Emerging from the Causal Drain

Richard Corry

This is a pre-print of an article published in *Philosophical Studies* 165 (1):29-47 (2013). The final publication is available at link.springer.com

Abstract

For over twenty years, Jaegwon Kim's Causal Exclusion Argument has stood as the major hurdle for non-reductive physicalism. If successful, Kim's argument would show that the high-level properties posited by non-reductive physicalists must either be identical with lower-level physical properties, or else must be causally inert. The most prominent objection to the Causal Exclusion Argument—the so-called Overdetermination objection—points out that there are some notions of causation that are left untouched by the argument. If causation is simply counterfactual dependence, for example, then the Causal Exclusion Argument fails. Thus, much of the existing debate turns on the issue of which account of causation is appropriate. In this paper, however, I take a bolder approach and argue that Kim's preferred version of the Causal Exclusion Argument fails no matter what account one gives of causation. Any notion of causation that is strong enough to support the premises of the argument is too strong to play the role required in the logic of the argument. I also consider a second version of the Causal Exclusion Argument, and suggest that although it may avoid the problems of the first version, it begs the question against a particular form of non-reductive physicalism, namely emergentism.

Keywords: *Jaegwon Kim, Non-reductive Physicalism, Causal Exclusion, Supervenience, Emergence, Causation.*

Introduction

In the philosophy of mind, and of the special-sciences more generally, non-reductive physicalism is an attractive position, since it holds out the promise of allowing us to have our cake and eat it too. For the non-reductive physicalist holds the hard-headed view that the world is nothing over-and-above the physical, but at the same time holds that the properties studied by the special sciences are real, causally active, properties that are not reducible to physics. Like the ability to eat one's cake without thereby destroying it, however, non-reductive physicalism may seem too good to be true. Indeed Jaegwon Kim (1989; 1990; 1992; 1993a; 1993b; 1998; 2003; 2005; 2006a; 2006b; 2006c) has famously argued that non-reductive physicalism is incoherent, and his "Causal Exclusion", or "Supervenience" argument has stood as the major hurdle for non-reductive physicalism for over twenty years.

There have, of course, been many objections to Kim's argument. In fact the literature on this debate is so extensive that I step in with some trepidation. The most prominent objection to the Causal Exclusion Argument—the so-called Overdetermination objection—points out that some analyses of causation render the argument unsound (Burge 1993; Horgan 1997; Loewer 2001; Crisp and Warfield 2001). Kim's response is that none of these analyses of causation capture the notion that is important in philosophy of mind (For example, Kim 2005, 38). And so the debate becomes one of the correct account of causation in this context.

In this paper, I take a bolder approach and argue that Kim's preferred version of the Causal Exclusion Argument fails, no matter what account one gives of causation. I also consider a second version of the Causal Exclusion Argument, and suggest that although it may avoid the problems of the first version, it begs the question against a particular form of non-reductive physicalism, namely emergentism.

The Causal Exclusion Argument

Non-reductive physicalism is the view that there are some properties—call them "high-level properties"—of which the following three claims are true:

> *Supervenience*: High-level properties strongly supervene on physical properties.
>
> *Irreducibility*: High-level properties are not reducible to, and are not identical with, physical properties.
>
> *Causal Efficacy*: High-level properties are causally efficacious. That is, their instantiations can, and do, cause other properties to be instantiated.

The Supervenience claim captures a minimum requirement for non-reductive physicalism to count as a form of physicalism at all. Physicalists hold that all properties, and in particular high-level properties, are grounded in physical properties. Whatever "grounded in" means in this context it plausibly implies that mental properties will supervene on physical properties. Kim spells out what he means by this as follows (where M represents a high-level property):

> If M is instantiated in *s* at *t* then, necessarily, there is some physical property P instantiated in *s* at *t*, and anything instantiating P at any time instantiates M at that time (Kim 2005, 33).

Note that this notion of supervenience involves more than just necessary covariation between mental properties and certain physical properties. The claim also requires that the mental and physical properties in question are instantiated by the same entity at the same time. This requirement of co-location is plausible as a requirement of physicalism, and Kim passes over it without comment. I do not intend to question the requirement here, but there are those who might endorse supervenience without the simultaneity requirement (see, for example, O'Connor and Wong 2005), and as we will see, the requirement of simultaneity plays an important role in the Causal Exclusion Argument, a role that Kim does not make explicit.

One final clarification of the supervenience claim is that the necessity Kim has in mind here is nomological necessity; *given the laws of nature* there is no way to have P without also having M etc.

The claim of irreducibility is obviously meant to capture the "non-reductive" component of non-reductive physicalism. There is quite a bit of literature on the question of what, exactly, one might mean by "reducible" in a context like this (Nagel 1949; Wimsatt 1979; Sarkar 1992), but fortunately for us the details will not matter. For Kim's argument only makes use of the claim that mental properties are not identical with physical properties.

Kim's argument has two stages. The first stage establishes that if higher-level properties are to act as causes at all, then they must be involved in "downwards causation". That is, they must cause changes in the underlying physical properties of the world. The second stage of the argument attempts to show that downwards causation is incompatible with the non-reductive physicalist's claims of supervenience and irreducibility.[1]

When presenting his arguments, the high-level properties Kim has in mind are mental properties. For the sake of consistency, therefore, I too will focus on mental properties from here on. But it should be understood that the arguments below apply to high-level properties more generally. All one must do is replace "mental" with "high-level".

Stage 1: Downwards Causation

Suppose that some mental property M causes mental property M* (I will follow Kim here and use the locution "property M causes property M*" as a shorthand for "the instantiation of property M causes the instantiation of property M*"). The *Supervenience* thesis implies that M* will have some physical supervenience base P*. Since M* cannot

---

[1] Recently, Kim (2006b) has used the name "Supervenience argument" to refer to the first stage, and "Exclusion Argument" to refer to the second.

be instantiated without some such P* being instantiated, it follows that one cannot bring about M* without also bringing about some P*. So mental-to-mental causation necessarily involves mental-to-physical causation. This situation is shown in Figure 1.

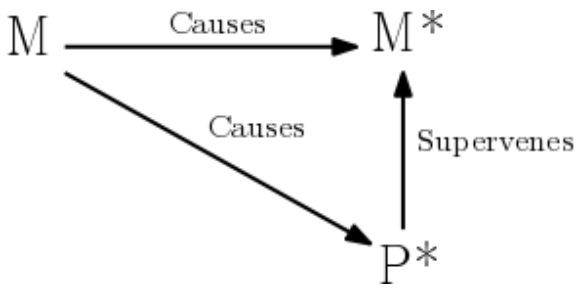

Figure 1 Downwards Causation

Kim actually takes the argument one step further before drawing his conclusion about downwards causation. He notes that P* is sufficient for M* and remarks that there is therefore some tension between M and P*—both seem to have a claim to be the reason for M* being instantiated. The only way to resolve the tension, he says, is to conclude that M causes M* *by* causing P*. So, in a sense, mental-to-mental causation *just is* mental-to-physical causation.

I have no quibble with this first stage of Kim's argument and will not discuss it further (but see Crisp and Warfield 2001; Wong 2010 for objections). I will note, however, that even if there is something wrong with the detail of the argument, downwards causation is

something that will be independently accepted by almost anyone who is committed to the causal efficacy of the mental. For we want to be able to say that my intention to raise my arm (a mental state) was the cause of my arm raising (a physical state). The claim that there are never cases of downwards causation is thus uncomfortably close to the very epiphenomenalism that is at stake in the Causal Exclusion Argument.

Stage 2: Excluding downwards causation

The second stage of Kim's argument seeks to establish that downwards causation is incompatible with *Supervenience* and *Irreducibility*. In this stage the following two principles make an appearance:

> *Exclusion*: No single event can have more than one sufficient cause occurring at any given time—unless it is a genuine case of causal overdetermination.(Kim 2005, 42)
>
> *Closure*: If a physical event has a cause at *t*, then it has a physical cause at *t*.(Kim 2005, 15)

Kim introduces these principles with very little argument, commenting that they will be accepted by almost everyone who is tempted by non-reductive physicalism. He treats *Exclusion* as an analytic truth, while *Closure* is taken to be both very plausible, and constitutive of physicalism. Perhaps unsurprisingly, the truth of these principles, particularly *Exclusion*, have come under debate in the literature and I will consider their merits shortly.

There are actually two logically distinct versions of the second stage of the Causal Exclusion Argument, though they have not been well distinguished in the literature. Both versions make use of *Exclusion*, but one, which I will call the *Direct Argument*, does not rely on *Closure*. In fact the Direct Argument is often presented in a way that makes use of *Closure*, but, as we will see, the reference to *Closure* is not essential. The Direct Argument is perhaps Kim's preferred version, as it is the one that appears in most of his publications on the topic. In what follows, however, I will show that the Direct Argument

(whether in its pure form, or mixed with *Closure*) involves a crucial non sequitur and does not pose a threat to any version of non-reductive physicalism.

The second version of Stage 2, which only appears in Kim's more recent writings, does make essential use of *Closure*, and I will therefore refer to it as the *Argument from Closure*. Because of it's reliance on *Closure*, this argument avoids the problems of the Direct Argument. I will show, however, that the Argument from Closure begs the question against one form of non-reductive physicalism, namely emergentism.

The Direct Argument

Consider a case in which some mental property M causes some physical state P*. By S*upervenience*, M has some supervenience base P, as in Figure 2.

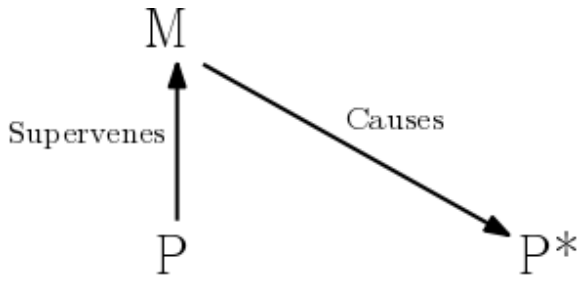

Fig. 2 Mental Causation

Since M nomologically supervenes on P, P is nomologically sufficient for M. Now suppose that causation is nomological sufficiency. Then M is, by supposition, nomologically sufficient for P*. But if P is nomologically sufficient for M and M is nomologically sufficient for P*, then P is nomologically sufficient for P*. Hence if causation is just nomological sufficiency, then P is a cause of P*. We get the same conclusion, says Kim, if causation is counterfactual dependence (in the sense that P* would not have occurred if P had not).

Now, Kim argues, the situation cannot be one in which there is a causal chain from P to P* via M, since the relation of supervenience between P and M is not a causal one.[2] Furthermore, P and M cannot be considered as parts of a jointly sufficient cause of P*, since P and M are each sufficient on their own. Hence M and P are *independent* sufficient causes of P*. But, P and M cannot both be independent sufficient causes of P*, since this would violate *Exclusion*.

The situation, then, is that if M has any causal efficacy at all, then either all its effects are genuinely overdetermined, or we have a violation of *Exclusion.* But says Kim, it would be bizarre if every case of mental causation were a case of genuine overdetermination. Furthermore, such a situation would make mental causes superfluous, they certainly wouldn't be adding anything new.

Ruling out systematic overdetermination, then, the supposition that M is causally efficacious leads to a violation of *Exclusion*. Since *Exclusion* is supposedly an analytic truth, it therefore follows that, M cannot have causal efficacy.

The Mixed Argument

I have presented the Direct Argument as a reductio of the supposition that M has causal efficacy. Kim only presents the argument this way in a few places (1993a; 1993b; 2006c). More commonly, Kim tends to present what is a mixture of the Direct Argument and the Argument from Closure (See his (1992; 2006a; 1998) and "completion 1" of the argument in (2005; 2003)). This mixed argument proceeds as follows: First Kim reasons as above to the conclusion that if M is a cause of P*, then so is P. He then employs Exclusion to conclude that either M or P must be disqualified as a cause. At this point, rather than performing a reductio to argue that it is M that must be disqualified, Kim

---

[2] This is supposed to be part of the concept of non-reductive physicalism, though O'Connor and Wong (2005) disagree.

makes use of Closure to choose between P and M. In particular, Closure tips the scales in favour of P, since if we drop P as a cause of P*, we must look for another physical cause, which will then be in competition with M, landing us back where we started.

This mixed version of the argument simply adds (superfluous) steps to the pure form of the Direct Argument, hence, if the pure Direct Argument fails, so too will the mixed version. For this reason I will focus my attention on the pure Direct Argument. Furthermore, since the Direct Argument makes no essential reference to *Closure*, I will not consider this principle further until I discuss the Argument from Closure.

A first objection

As it stands, there is a mistake in the Direct Argument. To begin to see the problem, note that the conclusion that P is nomologically sufficient for P* is derived by chaining together the supposition that P is nomologically sufficient for M with the supposition that M is nomologically sufficient for P*. So if causation is simply nomological sufficiency, then—contra Kim—P *is* a cause of M, and the causal link between P and P* is a causal chain that goes via M.

So Kim is wrong to claim that there is no causal chain from P to M to P*, but is this a problem for his argument? If we take the letter of the *Exclusion* principle, the answer is no. As stated above, the principle will rule out P and M both being causes of P*, since they occur at the same time. One might wonder, then, why, in every presentation of the direct argument, Kim goes to the trouble of arguing (mistakenly) that there is no causal chain from P to M to P*. The reason is that *Exclusion* is not intended to apply to cases involving causal chains. Kim himself states that "Two conditions can each be a sufficient cause of some single event by being different links in the same causal chain leading to the effect event" (1990, 40). So, for example, if I were to throw a brick at a window, we would happily cite both my action, and the striking of the window by the brick, as causes of the window breaking.

According to List and Menzies (2009, fn 7) the point of restricting *Exclusion* to simultaneous events is precisely to rule out its application to causal chains. There is some evidence that this was indeed Kim's motivation since, in the early discussion of the Exclusion argument, Kim endorses Alvin Goldman's suggestion that a similar exclusion principle proposed by Norman Malcolm be amended to apply only to simultaneous events, so as to rule out cases involving causal chains (Kim 1989, 82). Of course causal chains will be ruled out by this move only if simultaneous causation is not possible, and this impossibility is not guaranteed if we take causation simply to be nomological sufficiency.

Exclusion is a metaphysical descendant of Kim's earlier "Principle of Explanatory Exclusion" which states that "two or more *complete* and *independent* explanations of the same event or phenomenon cannot coexist "(1989, 89). Kim does not offer a definition of what is meant by "independent" here, but he does tell us that two causal explanations will fail to be independent if the explanans of one is causally dependent on the explanans of the other. Following this lead we can capture the spirit of *Exclusion* without making any assumptions about the possibility of simultaneous causation by restricting its application to *independent* sufficient causes, where A is independent of B only if neither is causally dependant on the other.

The problem with Kim's direct argument, then, is that if causation is simply nomological sufficiency, P and M are not independent in the relevant sense and *Exclusion* does not apply.

A similar problem arises if we take causation to be counterfactual dependence. Kim claims that P* is counterfactually dependant on P and hence is a cause of P. But why should we think this is the case? If causation is counterfactual dependence, then the supposition that M causes P* will tell us that P* is counterfactually dependant on M. Since M and P are distinct, to draw the conclusion Kim does, we also need to claim that M counterfactually depends on P. But there are two problems with this last claim. First,

the fact that M nomologically supervenes on P does not imply that M counterfactually depends on P. There may be more than one possible supervenience base for M, and if P had not been instantiated, one of these other supervenience bases might have been.[3] On the other hand, if M *does* counterfactually depend on P, and if causation is just counterfactual dependence, then P is a cause of M, and—as in the nomological sufficiency case—M is simply a link in a causal chain, so *Exclusion* does not apply.

So, if causation is either nomological sufficiency or counterfactual dependence (and if we ignore the possibility of alternative supervenience bases), then there is a causal chain from P to M to P*. So, although both P and M are sufficient causes of P*, they are not independent, and thus their joint existence does not violate *Exclusion*.

Clearly, the claim that the link between P and M is not causal is crucial to the Direct Argument. Kim simply stipulates that the link is not causal and proceeds from there. But the argument relies on a particular analysis of causation to establish that P is a cause of P*, so it is crucial that this analysis of causation is consistent with the stipulation. What the preceding considerations point out is that one cannot consistently make this stipulation whilst holding that causation is simply counterfactual dependence or nomological sufficiency. If P is not to be a cause of M, then we need a thicker notion of causation. In particular, the Direct Argument requires a notion of causation that will tell us that if M is a cause of P*, then P is a cause of P* but is not a cause of M.

There is, however, a simple way to constrain the notion of causation so that it can do the job. All we need do is take the simple counterfactual or nomological-sufficiency analysis of causation and add the requirement that effects cannot be simultaneous with their

---

[3] Kim recognises this objection but says, without explanation, that "we may assume, without prejudice, that no alternative physical base of M would have been available on this occasion". This assumption may not be so innocent. List and Menzies (2009), for example, have argued that *Exclusion* may fail when there are alternative supervenience bases available. For the purposes of this paper, however, I will grant Kim this point.

causes. Indeed Kim cites the simultaneity of P and M as a reason for claiming that P is not a cause of M (1998, 44).

We saw above that the Direct Argument establishes that P is nomologically sufficient for P*, and (ignoring the problems of multiple instantiation) that P* is counterfactually dependant on P. So, if we assume a counterfactual or nomological-sufficiency analysis of causation constrained only so that cause and effect cannot be simultaneous, then it will follow that P is a cause of P* so long as the two are not simultaneous. But M is, by supposition, a cause of P* and so cannot be simultaneous with P* and we saw above that Kim's notion of supervenience requires that P and M are simultaneous, so it follows that P is not simultaneous with P*. Thus, this minimally thicker notion of causation implies that P is a cause of P*, just as Kim requires. On the other hand, since P and M are simultaneous, P is ruled out as a cause of M and so we avoid the problem of a causal chain from P to M to P*.

A less plausible approach might be to constrain the counterfactual or nomological-sufficiency analyses of causation by simply stipulating that a property cannot be caused by it's supervenience base. A very similar argument to that in the previous paragraph would then show that if M is a cause of P*, then P is a cause of P*, but P is not a cause of M.

All of these slightly thicker notions of causation seem to deliver what Kim needs for the direct argument: they can be used to conclude that P is a cause of P*, and that P is not a cause of M and hence that there is no causal chain from P to M to P*. It might seem then that it is time to debate the plausibility of these analyses of causation, but as far as the Direct Argument is concerned, there is a deeper problem. For the notions of counterfactual dependence and nomological sufficiency—even when enriched with stipulations about non-simultaneity or the like—are simply not strong enough to support the principle of *Exclusion.*

Digging Deeper—Causation and Exclusion

It is well known that events can counterfactually depend on more than one other event. Indeed, this is typically the case: If I hadn't struck the match it wouldn't have lit, but nor would it have lit if there had not been oxygen in the air; if Thor hadn't thrown a brick at the window the window wouldn't have broken, but nor would it have broken if Winifred had not removed the plywood that was covering it; and so on. So if causation is just counterfactual dependence, then there is no reason to expect *Exclusion* to hold. Furthermore, adding a stipulation about non-simultaneity will not suddenly make *Exclusion* true either. So if causation is counterfactual dependence, possibly with a requirement that cause and effect not be simultaneous, then the failure of *Exclusion* is no reason to doubt that high-level properties can have causal efficacy.

The fact that counterfactual dependencies need not exclude one-another forms the basis for some versions of the Overdetermination Objection to the Causal Exclusion Argument.[4] This objection simply points out that *Exclusion* does not apply to counterfactual analyses of causation, and then argues that a counterfactual analysis of causation is appropriate in this context. Kim responds that in discussions of mental causation, the notion of causation that is of interest is much "thicker" than mere counterfactual dependence. Kim says that the notion of causation at play here is Elizabeth Anscombe's (1971) notion of *productive causation* and comments that this is "in many ways a stronger relation than mere counterfactual dependence" (Kim 2005, 18). Unfortunately neither Kim nor Anscombe provide much detail on this notion of causation. One obvious candidate for this stronger notion, however, is nomological sufficiency—for this is the other notion of causation that is explicitly mentioned in the Direct Argument, and *Exclusion* makes reference to "*sufficient* causes". However—and

---

[4] For example (Loewer 2001; Crisp and Warfield 2001)

this is a point that seems to have gone unmentioned in the literature—*Exclusion* is not true of nomological sufficiency either.

To see that nomologically sufficient conditions need not exclude each other, consider the Stern-Gerlach experiment, which can be used to detect the fundamental property known as spin. The experiment involves firing a beam of particles through an inhomogeneous magnetic field; the spin of each particle interacts with the magnetic field and the particle is deflected. Now, if the particles involved are spin ½ particles, such as electrons, protons and neutrons, each particle will be deflected in one of two directions, and the beam will thus be split in two. Thus being a beam of particles with spin ½ is nomologically sufficient for being split in two by the Stern-Gerlach apparatus. (Note also that a beam composed of particles with a different value of spin will not split in two—so splitting in two is also counterfactually dependant on the particles having spin ½).

But now consider that electrons and positrons are the only particles with a rest mass of 0.511 Mega Electron Volts (MeV), and that all electrons and positrons have spin ½. These two facts are plausibly consequences of the fundamental laws of nature (we don't yet have a satisfactory theory of why the fundamental particles are grouped into the types we observe, but it certainly doesn't seem accidental). But if it is a consequence of the fundamental laws of nature that all particles with a rest mass of 0.511 MeV have spin ½, then being a beam composed of particles with a rest-mass of 0.511 MeV is nomologically sufficient for being split in two by the Stern-Gerlach apparatus.

We thus have a situation in which two separate properties are each nomologically sufficient for the beam being split in two, in violation of *Exclusion*. In general, whenever two properties are connected as a matter of nomological necessity, we can expect cases in which each of these properties is nomologically sufficient for some third property. Thus *Exclusion* is false if causation is understood as nomological sufficiency (and once again, adding the stipulation of non-simultaneity will not change this).

One might object that *Exclusion* only applies to independent sufficient causes, and since having a rest mass of 0.511 MeV is nomologically sufficient for having spin ½, these two properties are not independent in the relevant sense. This tactic might indeed save the principle of *Exclusion*, but it would prove fatal for the Causal Exclusion Argument. For if being connected by nomological sufficiency renders two properties immune to the principle of *Exclusion*, then the principle cannot be applied to the pair consisting of a high-level property and its physical supervenience base, as the supervenience base is, by definition, nomologically sufficient for the high-level property.

So far, then, we have seen that counterfactual-dependence and nomological sufficiency are too weak to support either the logic of the Direct Argument or the principle of *Exclusion*. Enriching these concepts of causation with stipulations of non-simultaneity or non-supervenience will allow them to support the logic of the argument, but they are still not strong enough to support the principle of *Exclusion*. So are there richer notions of causation that will do the job?

One simple way of developing a richer notion of causation is to insist that A is a cause of B if and only if it is *both* the case that A is nomologically sufficient for B *and* that B counterfactually depends on A. However, List and Menzies (2009) have considered a notion of causation along these lines (which they call "difference-making") and they show that *Exclusion* is not true in general even of this stronger notion of causation. Their argument for this conclusion hinges on the possibility of multiply realisable high-level properties. But we can see that *Exclusion* is not an analytic truth about causation as difference-making even if we do as Kim suggests and put aside issues of multiple realisability. Consider again the Stern-Gerlach apparatus described above. We have seen that being a beam of particles of mass of 0.511 MeV and being a beam of particles having spin ½ are both nomologically sufficient for being split in two by the apparatus. It is also the case that if the beam is not composed of spin ½ particles it will not be split in two. So being composed of spin ½ particles counts as a difference-making cause of splitting in two. But now suppose that electrons and positrons were the only particles with spin ½. In

this case it would also be true that if the particles did not have mass 0.511 MeV, the beam would not split in two. Since, as we have seen, being composed of 0.511 MeV particles is also sufficient for being split in two, being composed of 0.511 MeV particles would also count as a difference-making cause of the beam splitting in two. Thus, in a world where only electrons and positrons have spin 1/2 *Exclusion* would be false.

Whether or not the Stern-Gerlach experiment provides a counter-example to *Exclusion* in the actual world, where there are other particles which have spin ½ is harder to judge. To do so we need to consider the truth of the counterfactual "If the beam had not been composed of particles with a mass of 0.511 MeV, then it would not have split in two." To assess this counterfactual we need to know which worlds are most similar to the actual world when the experiment is done with electrons or protons. If, in such worlds, the experiment is done with other spin ½ particles, the counterfactual will be false, if on the other hand, the most similar worlds involve the experiment being done with particles with some other spin value (many ions, for example, have values of spin greater than ½), or a different apparatus, then the counterfactual would be true. Surely there have been situations in which the nearest worlds are of the latter sort. So, combining nomological sufficiency and counterfactual dependence does not give us a notion of causation that supports *Exclusion*.

Suppose then that we simply take it as given that *Exclusion* is constitutive of whatever thick notion of causation is at play in discussions of mental causation, and following Kim, let us call this notion productive causation. From the discussion above, it follows that the existence of a relation of productive causation cannot be implied by relations of counterfactual dependence or nomological sufficiency, either separately or jointly, for these relations do not obey *Exclusion*. We have also seen that adding information about simultaneity or supervenience relations will not produce something that satisfies *Exclusion*. So productive causation is stronger than any mixture of counterfactual dependence, nomological sufficiency, non-simultaneity, and non-supervenience.

But consider again the Direct Argument. A crucial step of the argument is the inference from the information contained in Figure 2 to the conclusion that there is a causal relation between P and P*. But Figure 2 does not tie P and P* together with anything stronger than nomological sufficiency, counterfactual dependence, simultaneity, and supervenience. The information in Figure 2 is, therefore, insufficient to licence the conclusion that there is a relation of productive causation between P and P*.[5]

For a concrete example of the problem, consider David Armstrong's (1997) view of causation. Armstrong argues that causation is best understood as a relation that holds between universals. So, for example, to suppose that M caused P* would be to suppose that there is a relation of 'necessitation' between the universals *being a mind in mental state M* and *being a brain in physical state P**. Following Armstrong, let us represent the holding of this relation as N(M,P*). The relation, N, is a primitive, and not further analysable, though he argues that we have direct experience of instances. Armstrong uses the language of production and necessitation to describe N, and so it may well be the sort of "thick" notion of Causation that Kim has in mind (though he admits a weaker notion of causation defined as the ancestor of N).

I am not aware of Armstrong taking an explicit stand on the truth of *Exclusion*, however he does say the following:

> Suppose, then, that N(F,G) and N(G,H), where both are deterministic laws. It by no means follows, and will very likely not be true, that N(F,H). Given an instantiation of F, and given that the laws are iron ones, one can *infer* the instantiation of H in a suitable relation to the F. One can say that in these circumstances it is nomically necessary that H be so instantiated, and even that it is a law that Fs and Hs are so linked. But for N(F,H) to hold in these circumstances would be for the instantiation of H to be *overdetermined.*(Armstrong 1997, 234–5)

---

[5] Yes, one step in the chain from P to P* involves productive causation, but this will not help.

Although Armstrong here considers only one kind of situation, it seems fair to presume that, like Kim, he does not think that genuine overdetermination is generally uncommon. Let us, then, suppose that *Exclusion* is true of Armstrong's notion of cause. Do we now have a notion of causation that can do the work required in Kim's Direct Argument?

No. For Kim's argument to go through we need to be able to infer from the information given in Figure 2, that P is a cause of P*. Let S(P,M) represent the relation of supervenience that holds between M and P, then from the premises S(P,M) and N(M,P*) we need to infer that N(P,P*). But there is nothing about N that would license this inference. N is a primitive relation, and since we are supposing that $M \neq P$, it is at least logically possible that N could hold between M and P* without also holding between P and P*. Indeed the situation here seems very similar to the one Armstrong considers in the quote above, and so, I would suggest, Armstrong would deem it very unlikely that in this situation N(P,P*). Given P, we may be able to infer that P* will be instantiated. We may be able to say that in these circumstances it is nomically necessary that P* be so instantiated, and even that it is a law that Ps and P*s are so linked. But none of this implies that N(P,P*).

The problem here is general. Any notion of causation that is strong enough to satisfy *Exclusion* will necessarily be richer than any of the relations contained in Figure 2 and hence cannot be inferred to hold between P and P*. As far as the Direct Argument is concerned, then, the question of which analysis of causation is appropriate in the context of mental causation is beside the point. There is no notion of causation that will support the Direct Argument.

The Argument from Closure

The Argument from Closure uses *Closure* to infer that there is a relation of causation between P and P* without making reference to any specific analysis of causation (See (Kim 2006b), and "completion 2" of the argument in (Kim 2003) and (Kim 2005)). As

such it is worth exploring whether the Argument from Closure can avoid the problems of the Direct Argument.

To see why *Closure* is problematic for downward causation, suppose that some mental property M is a putative downward-cause of some physical event P*. *Closure* implies that P* has a physical cause—call it P—and *Irreducibility* implies that P is distinct from M. Thus there is some P, distinct from M, which causes P*. So the assumption that M is a downward-cause of P* leads to the conclusion that P* has two distinct sufficient causes: P and M. Once again, then, the assumption that M is an irreducible mental cause of P* leads to a violation of *Exclusion*, and hence the assumption must be wrong. So either M is reducible after all, or else it cannot act as a downward-cause. And since all mental causation involves downward causation (from Stage 1), this implies that irreducible mental properties cannot be causes at all, and hence that non-reductive physicalism is false.

I have presented the argument here as a reductio: Assuming M has causal efficacy leads to a violation of *Exclusion*, hence M does not have causal efficacy. As with the Direct Argument, however, Kim presents this argument slightly differently. He invokes *Exclusion* to conclude that one of P and M must not be a cause, and then invokes *Closure* a second time to argue that it is M that must be dropped. I prefer the reductio presentation, as it is simpler and clearer, but my comments below will apply to both versions.

Objecting to Closure

Strictly speaking, what the Argument from Closure shows is that non-reductive physicalism is incompatible with the conjunction of *Closure* and *Exclusion*. Let us put aside arguments over the truth of *Exclusion* and turn our attention instead to *Closure.*

So is there any compelling reason why *Closure* should should be accepted? Kim claims that “Most Philosophers, including anyone who considers himself or herself a physicalist

of any kind, accepts physical causal closure" (2006b, 195). He points out that if we deny closure, then we are asserting that "an ideally complete physical theory will not be able to give an account of all physical phenomena" (2006a, 200). Such a claim does seem like an unfortunate thing for a physicalist to have to say. Indeed, we would have good reason to deny that anyone who makes such an assertion is really a physicalist. In light of such considerations, let us, for the moment, grant that *Closure* is constitutive of physicalism.

In so far as they are physicalists, then, non-reductive physicalists cannot simply deny *Closure*. But there is a different strategy that could be taken here: non-reductive physicalists can accept *Closure* but deny that it applies in this case.

The Argument from Closure relies on the claim that P is distinct from M, for we will not have two competing causes if M and P are not distinct. The reasoning that leads to this claim seems straightforward: M is, by supposition, a mental property, *Closure* requires that P is physical property, and *Irreducibility* states that mental properties are not identical to physical properties. But this chain of reasoning only goes through if "physical" has the same meaning in the statement of *Irreducibility* as it does in the statement of *Closure*. However, it is not obvious that this condition holds.

The non-reductive physicalist is, after all, a physicalist. As such, she holds that high-level properties are not something external to the physical world. It would seem natural, then, for the non-reductive physicalist to insist that mental properties *are* physical, at least in the sense involved in *Closure*. This position might seem in conflict with *Irreducibility*, but this conflict is due to an equivocation over the word "physical". We can avoid such equivocation by interpreting *Irreducibility* to mean that high-level properties cannot be reduced to *other* physical properties. This is just the kind of position that is usually attributed to emergentism. Indeed, we can characterise emergentism as subscribing to the following three claims:

> *Supervenience (emergence)* Emergent properties strongly supervene on fundamental physical properties;

> *Irreducibility (emergence)* Emergent properties of a system are not reducible to, and are not identical with, the fundamental physical properties of the system's components;
>
> *Causal Efficacy (emergence)* Emergent properties grant novel causal powers to a complex system (over and above the powers granted by the supervenience base of the property).

Understood like this, the parallel between emergentism and non-reductive physicalism is clear. Indeed emergentism so-construed just is a form of non-reductive physicalism (and if my argument of the preceding paragraph is correct, it is the most natural form of non-reductive physicalism). For this reason Kim has claimed in numerous places that the Causal Exclusion Argument applies to emergentism (1992; 1993a; 2006a; 2006c).

However, the emergent version of *Irreducibility* claims only that emergent properties are distinct from the *fundamental* physical properties, not that they are distinct from all physical properties. Thus the way is open for the emergentist to accept *Closure*, but insist that M is a physical property in the relevant sense and hence that M is the physical cause of P*.

At this stage one might worry that the debate has degenerated into an argument about the meaning of the word "physical". It has long been recognised that there is no easy way to give an adequate definition of the word "physical" such that it does justice to physicalist intuitions (see Crane and Mellor 1990), but surely we can make some sense of Kim's worries that emergentism adds something to the physicalist picture of the world. Indeed one anonymous referee has complained that if emergentists are allowed to claim that their emergent properties are physical, then there is no reason why the substance dualist should not do likewise: claim that mental substances and properties are just as "physical" as those studied by physics and so embrace C*losure*. Given the lack of an adequate definition of "physical" it is hard to give any principled reason why the substance dualist should be denied this strategy. Nonetheless, there are some obvious features of substance dualism that would worry anyone who might call themselves a physicalist. First, substance dualism posits new substances that are different in kind to any of the

substances studied by physics, and which do not share many of the properties (like mass or position) that are typically had by the objects of physics. Second, substance dualism posits new basic kinds of property that are typically very unlike the properties described in physics. Third, the properties of mental substances are typically held to be somewhat independent of the arrangement of the substances and properties described by physics (so as to allow for free will). Thus substance dualism builds in a division that is naturally interpreted as a division between the physical and the non-physical.

Emergentism, on the other hand, does not necessarily share any of the above mentioned features of substance dualism. With regard to the first, emergentism does not posit any new kinds of substance. Everything, says the emergentist, is, or is composed from, the kinds of stuff that is the subject of fundamental physics (particles, fields, M-branes or whatever).With regard to the third, emergent properties supervene on fundamental physical properties, so fixing the fundamental physical state of the world also fixes any emergent properties. With regard to the second feature of substance dualism, emergentism too may add irreducible properties that are very different in kind to the fundamental physical ones (think qualia) but this is not necessarily so. Emergent properties could be the very same properties that are found in fundamental physics (for example if an entity composed entirely of electrons in the right configuration were to take on a positive electric charge), or could just add new forces that are similar to those described by fundamental physics (in the same way that the discovery of the strong and weak nuclear forces added new forces to physics).

The only thing that emergentism per se adds to the standard physical picture of the world is the idea that irreducible properties can attach to complex entities, so if the physicalist is going to object to emergentism, it is this feature that must be rejected. Thus, if the emergentist is to be accused of adding non-physical causes to the world, then "non-physical cause" must mean something like "a cause that is not traceable to the properties

of the individual objects of fundamental physics".[6] What Kim really means by *Closure*, then, must be something like:

> *Closure**: If a physical event has a cause at *t*, then it has a cause that is traceable to the properties of the individual objects of fundamental physics at *t*.

*Closure** avoids reference to physical properties and so sidesteps any debate over the meaning of "physical". Furthermore, this principle can do the work required of a closure principle in the Causal Exclusion Argument. The problem with *Closure** is that it is essentially a denial of the possibility that complex entities can have novel causal powers. Thus *Closure** is incompatible with the emergentist's claim of *Causal Efficacy*, and so begs the question against emergentism. Kim acknowledges this point himself, stating that "Most emergentists will not have a problem with the failure of the [sic] physical causal closure... For many emergentists that precisely was the intended consequence of their position." (1993a, 209).

So is there any independent reason to believe *Closure**? The denial of *Closure** does not imply that "an ideally complete physical theory will not be able to give an account of all physical phenomena", but, rather, that an ideally complete theory of low-level particles and their pairwise interactions will not be able to account for all physical phenomena. But this claim is precisely the point of emergentism, and so can hardly be counted as an independent reason for accepting *Closure**.

Perhaps the motivation for believing *Closure** is a worry that emergentism is somehow incoherent. If low-level laws already give a complete account of the behaviour of low-

---

[6] I have deliberately used the rather vague locution "traceable to" so as to allow the possibility that views of causation such as those put forward by Woodward (2003), Hitchcock (2007), or Loewer (2007) might count as physical in Kim's sense. These views, and others like them, see causation as a macroscopic phenomenon and deny its existence at the fundamental level. However all of these views trace macroscopic behaviour back to interactions at the fundamental level.

level entities, the thought might go, then adding new high-level laws will either make no difference, or the high-level laws will conflict with low-level laws. But this worry rests upon a mistake. Laws that describe causal powers are what Lewis Creary (1981) calls "laws of influence". Such laws describe the influence the instantiation of one property will have on another but do not purport to describe the ultimate behaviour of a system. The influences must be combined to determine the resulting behaviour.[7] Thus laws of influence cannot come into conflict (contra Cartwright(1983)). This is why the laws of gravitation do not conflict with the laws of electro-magnetism. Adding new laws of influence, emergent or not, just adds new influences that must be combined.[8] There is nothing incoherent about the emergentist's denial of *Closure**.[9]

Ultimately, then, the question of whether or not *Closure** is true is an empirical one. Of course it may turn out that *Closure** is true, in which case the Argument from Closure will be sound as an argument against emergentism. But in this case the argument will be vacuous, since the premise *Closure** is just the denial of *Causal Efficacy (emergence)* and hence of emergentism.

So there is no reason for the emergentist to accept the Argument from Closure. But what of non-emergent non-reductive physicalists? The three claims of emergentism above differ from the three claims of non-reductive physicalism in two ways. First, the terms "high-level" and "physical property" are replaced by "emergent" and "fundamental physical property". If instead we replaced "physical property" with "non-high-level physical property" then there is nothing for the non-reductive physicalist to object to here. But nothing in the argument above turns on these differences. The second point of

---

7 See Corry (2006; 2009) for further discussion.

8 Of course the composition laws may be non-trivial.

9 McLaughlin (1992) has made a similar point.

difference is that the emergentist version of *Causal Efficacy* claims not only that emergent properties have causal efficacy, but that these causal powers are above and beyond those provided by the fundamental physical properties. It is here that there is room for a non-reductive physicalist to distinguish herself from the emergentist.

A non-emergent non-reductive physicalist, therefore, will claim that although high-level properties have causal efficacy, high-level causal powers are nothing over and above the powers of the supervenience base. Thus a non-emergentist non-reductive physicalist will typically accept *Closure**.

For example one common form of non-reductive physicalism claims that mental properties are individuated functionally but are realised in each instance by some arrangement of fundamental physical properties. On this view mental properties have causal powers, but these powers are inherited from the powers of their physical realiser, meaning that *Closure** is not violated. However the mental properties are claimed to be distinct from the realising properties due to the possibility of multiple realisation: the very same mental property may be realised by different physical properties on different occasions.

Since non-emergentist non-reductive physicalism accepts *Supervenience*, *Irreducibility*, *Causal Efficacy*, and *Closure**, it would seem that Kim's argument from Closure should apply. If this is indeed the case, then we can conclude that non-emergentist non-reductive physicalism is incoherent, as Kim claims. In fact, though there is still some wiggle room. Like the Direct Argument, the Argument from Closure relies on *Exclusion*, and as we have seen *Exclusion* is not true in all accounts of causation, so it is open to the non-reductive physicalist to argue for a notion of causation which does not respect *Exclusion*. Alternatively, it might be argued that a mental property and its supervenience base are not independent in some sense relevant to *Exclusion.*

These strategies for defending non-emergentist non-reductive physicalism from the Argument from Closure account for much of the current literature on the topic. But note

that unlike the Direct Argument, the Argument from Closure does not place any restrictions on the analysis of causation, so we cannot argue—as I did against the Direct Argument—that there is no analysis of causation that will do the job. Thus the current defences of non-reductive physicalism tend to rely on particular accounts of causation that Kim can, and does, reject. The defence of emergentism given here, however, is intended to involve no assumptions that Kim himself would not accept.

Conclusion

What I have argued is that there is no notion of causation that will simultaneously satisfy *Exclusion* and support the reasoning of the Direct Argument. Thus only the Argument from Closure can have any traction against the non-reductive physicalist. But I have also shown that the Argument from Closure begs the question against emergentists. Thus emergentism is immune from both versions of the Causal Exclusion Argument. It is only non-emergentist non-reductive physicalism that has anything to fear from the Argument from Closure. It is here, on this remaining ground, that the much-discussed Overdetermination Objection to the Causal Exclusion Argument comes in to play.

The fact that emergentism emerges unscathed from the Causal Exclusion Argument without buying into any particular account of causation is a mark in its favour. As such it would be nice to provide a positive account of the kind of emergentism that is on offer here. A detailed account will have to wait for another occasion, but the considerations above suggest two important features.

First, we have seen that emergentism avoids the Argument from Closure by denying *Closure*.* Furthermore, as we saw above, it is the denial of *Closure** that distinguishes emergentism from other forms of non-reductive physicalism. The denial of *Closure** therefore is an important feature of emergentism and could be seen to provide content to the claim that an emergent property is “more than the sum of the parts”.

Second, consider again the discussion of the Stern-Gerlach experiment. The problem here for the Causal Exclusion Argument arose from the existence of a nomological tie between having a mass of 0.511 MeV and having spin ½. This kind of nomological tie is just the sort of relation that an emergentists might posit between a high level property and its supervenience base. In both cases there is a law of nature which states that two otherwise independent properties are related in such a way that whenever the second is instantiated, the first is too. The only difference is that in the case of mass and spin, both properties are simple, whereas in the case of a high-level property and its supervenience base, the supervenience base is a complex property. I suggest then that we conceive of the tie between an emergent property and its supervenience base as being of the same kind as the ties that exist between properties at the fundamental level.

If we do conceive of emergence in this way, then it is clear that we cannot write-off the Stern-Gerlach counter-example as somehow irrelevant to the discussion of non-reductive physicalism. The nomological ties between emergent properties and their supervenience-base will be of just the right kind to cause trouble for the Causal Exclusion Argument.

This paper has had three main goals: (1) to clarify Kim's Causal Exclusion Argument and show that it comes in two logically distinct versions; (2) to argue that one of these versions—the Direct Argument—fails; and (3) to show that the other version—the Argument from Closure—is ineffective against emergentism and will not tell against other forms of non-reductive physicalism if one does not share Kim's endorsement of a thick notion of causation. In conclusion then, I would suggest that non-reductive physicalists who are attracted to a thick notion of causation should explore the merits of emergentism.

Thanks to Peter Menzies for helpful comments on an earlier version of this paper.

References

Anscombe, G. E. M. (1971). *Causality and Determination: An Inaugural Lecture*. : Cambridge: Cambridge University Press.

Armstrong, D. (1997). *A World of States of Affairs*. Cambridge: Cambridge University Press.

Burge, T. (1993). Mind-Body Causation and Explanatory Practice. In: Heil, J. & Mele, A. (Eds.), *Mental Causation* (97-120). Oxford: Clarendon Press.

Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford: Oxford University Press.

Corry, R. (2006). Causal Realism and the Laws of Nature. *Philosophy of Science*, 73(3), 261-276.

Corry, R. (2009). How is Scientific Analysis Possible?. In: Handfield, T. (Ed.), *Dispositions and Causes* (158-188). Oxford: Oxford University Press.

Crane, T. & Mellor, D. H. (1990). There is No Question of Physicalism. *Mind*, 99(394), pp. 185-206.

Creary, L. (1981). Causal Explanation and the Reality of Natural Component Forces. *Pacific Philosophical Quarterly*, 62(2), 148-157.

Crisp, T. M. & Warfield, T. A. (2001). Kim's Master Argument. *Noûs*, 35(2), 304-316.

Hitchcock, C. (2007). What Russell Got Right. In: Price, H. & Corry, R. (Eds.), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited* (45-65). Oxford: Oxford University Press.

Horgan, T. E. (1997). Kim on mental causation and causal exclusion. *Philosophical Perspectives*, 11, 165-84.

Kim, J. (1989). Mechanism, Purpose, and Explanatory Exclusion. *Philosophical Perspectives*, 3, pp. 77-108.

Kim, J. (1990). Explanatory Exclusion and the Problem of Mental Causation. In: Villanueva, E. (Ed.), *Information, Semantics, and Epistemology* (36-56). Cambridge MA: Basil Blackwell.

Kim, J. (1992). “Downward Causation” in Emergentism and Non-Reductive Physicalism. In: Beckermann, A., Flohr, H. & Kim, J. (Eds.), *Emergence or Reduction?* . Berlin: Walter de Gruyter.

Kim, J. (1993a). The Non-Reductivist's Troubles with Mental Causation. In: heil null, J. & Mele, A. (Eds.), *Mental Causation* (189-210). Oxford: Clarendon Press.

Kim, J. (1993b). *Supervenience and Mind*. Cambridge MA: Cambridge University Press.

Kim, J. (1998). *Mind in a Physical World: An essay on the mind-body problem and mental causation*. Cambridge MA: Bradford Books.

Kim, J. (2003). Blocking causal drain and other maintenance chores with mental causation. *Philosophy and Phenomenological Research*, 67(1), 151-176.

Kim, J. (2005). *Physicalism, or something near enough*. Princeton: Princeton University Press.

Kim, J. (2006a). Being Realistic About Emergence. In: Clayton, P. & Davies, P. (Eds.), *The Re-Emergence of Emergence* . Oxford: Oxford University Press.

Kim, J. (2006b). *Philosophy of Mind*. Cambridge MA: Westview Press.

Kim, J. (2006c). Emergence: Core ideas and issues. *Synthese*, 151(3), 547-559.

List, C. & Menzies, P. (2009). Nonreductive Physicalism and the Limits of the Exclusion Principle. *Journal of Philosophy*, CVI(9), 475-502.

Loewer, B. (2001). Review of J. Kim, Mind in a Physical World. *Journal of Philosophy*, 98(6), 315-324.

Loewer, B. (2007). Counterfactuals and the Second Law. In: Price, H. & Corry, R. (Eds.), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited* (293-326). Oxford: Oxford University Press.

McLaughlin, B. P. (1992). The Rise and Fall of British Emergentism. In: (Ed.), *Emergence or Reduction?: Prospects for Nonreductive Physicalism* . : De Gruyter.

Nagel, E. (1949). The meaning of reduction in the natural sciences. In: Stauffer, R. (Ed.), *Science and civilization* (99-135). Madison: University of Wisconsin Press.

O'Connor, T. & Wong, H. Y. (2005). The Metaphysics of Emergence. *Noûs*, 39(4), 658-678.

Sarkar, S. (1992). Models of reduction and categories of reductionism. *Synthese*, 91(3), 167–194.

Wimsatt, W. C. (1979). Reductionism and reduction. In: (Ed.), *Current research in philosophy of science* (352–377). East Lansing: Philosophy of Science Association.

Wong, H. Y. (2010). The Secret Lives of Emergents. In: Corradini, A. & O'Connor, T. (Eds.), *Emergence in Science and Philosophy* (7-24). New york: Routledge.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.