompatible claims is false, or else somehow arguing that they are not incompatible after all. Solving the exclusion problem decidedly does not require *defending* any of those premises. In particular, solving the exclusion problem does not require defending the causal efficacy of mental events and properties.

This makes it different from the various other problems about mental causation—such as concerns about whether Cartesian souls or second-order functional properties are the right sorts of things to cause anything—with which it can easily be entangled. The point of the exclusion problem is not that there is a special problem establishing the causal efficacy of the mental, but instead that the *assumption* that it is efficacious leads to trouble (see my 2003, 471-472).[5] My point here is that responding to the exclusion problem requires less than is sometimes supposed. It does not require providing a positive story about how the mental manages to be causally efficacious. Telling such a story is of course required by a full defense of mental causation from all challengers, but not by a defense from the exclusion problem in particular.

As I have said, the exclusion problem is not always presented as a package of inconsistent claims. It is sometimes instead presented as an argument against Distinctness in particular. When it is framed like that, it is supposed to show that the mental is not distinct from

[4] It is purely a terminological matter whether this is formulated as stating that the effects of mental causes are not overdetermined at all, or as stating that they are not overdetermined in some particularly *bad* way. The important point is the insistence on the disanalogy.

[5] The point here can also be made by virtue of the standard metaphor about how the physical 'does all the causal work'. The problem is not that the mental lacks the requisite skills; the problem is rather that there are no job openings. It is one thing to be fit for work, and quite another to actually find a job. The exclusion problem comes in at the *second* stage; the other arguments enter at the first.

the physical after all. But that conclusion is somewhat ambiguous, and so is the role of the exclusion argument in the philosophy of mind literature. The argument gets used for two rather different purposes. Sometimes it is used as an argument for physicalism, as against property or substance dualism (see, e.g., Papineau 1995, 2001).[6] Sometimes it is used as an argument for reductive physicalism, as against nonreductive physicalism (see, e.g., Kim 1989, 1993, 1998, 2005). That is, sometimes it is used to defend physicalism, and sometimes to defend a *version* of physicalism. These two uses can blur into each other, and it is not always obvious which a given author is doing. However, they need to be carefully distinguished. Doing so opens up the possibility of trying to preserve one while rejecting the other. As a proud card-carrying physicalist, that is what I would like to do.

The thought is that we physicalists should set our sights higher than we have in the past. We should not merely argue that we are not in trouble over the exclusion problem; we should argue that we are not in trouble *while the dualist still is*. That is, we should do our best to deny that we are in the same boat as emergentists vis à vis the exclusion problem and the commitment to so-called 'downward causation' (*pace* Crane 2001, Kim e.g. 1989, 1993a). We should do our best to deny that the exclusion argument is a good argument for reduction, while nonetheless insisting that it *is* a good argument for the claim that the mental is nothing over and above the physical. At least, that would be the ideal conclusion. It would be better than the claim that the exclusion problem is so deeply flawed that it is not a genuine problem for anyone. After all, actual arguments for physicalism are rather hard to come by, and we should not throw the baby out with the bathwater.

I am going to argue for something slightly short of this ideal conclusion. The trouble is that it is not clear that dualists need to accept Completeness. Physicalists presumably do; physicalism itself arguably entails it. But it is not clear that dualists do. Many, notably Descartes, reject it, and those that do reject it do not contradict themselves in doing so. So perhaps dualists can escape the exclusion problem by claiming that some physical effects have purely mental causes. Perhaps. But perhaps not. The question is not just whether dualists can *consistently* reject Completeness, but whether they can *plausibly* reject it. It is not clear that they need to endorse Completeness, but it is also not clear that they can happily deny it and walk away whistling. Many contemporary property dualists, including David Chalmers (1996, 150),

[6] That is, it is used to get from the completeness of physics to physicalism proper.

do endorse Completeness. Indeed, the current Chalmers-inspired trend towards 'naturalistic,' scientifically responsible forms of dualism would seem to be a trend towards forms of dualism that are much friendlier to Completeness. It is an interesting and important project, I think, to see whether even dualists have compelling reason to accept that physics is causally complete. Perhaps the best reasons to accept that claim do not presuppose physicalism. David Papineau comes close to showing this, with his illuminating discussion of the scientific history of Completeness (1995, 2000, 2002). However, it is not a project I am going to undertake here.

Since I am not going to argue that even dualists must accept Completeness, I cannot quite argue that the exclusion problem constitutes a successful argument against dualism. And since I cannot quite argue that the exclusion problem works against the dualist, I obviously cannot quite argue that it works against the dualist but does not work against the nonreductive physicalist. However, I am still going to do better than simply argue that it does not work against the nonreductive physicalist. I have already done that (2003). Here, I am going to argue for something in between that conclusion and what I now take to be the ideal conclusion. What I am going to argue is that the nonreductive physicalist's best strategy for avoiding the exclusion problem is not available to dualists. Defending nonreductive physicalism does not require defending full-blown dualism, too.

All physicalists have a well-motivated solution to the exclusion problem that no dualist has. Physicalists' best option is to deny Exclusion, and thereby endorse a strategy that I call 'compatibilism', and have defended in more detail elsewhere (2003). Dualists, I shall argue, must accept Exclusion. They therefore really do have to choose between denying Efficacy, denying Nonoverdetermination, and denying Completeness. That is, they must either endorse epiphenomenalism, claim that the effects of mental causes are systematically overdetermined in the standard firing squad way, or else claim, with Descartes, that the mental injects itself into the physical causal order. None of these options are particularly appealing. Indeed, I claim that if a case can be made that dualists should accept Completeness after all, the fact that they must also accept Exclusion would amount to an argument for the ideal conclusion. It would mean that the exclusion problem *is* a good argument against dualism, though it does not succeed against nonreductive physicalism.

My primary goal in this paper is to argue that only physicalists can be compatibilists. Here is the rough outline of the argument. There are two *prima facie* ways to motivate the

central compatibilist claim that Exclusion is false. However, only one of them is successful, and it is not open to dualists. In arguing that one of the strategies does work, I will argue that, despite possible appearances, the force of the exclusion problem does not rest on any particular account of causation. In arguing that the successful version of compatibilism is only open to physicalists, I will rely upon the claim that the physicalist and the dualist mean rather different things when they endorse Distinctness. When nonreductive physicalists deny that mental properties are physical, they are saying something much weaker than dualists are. Physicalists have a clear argument for the falsity of Exclusion. But because dualists mean something rather different by Distinctness, they wind up with no argument against Exclusion at all.

As those last remarks make clear, however, pursuing this line of argument requires sorting out what the various positions are, and what the labels mean. Here, then, is the plan for the rest of the paper. In the next section, I will briefly clarify the relations between dualism and physicalism, both reductive and nonreductive. In sections 3 through 5, I will argue in detail that compatibilism requires physicalism. In section 6, I will turn to some objections and replies.

2. Taxonomy: Reductive Physicalism, Nonreductive Physicalism, Property Dualism

So what is physicalism, anyway? It is notoriously hard to define the view adequately, but I can at least offer up the same slogans as everyone else. Physicalists not only endorse the completeness of physics, but also think that all the facts are physical facts—that there is nothing 'over and above' the physical. Physicalists believe that everything globally supervenes[7] on the physical as a matter of metaphysical necessity. More precisely, physicalists typically endorse a thesis like the following:

> Any world which is a minimal physical duplicate of our world is a duplicate *simpliciter* (Jackson, 1998, 12),

where a minimal physical duplicate is what results from duplicating all the physical facts and "stopping right there". This allows the possibility of worlds physically like ours, but with ghostly 'extras', and thus does not require that physicalism be *necessarily* true. It is a contingent truth about the actual world. (See also Lewis 1983 and Chalmers 1996; see Hawthorne 2002 for interesting challenges to all three definitions). Given the actual physical facts and physical laws—and no extras—everything else follows necessarily.

[7] I invoke global supervenience because it is both standard and convenient. As I have argued elsewhere (2004), however, any claim made with global supervenience can also be formulated in terms of strong supervenience.

Crucially, note that physicalists deny that there are special psychophysical laws in addition to the physical ones—breakable laws that merely link or *tether* the mental to the physical. That is a dualist claim. Physicalists instead think that mental events and properties are not truly distinct existences that can be snipped away from their physical bases. There is no room for any wedge. That is why the metaphysical necessity of the supervenience claim—rather than the mere nomological necessity endorsed by some dualists (e.g. Chalmers 1996)—is of crucial importance to their view.

Now, many physicalists do endorse a claim that can sound vaguely dualist—namely, "mental properties are not identical to physical ones". These are nonreductive physicalists, and it is their endorsement of this claim that makes them appear vulnerable to the exclusion problem. It is, after all, the Distinctness premise. Both nonreductive physicalists and property dualists endorse Distinctness, although they have different motivations for doing so.

Property dualists typically endorse Distinctness for the same reason that they reject physicalism—namely, they do not think that consciousness can be explained in physical terms. Nonreductive physicalists, in contrast, typically endorse Distinctness for a combination of reasons having to do with the purported multiple realizability of mental state-types, and with the semantics of mental terms. They typically think that words like 'pain' rigidly designate a second-order functional property—the property of having some physical property or other that plays a particular causal role. Reductive physicalists, in contrast, do identify mental properties with first-order physical properties. The most plausible version, best articulated by David Lewis (especially 1978, 2000b), accommodates multiple realizability intuitions by taking terms for mental state-types to be nonrigid designators that can refer to different first-order physical properties in different contexts.[8]

Now, I myself think that there are many interesting complexities here, and suspect—somewhat heretically—that the distinction between reductive and nonreductive physicalism is probably not metaphysically very deep. However, I am not going to argue that here. What I *do*

[8] This position is sometimes called 'realizer functionalism.' It is more plausible than the position standardly attributed to early type-identity theorists like U. T. Place (1956) and J. J. C. Smart (1959), which simply takes a term like 'pain' to rigidly designate a first order physical property like C-fiber stimulation. This view identifies *pain itself*, rather than a 'local' property like pain-in-humans, with C-fiber stimulation, and is consequently subject to the multiple realization objection (Putnam 1973).

However, it is at least debatable that these so-called 'identity theorists' actually had something closer to Lewis-style realizer functionalism in mind. Consider, for example, Smart's insistence that the identity between pain and C-fiber firings was merely contingent (1959, 147.) We could take this as an unfortunate pre-Kripkean failure to recognize the necessity of identity, but we could also take it as an indication that he was not using 'pain' as a rigid designator.

want to argue is that even if the line between reductive and nonreductive physicalism is indeed important, it is *less* important than the line between physicalism and non-physicalism. My claim is that the commonalities between reductive and nonreductive physicalists swamp their differences, at least as far as the exclusion problem is concerned.

One can picture nonreductive physicalism as occupying middle ground between reductive physicalism and property dualism. After all, there are two choice points here. One is the question of the Distinctness premise—are mental properties identical to physical ones? The other is the question of physicalism—does the mental, along with everything else, supervene with metaphysical necessity upon the physical? The reductive physicalist says 'yes' to both. The nonreductive physicalist says 'no' to the former and 'yes' to the latter. The property dualist says 'no' to both. (Terminological caveat: one can occasionally find the label 'nonreductive physicalism', or at least 'nonreductivism', used in the way that I use 'property dualism'. See Chalmers 1995, and, fascinatingly, Kim 2005, 49.[9]) However, this "middle ground" occupied by nonreductive physicalism does not lie halfway between the other two views. It lies closer to reductive physicalism, because the decision that matters is the one the reductive and nonreductive physicalist share—namely, that physicalism is true.

One way to put the basic upshot of this paper is that whether mental and physical properties are numerically distinct matters less than whether physicalism is true. Another way to put the same point is that the property dualist and the nonreductive physicalist have rather different claims in mind when they endorse Distinctness. The nonreductive physicalist is simply asserting that mental and physical properties are numerically distinct. The property dualist, in contrast, is asserting both that the properties are numerically distinct *and* denying that any dependence relation with modal strength holds between them. (See Stoljar, this volume, for more detail on the importance of the difference between various distinctness claims.)

Whichever way we put the point, what matters is that nonreductive physicalists endorse, and dualists reject, the claim that everything supervenes with metaphysical necessity on the physical facts and laws. In the next three sections, I will argue that this makes all the difference.

---

[9] This passage appears to suggest that Kim always had property dualism in mind when he used the label 'nonreductive physicalism'—which, if true, changes the impact of many of his claims. Here is the passage:

> I think we can set aside the possibility that mind-body supervenience is logically or metaphysically necessary, since such a view is essentially a reductionist view, and we are here considering [mind-body supervenience] as a part of nonreductive physicalism (2005, 49).

He also says in a footnote that "there are independent reasons for thinking that mind-brain supervenience, if it holds, must be construed as nomological, not logical or metaphysical, supervenience" (49n15).

3. Denying the Exclusion Principle: Compatibilism

Exclusion, recall, says that no effect has more than one sufficient cause unless it is overdetermined. The idea behind the name is just that the existence of one sufficient cause somehow blocks or excludes all others, except in the presumably rare case of overdetermination. To deny this claim is to insist that there is more to overdetermination than that—to claim that in certain circumstances effects can have several sufficient causes and nonetheless not count as overdetermined in the paradigmatic firing squad sort of way.[10] The resulting view is both popular and plausible. It is the view that I have called Compatibilism, and defended in more detail elsewhere (2003).

As I argued there, however, we cannot simply *proclaim* Exclusion false; we have to at least make a modicum of an effort to say *why* it is false. There are two strategies for doing so, based on two diagnoses of what overdetermination requires. To see this, think about some paradigm cases of overdetermination: firing squads, two children (presumably named Billy and Suzy) simultaneously throwing rocks at a window, your alarm clock sounding at just the same moment as a jackhammer starts up outside your window. Paradigm cases like these seem to involve a) two completely distinct b) transfers of energy. That is, there are two separable components in there—first, that there are *two* causes, and, second, that the causes are *ooomphy* in some sense that I will try to say a bit more about later. A compatibilist can focus on the latter, and reject the ooomphy notion of causation in favor of something along the lines of a pure counterfactual dependence notion. Or he can focus on the former, and claim that in the relevant cases the requirement of twoness is not quite met—i.e., he can claim that in the relevant cases the causes are tightly related in some way that defuses the threat of overdetermination.

So one strategy for denying the exclusion principle focuses on the notion of causation in play, and the other focuses on the relation between the causes. Only one of them works, and only one of them is available to a dualist. Unfortunately, it's not the same one. The one that works is the strategy that claims that certain kinds of tightly related causes can both be causally sufficient for the same effect without overdetermining it. In section 4, I will briefly sketch my

[10] Again, it does not matter whether I say that sometimes effects can have more than one sufficient cause without being overdetermined, or whether I say that sometimes effects can have more than one sufficient cause without being overdetermined in some particularly bad way (see note 4). It is a purely terminological matter whether I say that not all double causing is overdetermination, or whether I say that all double causing *is* overdetermination, but not all overdetermination is on a par.

own preferred version of this strategy, and argue that no dualist can avail herself of it. In section 5, I will argue that although the dualist certainly *can* move to a pure dependence notion of causation, doing so does not defeat the Exclusion principle. Doing so consequently cannot itself ground compatibilism.

4. Motivating Compatibilism I: Tight Relations Between the Causes

The thought that the mental and physical are so tightly related that they do not overdetermine their effects is a natural one, and I am not the only person to have it (see in particular Yablo 1992, Shoemaker 2001, Pereboom 2002).[11] But unlike some people who push this line, I think that we both must and can say more about just *why* certain kinds of tight relation moot the threat of overdetermination. I think it is unsatisfactory to say, while emphasizing one's nonreductivism, that mental event or property *m* is not identical to narrowly physical event or property p, and then to say in practically the same breath that of course *m* and *p* do not causally compete in any way. If they are distinct, the threat of competition must be *argued against*. We cannot just assert that we can have it both ways. We nonreductive physicalists must properly shoulder the burden of proof and say *why* these intimately-related-but-*distinct* causes do not overdetermine their effects. This is what I tried to do in my earlier paper (2003).

What I claimed is that overdetermination requires the nonvacuous truth of certain counterfactuals. In order for two causes, *m* and *p*, to overdetermine some effect *e*, it must be nonvacuously true that

(O1) if *m* had happened without *p*, *e* would still have happened: $(m \,\&\, \sim p) \,\Box\!\!\rightarrow e$, and
(O2) if *p* had happened without *m*, *e* would still have happened: $(p \,\&\, \sim m) \,\Box\!\!\rightarrow e$.[12]

A couple of caveats here. First, this is only meant to be a necessary condition, not a sufficient one. In particular, overdetermination also requires that *m* and *p* both be causally sufficient for

[11] I take it that in trying to say more about *why* certain pairs of causes do not overdetermine their effects, I am addressing a question that Stephen Yablo (1992) does not address. I think he can more or less take my view on board if he likes. In contrast, both Derk Pereboom (2002) and Sydney Shoemaker *are* addressing the same question as me (2001). However, both of their views are more metaphysically committal than mine. My approach is much more neutral on questions about the nature of properties, causal powers, the constitution relation and the like. Perhaps this is a weakness; perhaps it is a strength. Regardless, all four of us are compatibilists, with views in the same rough vicinity.

[12] This test is supposed to be fully general; I only name the causes 'm' and 'p' in order to streamline the ensuing discussion. Also, the counterfactuals can be tweaked in various ways to account for the fact that a version of the exclusion problem can be run on mental *properties*.

e.[13] Second, this is not meant to require a counterfactual analysis of causation; it is simply a test for overdetermination that reflects our everyday reasoning about causation and overdetermination. Take any case you like. If only one of the putative causes really was a cause, only one of the counterfactuals will be true. If they were joint causes, both of the counterfactuals will be false. If *m* and *p* are in fact the very same event, both counterfactuals will be vacuous. And if *e* really is overdetermined by *m* and *p*—think firing squads, Billy and Suzy throwing rocks at the window, etc.—both counterfactuals will be nonvacuously true.

If I am right that the nonvacuous truth of these counterfactuals is necessary for overdetermination, the next step is clear.[14] The question is whether the dualist, the physicalist, or both, can deny the nonvacuous truth of at least one of the counterfactuals. I shall argue that although the physicalist can deny the nonvacuous truth of (O2), the dualist cannot deny the nonvacuous truth of either (O1) or (O2). I shall only take a quick look at the status of (O1) before turning to a more detailed discussion of the status of (O2).

First, (O1). The status of (O1) is complicated for the physicalist (see my 2003, 481-484). Luckily, however, I can dodge those complications here, because the status of (O1) is not particularly complicated for the dualist. She will not claim that it is either vacuous or false. She will not claim that it is vacuous, because she thinks that *m* can indeed happen without *p*. It can certainly happen without *p* in particular, and she will also think that it can happen without any physical realizer at all. Even physicalists, recall, usually think that physicalism is contingent. Cartesian souls are possible, just not actual; worlds where they exist are worlds in which

---

[13] It is perhaps worth emphasizing again that I am not arguing that mental events or properties *can* be causally sufficient for anything; I am arguing that the assumption that they can be does not lead to widespread overdetermination.

[14] Martin Jones has raised the following counterexample to my claim that the nonvacuous truth of the counterfactuals is necessary for overdetermination. Suppose that there is a small firing squad of just two shooters, Billy and Suzy, with their weapons trained upon the victim. Suppose further that Billy is standing closer to the victim than Suzy is. Billy is a sensitive chap, however, and wants to avoid being the only person to shoot the victim. So he waits a split second to make sure that Suzy has fired her gun, and only then fires his. If Suzy does not fire when the command is given, Billy fires into the air. But if Suzy does fire, he aims properly and fires at the victim too. Although he fires later than Suzy, his bullet nonetheless strikes the victim at the same moment that Suzy's does. This certainly looks like a case of overdetermination—after all, the victim gets shot with two bullets! Yet one of the two counterfactuals is (non-backtrackingly) false. Had Billy fired his gun and Suzy not fired hers, the victim would not have died. So it looks like this is a counterexample to the claim that overdetermination requires that both counterfactuals be nonvacuously true.

There are several possible responses to this kind of 'staggered' overdetermination case. One is to insist that the relevant event here is Billy's firing *at the victim*, not his firing full stop. The overdetermination counterfactuals are nonvacuously true for that choice of cause. Another, which is more neutral about the individuation of events, is to grant that the death is not overdetermined by Billy and Suzy's firings, and to claim that it is instead overdetermined by some *intermediate* pair of events for which the counterfactuals are nonvacuously true. Billy and Suzy's firings count as overdetermining the death in a slightly derivative sense, because they cause events that nonderivatively overdetermine it.

physicalism is false, and mental properties can be instantiated without being physically realized. And the dualist will not want to say that (O1) is false, either. Doing so certainly appears to undermine the claim that *m* is causally efficacious with respect to *e*. To say that (O1) is false is to say that if *m* were to happen without *p*, *e* might not occur. But that suggests that *p* is required, that *m* is not in fact good enough to do the work. Thus there is a real tension here between saying that (O1) is false, and that *m* is causally sufficient for e.

Let us move on to (O2). I have argued elsewhere (2003) that the *physicalist* gets to say that (O2) will come out either false or vacuous in all cases of mental causation, depending on what sort of physical events or properties he takes to be causally sufficient for the effect in question. In the remainder of this section, I would like to argue that the same is not true of the dualist. The dualist cannot claim that (O2) is either vacuous or false.

Let's start with the easy bit. It is clear that only the physicalist can say that (O2) ever comes out vacuous. The dualist cannot, because she does not think that there are any physical events or properties that metaphysically necessitate mental ones. She precisely thinks that there are—at best!—contingent psychophysical laws that link the two. So the dualist denies that there is any legitimate substitute for *p* that would make the antecedent metaphysically impossible. She at most thinks that there are choices of *p* that would make the antecedent *nomologically* impossible. So the dualist cannot claim that any instance of (O2) is vacuous.

The interesting and complicated question is whether the dualist can claim that (O2) is false, despite thinking that *p* is causally sufficient for the effect. I do not see how. Here, just as in the case of (O1), there is a real tension between the falsity of the counterfactual and the efficacy of the putative cause held constant. However, the *physicalist* can escape this tension, and say that (O2)'s falsity is consistent with p's causal sufficiency for e. Let me sketch my story about how the physicalist can do that, and then explain why the dualist cannot say the same thing.

The first move the physicalist has to make towards establishing the falsity of (O2) is to convince us that he is not committed to thinking that all instances of (O2) are vacuous. He is not. If he holds a particular view about the nature of causal sufficiency, he can think that some physical event *p* is causally sufficient for effect *e*, that some mental event *m* is as well, and that *p* fails to necessitate *m*. After all, most of the events and properties that we talk about when we talk about the exclusion problem–things like patterns of neural activity, or properties like *being a*

C-fiber firing—do not necessitate anything mental. These ordinary, everyday events and properties tend to be spatio-temporally localized, and they only guarantee the existence of the mental events and properties that they 'realize' *given certain background conditions*. For example, it is perfectly possible for C-fiber firings to occur without pain. They could be hooked up rather differently, or not hooked up to anything at all. Context matters. It is not an accident that physicalism is usually characterized by means of a *global* supervenience thesis rather than a local one.

There is, in short, an important mismatch between the sorts of physical properties and events that are typically invoked in instances of the exclusion problem, and those that constitute the supervenience base for the mental. It is only complicated extrinsic physical properties, and physical events with complicated extrinsic essences, that will metaphysically necessitate mental ones. Thus as long as it is legitimate to plug the more intrinsic, everyday physical events and properties into the counterfactual (O2), it will not come out vacuous. Whether it is legitimate to do so depends on whether such things ever count as causally sufficient for anything, which in turn depends upon one's views about the nature of causal sufficiency. If you think that causal sufficiency is a kind of strict sufficiency, according to which only big complicated sums of everyday events, background conditions, causal intermediaries and the like count as causally sufficient for anything, then the only physical occurrences that will ever be causally sufficient for action will be the complicated nonlocal occurrences. These do guarantee the mental ones, and thus all instances of (O2) come out vacuous. If, on the other hand, you think that causal sufficiency is mere sufficiency in the circumstances, then the more ordinary, localized physical events and properties will count as causally sufficient for action, and not all instances of (O2) will be vacuous.

So even the physicalist can indeed say that there are nonvacuous instances of (O2). These are claims like the following:

> had these C-fibers fired occurred without the pain, my hand would still have jerked back from the stove.

However, these claims are typically *false*—by physicalist lights, anyway. The idea here is simple. The context within which the physical event or property guarantees the mental one *is the same as the background conditions within which it brings about its effects*. So the C-fibers can perfectly well fire without the pain. They could be wired up differently, or perhaps twitching

away in a petrie dish. But in such a situation, they will not at all cause the sorts of things they actually cause—they will not cause me to pull my hand away from the stove, and they will not cause me to jump around swearing like a sailor. For localized choices of *p*, then, *p* can indeed happen without *m*, but if it did, there is no reason to expect the occurrence of *e*. Those instances of (O2) are false. So says the physicalist, anyway.

Let's pause for a quick rundown: the physicalist says that if your notion of causal sufficiency requires you to plug in complicated extrinsic properties and events for *p*, then (O2) is vacuous. If, on the other hand, your notion of causal sufficiency allows you to plug in more ordinary sorts of events and properties, instances of (O2) will not be vacuous, but will typically be *false*. Now, we have already seen that the dualist cannot claim that any instance of (O2) is vacuous. So can she claim that all instances are false? Can she adopt this strategy, and say with me that if the physical cause had occurred without the mental one, it would not have caused the same effects?

Not without abandoning standard ways of evaluating counterfactuals. For the dualist, the closest world in which the C-fibers fire without pain is not a world in which various surrounding physical facts go differently. It is not a world in which the C-fiber stimulation takes place in a petri dish, or otherwise without crucial background conditions that actually obtain. It is instead a world in which the psychophysical law that links appropriately situated patterns of C-fiber stimulation to pains is violated. It is not a full-blown *zombie* world, mind you—that would clearly involve the kinds of "big, widespread, diverse violations of law" that Lewis says it is of the first importance to avoid (1979, 47). It is instead simply a world in which just that particular physical occurrence fails to give rise to the sort of mental one that usually accompanies it. That is merely a "small, localized, simple violation of law," that allows us to "maximize the spatio-temporal region throughout which perfect match of particular fact prevails" (47-48). This one tiny little violation of psychophysical law is a lot easier to accomplish—if it can be accomplished at all—than a big sweeping change in circumstances.

Crucially, of course, the nonreductive physicalist does not think it can be accomplished at all. As I have already emphasized, he thinks it is a mistake to think of psychophysical laws as contingent nomological connections between distinct things. The dualist and the nonreductive physicalist disagree about what is possible, about what worlds there are—and this forces them to disagree about which is the closest world in which the antecedent of (O2) is true. For the

relevant choices of *p*, the closest *p* & ~*m* world that the nonreductive physicalist recognizes is *not* an *e* world. But for those same choices of *p*, the closest world that the *dualist* recognizes *is* still an *e* world. Nothing physical changes at all; given Completeness, *e* still occurs.[15]

Consequently, the dualist cannot say that (O2) is either false or vacuous, and therefore cannot motivate compatibilism in this way. For the dualist, cases of mental causation *do* meet the necessary condition on overdetermination. She thus has no argument for the claim that mental and physical events and properties are so intimately related that they can both be causally sufficient for the same effect without overdetermining it. She has given us no reason to think Exclusion is false of mental and physical causes. In short, it *matters* that the dualist does not think that the connection between physical facts and mental facts is as tight as the nonreductive physicalist does. A mere nomological connection does not fly.[16]

5. Motivating Compatibilism II: The Notion of Causation

Let us move on, then, to the other strategy for motivating compatibilism. Recall that I said there were two strategies—one that focuses on the intimate relation between the putatively competing causes, and one that focuses on the notion of causation in play. The latter strategy claims that the plausibility of the exclusion principle, and thus the force of the exclusion argument as a whole, turns upon a mistaken view about the nature of causation—namely, that it involves some kind of *ooomph* over and above mere counterfactual dependence.

I shall not say anything about whether this really *is* a mistaken view about the nature of causation, and I also shall not argue that the dualist faces any special difficulty claiming that it is mistaken. Presumably, she can wade into the causation literature and emerge with whatever view she likes. I certainly see nothing stopping her from adopting the sort of pure dependence view that is allegedly friendly to compatibilism. Instead, I will argue that this strategy simply does not work. Rejecting oomphy causation does not in fact provide any reason to think that the exclusion principle is false. The force of the exclusion problem does not turn upon any substantive view about the nature of causation.

---

[15] Of course, the dualist may at the end of the day want to avoid the exclusion problem by denying Completeness. But the question at the moment is whether she can avoid the exclusion problem *without* doing so, by means of compatibilism.

[16] It turns out that Barry Loewer has given a very similar argument for a slightly different conclusion. In his case, it is for the claim that the dualist cannot say that ~*m* $\Box\!\!\rightarrow$ ~*e* is true, rather than (as for me) for the claim that the dualist cannot say that *p* & ~*m* $\Box\!\!\rightarrow$ *e* is false. See 2001, 51-52.

The two views about causation I have in mind are those that Ned Hall has called 'dependence' and 'production' (2004).[17] The rough distinction is this. According to the production view, causation is a matter of the transfer of energy, or—to use the slang of its detractors—the transfer of 'causal juice', 'oomph', or 'biff'. This is the kind of view according to which causes generate their effects by means of a connecting process (Salmon 1984, Dowe 2000). It entails that there is no such thing as causation by omission or double prevention. According to the dependence view, in contrast, causation is *purely* a matter of counterfactual dependence (or probability-raising, or something of the sort). Patterns of counterfactual dependence do not indicate underlying ooomphy causes, but fully constitute causal reality.

People sometimes suggest, both in conversation and (to some extent) in print that the exclusion problem does not get off the ground on the pure dependence notion of causation.[18] But while I certainly agree that the production view is often in the background of discussions of the exclusion problem—Kim admits as much (2002, 675)[19]—I do not agree that the exclusion problem itself actually *requires* it. I do not agree that rejecting it makes the problem go away.

My claim here is ripe for misinterpretation, so let me be clear about what it is that I am disputing. I am not denying that a dependence notion of causation might be handy in establishing the causal efficacy of the mental in the first place. That is, it might well help block the worries about the causal relevance of mental properties that arose around Davidson's anomalous monism (see Lepore and Loewer 1989), and it will quite likely also help block Princess Elizabeth's worries about the causal powers of Cartesian souls.[20] Whether it can or not depends on whether it is good enough as an account of causation full stop. However, those questions are not currently on the table (see section 1). The only question that *is* on the table is whether a pure dependence notion of causation can defuse the threat of overdetermination by falsifying the exclusion principle. The question on the table is whether thinking that the presence of one cause excludes others requires thinking of causation like causal juice of which some

[17] Hall himself thinks that our causal intuitions are not univocal, and that we actually have two concepts of causation.
[18] Loewer 2002 is a possible example, though he does not in the end endorse the strong claim in the main text above (personal communication).
[19] Kim says that "Loewer is right… in saying that my thinking about causation and mental causation involves a conception of causation as 'production' or 'generation'" (2002, 675). He goes on to try to defend the production model against Loewer's claim that contemporary physics has no place for such a notion. I think Kim is right to admit this, but wrong to assume that the pure dependence notion alone would dissolve the problem completely.
[20] While it is very hard to see how nonphysical, nonextended souls could actually *transfer energy* to physical things like neurons, it would not be very hard to argue that there are counterfactual connections between, say, acts of will and the contraction of muscle fibers. See my 2007, section 2.

effects get a double dose. Is it true that if we reject that in favor of dependency, the exclusion principle will fall away, bringing the exclusion problem with it like a house of cards?

No. This second strategy for defending compatibilism does not work. Moving to a pure dependency notion of causation is not sufficient to establish that allegedly competing mental and physical causes do not overdetermine their effects. I actually do not think that think that it is necessary, either—as long as one believes that mental and physical causes are appropriately intimately related, I suspect that one can think that causation is as 'oomphy' as one likes and nonetheless claim that mental and physical causes do not overdetermine their effects—but I will set that aside for now.[21] All I will argue is that moving to a pure dependence notion is not by itself enough. The real work must be done by an appeal to the relation between the causes, à la the first strategy.

To see this, note that that most believers in a pure dependence theory *also believe that genuine overdetermination occasionally happens*. They think that classic firing squad cases do happen, and that they are importantly different from cases of mental causation. Consider, for example, the familiar point that simple counterfactual theories, according to which *c* is a cause of *e* just in case *e* would not have happened if *c* had not happened (Lewis 1973), do not allow overdetermining causes to count as causes at all. Such views are forced to say that all apparent cases of overdetermination are really cases of joint causation. They are forced into what Jonathan Schaffer calls the 'collectivist' view of overdetermination rather than the 'individualist' view (2003). But—and this is the crucial point—everyone thinks that this is a *problem*, and starts looking for a less simple counterfactual theory. Lewis did, for example (cf. 1986, 2000a). Everyone agrees that the right version of a dependence theory must accommodate genuine overdetermination.

But that means that the dependence theory alone cannot dismiss the charge that some particular effect is overdetermined. It says that sometimes effects have two causes and are

[21] The trick would be to claim that mental property instances (or events, etc.) and their physical realizers *only provide one injection of oomph*. Events, properties, and the like are individuated differently than are transfers of oomph. Two events, one dose of oomph. To see the idea, imagine two events, one a proper part of the other, such that the part constitutes what might be called an 'efficacious core': the other parts of the larger event are wholly inert. One might well want to say that both the larger and the smaller event are causally sufficient for some effect, but do not overdetermine it. And in such a case, surely one could say that even if causation was literally the transfer of a magic pellet from one event to the other. That was all by way of cartoon analogy, but do note that Sydney Shoemaker can probably think of causation as being as oomphy as he likes, while nonetheless maintaining his compatibilist solution to the exclusion problem (2001, forthcoming). Thus while I myself do not actually endorse the metaphysics of realization that adopting this strategy for the mental/physical case would require, I nonetheless think it is worth mentioning.

overdetermined, and that sometimes effects have two causes and are not overdetermined. *No* theory of causation that allows both cases can *all by itself* distinguish between them. Only information about the two causes—and, in particular, how they are related—can do so. A mere insistence that causation is not ooomphy cannot do the job; it cannot distinguish cases of mental causation from cases in which a person is simultaneously hit with two bullets from two independent shooters. So the mere appeal to a pure dependence theory of causation cannot itself establish that the exclusion principle is false and compatibilism is true. It cannot show that mental and physical causes do not overdetermine their effects.

Indeed, I am inclined to suspect that the only way in which the dependence view of causation can help is because anyone who endorses it will be amenable to my counterfactual test for overdetermination, and will consequently be amenable to my own version of the *first* strategy for motivating compatibilism. Be that as it may, the fact is that the only way to properly motivate compatibilism is by appeal to the tight relation between mental and physical causes. And once we go beyond simply asserting that tightly related causes cannot overdetermine their effects, and provide an actual *test* for overdetermination that some pairs of causes pass and others fail, we can see that *compatibilism requires physicalism*. The dualist cannot avail herself of the nonreductive physicalist's solution to the exclusion problem.

6. Objections and Replies

Some readers will object to my claim that my version of compatibilism works for physicalists. Other readers will object to my claim that it does not work for dualists. That is, some will protest that not even physicalists can dodge the exclusion problem in the way I have suggested. Some will agree that physicalists have a viable answer, but will insist that dualists can in fact help themselves to it too. Although the former sort of complaint is really more targeted against my 2003 than against my claims in this paper in particular, I will consider two of each sort of complaint, in reverse order.

1. *Your claim that dualists cannot endorse your compatibilist solution seems to rest on a rather small point. The Lewis-Stalnaker semantics for counterfactuals has to bear a lot of weight here. Can't I just reject it?*

If the dualist rejects the standard semantics for counterfactuals, she can disagree with the physicalist about which worlds there are without disagreeing with him about which is the closest

world in which the antecedent of (O2) is true. That is correct. The only question is whether that is an acceptable price to pay. It is not.

Bear in mind that not just any quibble with the Lewis story will help the dualist. The dualist's goal here, recall, is to insist that (O2)—(p & $\sim m$) $\Box\!\!\rightarrow$ e— is false despite the fact that she thinks there are metaphysically possible worlds in which p happens without m and without any change in surrounding circumstances or preceding occurrences. Consequently, what she needs to do is reject Lewis' way of reckoning the relative similarity of possible worlds. In particular, what she needs to do is insist that worlds with quite different matters of particular fact but perfect nomological match are closer to the actual world than are worlds with localized nomic violations but perfect match of particular fact.

But this similarity metric gets other counterfactuals wrong—and, indeed, gets other *causal* counterfactuals wrong. The dualist who opts for the modified similarity metric will be committed to backtracking evaluations being the norm rather than the exception. Consequently, any such dualist who is also tempted towards a counterfactual analysis of causation will have to work out a different response to the problems of effects and epiphenomena than that which Lewis offers. I simply direct the reader to Lewis' own defense of his similarity metric (especially 1973, 170-1; 1979, 43-48); I have nothing of substance to add. The dualist should listen to Lewis, and should not jettison that metric in order to count (O2) false. The cost is too high.

2. Suppose you're right that motivating compatibilism requires endorsing the claim that the mental supervenes upon the physical with metaphysical necessity. That's not to say that compatibilism requires physicalism.

The objection here is an objection to the definition of physicalism upon which I am relying. Although there is widespread agreement that physicalism does require the claim that everything supervenes upon the physical with metaphysical necessity, a number of people think it also requires more. They think that the definition I have provided, though necessary, is not sufficient for physicalism. If this is correct, it would follow that supervenience with metaphysical necessity is compatible with dualism, and thus that dualists who are willing to endorse it can be compatibilists after all.

Now, if worst comes to worst, I am willing to downplay my claim. If a version of dualism that accepts that the mental supervenes on the physical with metaphysical necessity is

both coherent and well-motivated, I will allow its proponents to help themselves to my solution to the exclusion problem. After all, it is the metaphysically necessary supervenience claim, and not any further requirements on physicalism, that is doing the work. If necessary, then, I am willing to downgrade my claim from

- compatibilism requires physicalism.

to

- compatibilism requires the metaphysically necessary supervenience claim.

But I am only willing to do this if it is necessary, and I am not convinced that it is. I do not think that there is any real reason to deny that the metaphysically necessary supervenience claim is sufficient for physicalism, and some reason to think that it indeed is sufficient.

Why would anybody think that it is not sufficient for physicalism? Let me quickly canvass a variety of reasons, some of which can be found elsewhere and some of which cannot. First, Jessica Wilson (2005) argues against the sufficiency claim in two stages. She begins by arguing that physicalists should be necessitarians about the laws of nature (with Shoemaker 1980, Swoyer 1982), and then argues that necessitarianism collapses the distinction between nomological and metaphysical necessity. But even assuming both the controversial necessitarian premise and the ensuing merging of the two grades of necessity, it is not clear why it would follow that supervenience with metaphysical necessity is not sufficient for physicalism. The mere claim that there is no real distinction between nomological and metaphysical necessity can only show that there cannot be any nomologically-but-not-metaphysically-necessary supervenience relations—and thus that Chalmers' version of property dualism (1996) is not coherent. It cannot itself show that a position that endorses a nomological-*and*-metaphysically-necessary supervenience claim can legitimately count as dualist. In fact, perhaps the proper upshot of Wilson's premises is that genuine dualists have to think that all connections between physical properties and mental ones have to be *completely*—even nomologically—contingent. Thus I do not think that Wilson has provided reason to believe that the metaphysically necessary supervenience claim is consistent with dualism.

Second, one might argue that metaphysically necessary supervenience cannot be sufficient for physicalism by appeal to a variety of technical features of supervenience itself. All of these are reasons to think that supervenience does not guarantee that everything that happens genuinely *depends* upon what happens at the most basic physical level, as physicalism surely

requires. For example, supervenience can hold symmetrically, but dependence is usually thought to be asymmetric. Further, there are various odd versions of global supervenience that are too weak to count as genuine dependence relations, *even if* they hold with metaphysical necessity (see my 2004). Neither of these are real concerns, however. At worst, we would simply need to specify which version of global supervenience is used to characterize physicalism, and add 'and not *vice versa*'—e.g. it is metaphysically necessary that everything strongly globally supervenes upon the physical, and not *vice versa*.

A more important threat to supervenience's ability to capture dependence claims is posed by necessary existents. Anything that exists necessarily exists regardless of what else exists, or what properties other things have. It follows that necessary existents supervene on anything whatsoever. For example, every two worlds that are just alike vis à vis the distribution of rutabagas will be just alike vis à vis whatever necessary existents you wish to countenance—perhaps God, or the number three. So God and the number three supervene on the distribution of rutabagas. But surely they do not in any intuitive sense *depend* on the rutabagas—we are assuming that they exist no matter what.

This is a real issue. I am not sure what best to say about it. I simply note two points. First, the very fact that there is no sense in which God or platonic numbers or what have you depend upon the physical means that physicalists should view them with suspicion, and arguably should repudiate them altogether (see Jackson 1998, 22-23). Second, even if the case does show that one set of properties can supervene upon another without depending on it, it does not obviously show that the *mental* can supervene on the physical without depending upon it. After all, no one thinks that mental properties or particular mental states exist necessarily. So this line of thought is not obviously relevant here.

A third argument derives from the idea that there are more informative characterizations of physicalism to be had. A variety of people have suggested that if everything supervenes with metaphysical necessity on the physical, there must be some explanation of why it does. Supervenience itself is simply a relation of property covariation, and it is not in general plausible to say that it is just a brute fact that two sets of properties covary with each other (see Blackburn 1984, 186; Horgan 1993; Kim 1993c, 167-168, Melnyk 2003). Andrew Melnyk, for example, thinks that the supervenience of the mental on the physical is best explained by the fact that each

instance of a mental property either is or is realized by an instance of a physical property. He consequently thinks that physicalism is best characterized in terms of realization.

Now, that is all well and good. I agree that supervenience claims typically require explanation, and am happy to grant for the sake of argument that realization provides the best explanation of the physicalist's claim that the mental supervenes on the physical with metaphysical necessity. But it is important to see that what this sort of argument at best shows is that the metaphysically necessary supervenience claim is not a sufficiently informative *characterization* of physicalism. It cannot show that the metaphysically necessary supervenience claim is not sufficient for the *truth* of physicalism. After all, it might be the case that metaphysically necessary supervenience guarantees that realization holds, which in turn means that physicalism is true.

Melnyk himself denies that metaphysically necessary supervenience guarantees that realization holds. He suggests, as does Frank Jackson, that the metaphysically necessary supervenience claim is consistent with dualism (Melnyk 2003, 58; Jackson 2006, 243). However, neither really argues for this. They both seem to take it to be obvious that a dualist could endorse that rather strong claim. However, I don't think it *is* obvious. Indeed, I find it quite hard to make sense of a version of dualism according to which the mental is both genuinely distinct from and metaphysically necessitated by the physical. I cannot understand what makes such a view properly *dualist*, because I cannot understand how genuinely distinct mental properties (let alone full blown souls!) could be metaphysically necessitated by the physical. I realize that this is not really an argument, of course. Settling the dispute in favor of either party probably requires settling the status of the Humean dictum that there cannot be necessary connections between completely distinct existences, and I do not intend to do that here. I simply note that I am not convinced that the view is coherent.

Further, it is not clear why anyone would endorse it even if it is coherent. Note, in particular, that this sort of dualist cannot appeal to the conceivability of zombies to help motivate their view. In fact, they would need to respond to the argument just as much as a physicalist does! Zombies are *not* metaphysically possible on this view. So the dualist would have to explain why they seem possible. Presumably, she would either claim that they will not seem possible upon properly informed reflection, or else that although they will always continue to seem possible, this indicates something about the structure of our concepts rather than about the

structure of reality. These are familiar options; they are the same options that the *physicalist* has. Indeed, we can borrow Chalmers' labels for types of physicalism (2002), and call these type A and type B metaphysical-necessitation-dualism, respectively.

So I am not convinced by any of the arguments in favor of requiring more from physicalism; I do not think that dualism is consistent with the metaphysically necessary supervenience of everything on the physical. I consequently do not see any compelling reason to backpedal from my claim that compatibilism requires physicalism.

3. Haven't you been paying attention? Kim's latest version of the problem is the 'supervenience argument'. His claim is that not only does the supervenience of the mental on the physical not help*, it actively* hurts*—it only makes the problem worse.*

It is true that I have focused upon the original exclusion problem, which Kim (1989a) extracted from Norman Malcolm (1968), and refined and ingeniously defended through the 1980s and 90s. It is also true that Kim has recently shifted his attention to what he calls the 'supervenience argument' (1998, 38-47; 2005, ch. 2), and has been quite explicit that he does not expect supervenience to do the nonreductive physicalist any good.[22]

However, the supervenience argument is best understood as a further twist on the exclusion problem. The exclusion problem proper only challenges the idea that mental causes sometimes have physical effects. The supervenience argument extends the reach of the exclusion problem to also challenge the claim that mental causes sometimes have *mental* effects. I am going to skim over the details, but here is the basic idea. Suppose some mental occurrence M_2 has a mental cause M_1. Either M_1 causes M_2 directly, or it does so by causing M_2's physical supervenience base P_2 to be instantiated. If it causes M_2 directly, we have a tension between the two 'determiners' of M_2. The competition between M_1 and P_2 is not exactly *causal*, and would only be ruled out by a modified version of the exclusion principle, but the point seems clear. P_2 itself guarantees M_2, so what need of M_1?[23] Alternatively, M_1 could cause M_2 by causing its supervenience base P_2 to be instantiated. Even aside from the appeal to some non-causal exclusion principle, it seems plausible to require this (see Kim 1998, 42-43). But the causal

[22] The supervenience argument does not affect dualism, with the possible exception of the unmotivated and possibly incoherent 'metaphysical-necessitation-dualism' discussed in conjunction with objection 2.
[23] In the latest version of this argument, Kim explicitly refuses to formulate any exclusion-like principle here (2005, 41n8).

completeness of physics tells us that P_2 must also have a sufficient physical cause… and we are back at the original exclusion problem.

However exactly it is fleshed out, then, the supervenience argument rests upon the exclusion problem. The point of the supervenience argument is simply that physicalists must think that mental-to-mental causation requires mental-to-physical causation. Since the exclusion problem challenges the possibility of mental-to-physical causation, it gives rise to a challenge to the possibility of mental-mental causation as well. But if the exclusion problem can be solved, the supervenience argument does not get off the ground. More precisely, if the exclusion problem can be solved in a way that does not presuppose the possibility of mental-to-mental causation, the supervenience argument does not get off the ground. (In my view, one important lesson of the supervenience argument is that 'dual explanandum' responses to the exclusion argument are in trouble (see my 2007).) I have argued for a solution to the exclusion problem that does not presuppose the possibility of mental-to-mental causation. The supervenience argument poses no additional threat.[24]

4. You're only getting out of the problem—if you are—by giving up on mental causation. You haven't said anything about how the mental can really be causally efficacious, and it is starting to feel as though its efficacy is only derivative. Isn't this at best a Pyrrhic victory?

Two points. First, recall my remarks in the first few pages to the effect that solving the exclusion problem does not require providing a positive account of the efficacy of mental events and properties. Second, the objector's underlying thought is correct: no one can say that mental and physical causes are completely independent of each other, and yet do not overdetermine their mutual effects. That is the truth at the heart of the exclusion problem.

Thus I am happy to acknowledge that the dualist has something the nonreductive physicalist does not have—namely, the claim that the mental is *independently* causally efficacious. Perhaps doing without independent efficacy is a disturbing thought. But the fact is that it is a mistake to think that a *physicalist* can say anything else. Physicalists need to bite this bullet for reasons having nothing to do with the exclusion problem. It is a direct consequence of

[24] Another way to put a similar point might be this. The supervenience of the mental on the physical only makes the original exclusion problem worse if the original exclusion problem works in the first place. I claim that the original exclusion problem only works if it is not the case that the mental supervenes upon the physical with metaphysical necessity. It follows that only views according to which the mental *does* supervene upon the physical, but *not* with metaphysical necessity, face an additional challenge from the supervenience argument. In short: the supervenience argument makes matters worse for Chalmers-style property dualists. It does not make matters worse for me.

their physicalism. Kim is surely right that physicalists need to accept something like his 'causal inheritance principle' (e.g. 1992, 326; 1998, 54).[25] That is, he is right to emphasize that physicalists cannot believe in causal powers that "magically emerge at a higher level and of which there is no accounting in terms of lower-level properties and their causal powers and nomic connections" (1992, 326). That is part of what it is to be a physicalist.

So the objector here needs to either stop deluding himself about the consequences of his physicalism, or else decide that he prefers dualism, all things considered. But if he prefers dualism because he thinks it is the only way to avoid epiphenomenalism, he must either deny the completeness of physics, or accept rampant overdetermination. That is the lesson of the exclusion problem. Compatibilism is not an option for him.

7. Conclusion

I have argued that the exclusion problem does not exert the kind of force on a physicalist that it does on a dualist. Dualists really do need to choose between systematic overdetermination, epiphenomenalism, and the incompleteness of physics. Nonreductive physicalists do not. Thus although the argument does provide pressure towards physicalism, it does not provide pressure towards reductive physicalism. A proper understanding of nonreductive physicalism—an understanding that puts the right emphasis on the 'physicalism', and does not get distracted by the 'nonreductive'—makes the exclusion problem look a lot less threatening.

One lesson to be drawn, then, is that these familiar diagrams are dangerous:

M1 → M2

↑ ↑

P1 → P2

They blur the line between emergentism and nonreductive physicalism, and mislead us into thinking that both views are in the same boat vis à vis the exclusion problem—which they are not. They obscure the fact that the upward arrows that symbolize Distinctness come to something rather different for those who endorse physicalism than for those who deny it. The

[25] The physicalist is only committed to the 'subset' version of the causal inheritance principle proposed in 1998, not the stronger 'identity' version of 1992.

dualist really does need to choose between denying Efficacy, Nonoverdetermination, and Completeness. The physicalist does not. And if, as seems likely, the dualist does have reason to endorse Completeness, I can get even closer to the ideal conclusion I discussed back in section 1. The exclusion argument is an enormous problem for dualist; not for those who say—and mean it—that the mental is nothing over and above the physical.

References

Bennett, Karen. 2003. Why the exclusion problem seems intractable, and how, just maybe, to tract it. *Noûs* 37: 471-497.

———. 2004. Global supervenience and dependence. *Philosophy and Phenomenological Research* 68:3, 501-529.

———. 2007. Mental causation. *Philosophy Compass* 2: 316–337.

Blackburn, Simon. 1984. *Spreading the Word*. Oxford: Oxford University Press.

Boyd, Richard. 1980. Materialism without reductionism: what physicalism does not entail. In Ned Block (ed.), *Readings in the Philosophy of Psychology, Vol. I*. Cambridge, MA: Harvard University Press.

Burge, Tyler. 1979. Individualism and the mental. *Midwest Studies in Philosophy* 4: 73-121.

Chalmers, David. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

———. 2002. Consciousness and its place in nature. Reprinted (2002) in David Chalmers, ed., *Philosophy of Mind: Classical and Contemporary Readings*. New York: Oxford University Press, 247-272.

Crane, Tim. 2001. The significance of emergence. In Carl Gillett and Barry Loewer, eds., *Physicalism and its Discontents*. Cambridge: Cambridge University Press, 207-224.

Crane, Tim, and Mellor, Hugh. 1990. There is no question of physicalism. *Mind* 99: 185-206.

Crisp, Thomas, and Warfield, Ted. 2001. Kim's master argument. *Noûs* 35: 304-316.

Dowe, Phil. 2000. *Physical Causation*. Cambridge: Cambridge University Press.

Hall, Ned. 2004. Two concepts of causation. In J. Collins, N. Hall, and L. A. Paul, eds., *Counterfactuals and Causation*. Cambridge, MA: MIT Press.

Hawthorne. John. 2002. Blocking definitions of materialism. *Philosophical Studies* 110: 103-113.

Hempel, Carl. 1980. Comments on Goodman's *Ways of Worldmaking*. *Synthese* 45:193-9.

Horgan, Terence. 1993. From supervenience to superdupervenience: meeting the demands of a material world. *Mind* 102: 555-586.

Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Clarendon.

———. 2006. On ensuring that physicalism is not a dual attribute theory in sheep's clothing. *Philosophical Studies* 131: 227-249.

———, and Phillip Pettit. 1990b. Program explanation: a general perspective. *Analysis* 50: 107–117.

Kim, Jaegwon. 1989a. Mechanism, purpose, and explanatory exclusion. Reprinted (1993) in *Supervenience and Mind*. Cambridge: Cambridge University Press, 237-264

———. 1989b. The myth of nonreductive physicalism. Reprinted (1993) in *Supervenience and Mind*. Cambridge: Cambridge University Press, 265-284.

———. 1992. Multiple realization and the metaphysics of reduction. Reprinted (1993) in *Supervenience and Mind*. Cambridge: Cambridge University Press, 309-335.

———. 1993a. The nonreductivist's troubles with mental causation. Reprinted (1993) in *Supervenience and Mind*. Cambridge: Cambridge University Press, 336-357.

———. 1993b. Postscripts on mental causation. In *Supervenience and Mind*. Cambridge: Cambridge University Press, 358-367.

———. 1993c. Postscripts on supervenience. In *Supervenience and Mind*. Cambridge: Cambridge University Press, 161-171.

———. 1998. *Mind in a Physical World*. Cambridge, MA: Bradford.

———. 2002. Responses. *Philosophy and Phenomenological Research* 65: 671-680.

———. 2003. Blocking causal drainage and other maintenance chores with mental causation. *Philosophy and Phenomenological Research* 67:1, 151-176.

———. 2005. *Physicalism, or Something Near Enough*. Princeton, NJ: Princeton University Press.
Lepore, Ernest, and Loewer, Barry. 1987. Mind Matters. *The Journal of Philosophy* 84: 630–642.
Lewis, David. 1966. An argument for the identity theory. *The Journal of Philosophy* 63: 17-25.
———. 1973. Causation. Reprinted (1986) in *Philosophical Papers Volume II*. NY: Oxford, pp. 159-172.
———. 1978. Review of Putnam. In Ned Block, ed., *Readings in the Philosophy of Psychology Volume I*. Minneapolis: University of Minnesota Press, 232–33.
———. 1979. Counterfactual dependence and time's arrow. Reprinted in *Philosophical Papers Vol. II*, NY: Oxford, 32-66.
———. 1983. New work for a theory of universals. *The Australasian Journal of Philosophy* 61: 343-377.
———. 1986. Postscripts to "Causation." In *Philosophical Papers Vol. II*, NY: Oxford, 172-213.
———. 2000a. Causation as influence. *Journal of Philosophy* 97: 182-197.
———. 2000b. Reduction of mind. In S. Guttenplan, ed., *A Companion to the Philosophy of Mind*. Oxford: Basil Blackwell, 412-31.
Loewer, Barry. 2001. From physics to physicalism. In Carl Gillett and Barry Loewer, eds., *Physicalism and its Discontents*. Cambridge: Cambridge University Press, pp. 37-56.
———. 2002. Comments on Jaegwon Kim's *Mind in a Physical World*. *Philosophy and Phenomenological Research* 65:3, 655-663.
Malcolm, Norman. 1969. The compatibility of mechanism and purpose. *The Philosophical Review* 78: 468-482.
Melnyk, Andrew. 2003. *A Physicalist Manifesto: Thoroughly Modern Materialism*. Cambridge: Cambridge University Press.
Merricks, Trenton. 2001. *Objects and Persons*. Oxford: Clarendon Press.
Papineau, David. 1995. Arguments for supervenience and physical realization. In Savellos and Yalçin, eds., *Supervenience: New Essays*. Cambridge: Cambridge University Press, 226-243.
———. 2001. The rise of physicalism. In Carl Gillett and Barry Loewer, eds., *Physicalism and its Discontents*. Cambridge: Cambridge University Press, 3-36.
———. 2002. *Thinking About Consciousness*. Oxford: Oxford University Press.
Pereboom, Derk. 2002. Robust nonreductive materialism. *Journal of Philosophy* 99: 499-531.
———. and Kornblith, Hilary. 1991. The metaphysics of irreducibility. *Philosophical Studies* 63: 125-145.
Place, U. T. 1956. Is consciousness a brain process? *British Journal of Psychology* 47: 44-50.
Putnam, Hilary. 1973. The nature of mental states. Reprinted (2002) in David Chalmers, ed., *Philosophy of Mind: Classical and Contemporary Readings*. New York: Oxford University Press, 73-79.
Salmon, Wesley. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.
Schaffer, Jonathan. 2003. Overdetermining causes. *Philosophical Studies* 114: 23-45.
Shoemaker, Sydney. 1980. Causality and Properties. In *Time and Cause*, ed. Peter van Inwagen, Dordrecht: D. Reidel: 109–35.
———. 2001. Realization and mental causation. In Carl Gillett and Barry Loewer, eds., *Physicalism and its Discontents*. Cambridge: Cambridge University Press, 74-98.
Smart, J.J.C. 1959. Sensations and brain processes. *The Philosophical Review* 68: 141-156.
Sturgeon, Scott. 1998. Physicalism and overdetermination. *Mind*107: 411-432.
Stoljar, Daniel. This volume. Distinctions in distinction.
Swoyer, Christopher. 1982. The Nature of Natural Laws. *Australasian Journal of Philosophy* 60: 203–223.

Unger, Peter. 1979. There are no ordinary things. *Synthèse* 4: 117-54
Van Inwagen, Peter. 1990. *Material Beings*. Cornell University Press, Ithaca.
Wilson, Jessica. 2005 Supervenience-based characterizations of physicalism. *Nous* 39: 426-459.
Yablo, Stephen. 1992. Mental causation. *The Philosophical Review* 101: 245-280.