

Why the Exclusion Problem Still Seems Intractable for the Counterfactual Compatibilist

Melina Tsapos

Supervisor: Robin Stenwall



LUND
UNIVERSITY

“There are other philosophical puzzles about consciousness, but this seems to me the most immediate. We will be ill placed to understand anything about consciousness if we cannot understand its relation to the physical realm”

~David Papineau, 2002

Contents

Contents	3
Introduction	4
1. The Causal Exclusion Problem	5
2 Implications from Denying Causal Exclusion	6
2.1 Injecting Supervenience	7
2.2 Supervenience and Physicalism	9
3 The Counterfactual Compatibilist Solution	10
3.1 Overdetermination	11
3.2 Vacuity and Falsity	16
3.3 Does p Need m's Help?	18
4 Not Compatible with Physicalism	21
4.1 A Robust Principle of Physical Closure	22
4.2 Why Supervenience Is Not Enough	27
4.3 Bennett's Reply to Objections	34
5 Conclusion	36
References	38

Introduction

This essay is about The Exclusion Problem, and in particular the solution suggested by the Compatibilist, and about why it won't work. This problem, the causal exclusion problem, is different from other problems about mental causation, including the anomalism problem and the extrinsicness problem, in that it does not question whether the mental is inherently unsuited for causing anything because it is not spatially extended or cannot be bridged by strict laws and so on. Rather, the worry is how the mental manages to be causal at all when effects caused by the physical as well as the mental seem to be overdetermined, since they are already fully accounted for by the physical causes. The compatibilist, not willing to reduce the mental, argues that both the physical and the mental can be sufficient causes despite the overdetermination worry; thus the mental and the physical are both genuinely efficacious while also genuinely distinct. Jaegwon Kim has famously argued that commitments to physicalism and its metaphysical constraints leaves mind-body property dualism as the only item that can be negotiated away (Kim, 2005, p. 22). It seems like the compatibilist wants to have her cake and eat it too. In this essay I wish to explore whether the compatibilist can hold a physicalist position while also claiming that the mental is at the same causal level, so to speak, as physical causes. Can she maintain her position that two distinct sufficient causes do not involve overdetermination, or at least not overdetermination that we can't live with¹, and that nonetheless she stays true to her physicalist commitments?

The subset of compatibilist that I will discuss in this essay is the one operating within a counterfactual theory of causation, namely counterfactual compatibilist. In particular I will be examining Karen Bennett's argument in "Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract It" (2003). I will argue that she fails to solve the problem because the assumptions are incompatible. The problem is that to maintain physicalism we cannot

¹ A note on the terminology regarding overdetermination, what Bennett sometimes called "bad overdetermination", and is called just overdetermination by Kim, is that when I will speak about "permissive overdetermination" or "genuine overdetermination" it is all references to overdetermination being unproblematic in the mental/physical case. Some compatibilists argue for "no overdetermination", which again has the same implications as all the terms just mentioned about overdetermination that supposedly does not block mental causation, although it has a slightly different understanding.

weaken the assumption that every physical thing² (often called the physical causal completeness principle) has only a physical cause. I am in agreement with Kim here, and I will argue that if we accept the criteria needed to stay within a physicalist position, well, then really our only option is reductionism (if not identity). And perhaps even more pressing is the issue of grounding the mental, and why supervenience is not enough. Before turning to these arguments some background will be provided. In the next section we will consider the causal exclusion problem and the motivation for why some philosophers are in favour of the causal exclusion principle, a principle which the compatibilist clearly rejects. This will be helpful in the next section when we discuss Bennett's counterfactual compatibilist argument, to properly evaluate the compatibilist position.

1. The Causal Exclusion Problem

The exclusion problem is the conjunction of four individually plausible, but jointly inconsistent principles. The four principles are as follows:

- 1) *Mental Efficacy*: Some physical effects have mental causes.
- 2) *Mental Irreducibility*: Mental causes are distinct from, and not reducible to physical causes.
- 3) *Physical Causal Closure*: Every physical occurrence has a physical cause.³
- 4) *Causal Exclusion*: No effect has more than one sufficient cause at any one time.

Each principle above is individually plausible, but taken together they form an inconsistent set; and different problems arise depending on assumptions about the principles. To solve the causal exclusion problem we can reject one of the claims; and yet each has its own consequences. The question is therefore which we should reject. Let's consider the options. To reject 2) is just what the reductionist argues. They believe the evidence is in favour of 1), 3) and 4); therefore only 2) is negotiable (Kim 2005). The nonreductionist argue that mental

² The overdetermination problem that this essay is about is neutral between events and properties. However, I should make it clear that in my discussion of mental properties and events I refer to both qualia and the conceptual.

³ As part of this principle, I will argue, we need a reading that includes a more robust physicalism, and it requires that physical causality is kept within a complete physical domain. But for now we will follow Bennett's reading that all closure requires is that we can account that for every physical phenomena there is a physical cause.

causation is multiple realizable⁴ and this modal difference just ensures their distinctness from physical properties (Kripke 1980; Pereboom & Kornblith 1991; Yablo 1992). The epiphenomenalist argues that principle 2) is not an option to reject, and neither is 3) or 4), which leaves 1), *Mental Efficacy*, as the only one that is false. The epiphenomenalist endorses the claim that the mental never causes anything at all. Rejecting the third principle, *Physical closure*, is to deny the completeness of physics. Endorsing that physics is causally incomplete is to say that the full explanation of some physical occurrence has to reach beyond the physical realm. This is what the substance dualist embraces and this option is not available for the physicalist (Chalmers 1996; Papineau 2001). Finally, if we want to embrace 1) *Mental Efficacy*, 2) *Mental Irreducibility* and 3) *Physical Causal Closure*, we may reject the fourth principle, *Causal Exclusion*. We would claim that effects, including behavioural effects have more than one single sufficient cause. I will argue that this leads to the weakening of principle 3), but principle 3) is what makes the exclusion problem a problem for the compatibilist in the first place. Nevertheless, this is just what the compatibilists argue (Bennett 2003; Yablo 1992; Sider 2003), and so it is this latter position that will be of concern for us for the remainder of this essay. To properly evaluate the position it will be helpful to take a closer look at and provide reasons for the causal exclusion principle, why some deny it while others don't, and the consequences of both.

2 Implications from Denying Causal Exclusion

The mind-body problem, or “*Weltknoten*” as Schopenhauer famously called it, is just that, a “world-knot” regarding the relationship between mind and matter; the problem of finding a place for the mind in a world that is fundamentally physical. The task for the non-reductive physicalist is to untangle the world-knot; to find a way to accommodate the mental within a physicalist scheme, while at the same time preserving it as something distinct. After all, the very possibility of human agency, our beliefs, desires, intentions and decisions presupposes the reality of mental causation. Not to mention the influence of the functionalist accounts of mentality and the multiple realizability they entail, which has prompted the view that mental properties are not identical with neural properties, as the dominant view in the philosophy of

⁴ Multiple realizable: mental property M could be possessed in a variety of physical ways.

mind. However, most of these non-reductionist also claim to be physicalist; meaning that they agree that all phenomena, including mental phenomena, depend on and obtain solely in virtue of physical phenomena. Naturally, this view may initially seem incoherent. If mental properties are not physical, then mental phenomena, instantiating mental properties, must be something more or at least something other than physical phenomena, and the nonreductionist would have to explain just how the mental depends on the physical. By consistently denying that mental properties are physical, it is only reasonable to expect an explanation for how she remains true to her physicalist commitments.

2.1 Injecting Supervenience

Many philosophers turn to the notion of supervenience to capture the autonomy of the mental while acknowledging the primacy of the physical (both properties and events). Donald Davidson in *Mental Events* (1970) made use of the supervenience principle. First introduced as the philosophical technical term by the British emergentist in the early 20th century, supervenience is taken as a relation between two sets of properties. The supervenience is looked to for a formulation of physicalism that is free of the reductionist commitments. And this was precisely the line of thinking that appears to have motivated Davidson to inject the supervenience principle (in *Mental Events*) into the discussion of the mind-body problem.

There are numerous ways to formulate the supervenience of mental properties in the literature, and there is an abundance of descriptions of different brands of the principle. For now, it is enough to know the supervenience principle often called “strong supervenience”, which will be argued to be a minimal requirement for physicalism (Kim, 1993, 2005). The thesis of the principle can be stated as:

Mind-body supervenience Mental properties supervene on physical properties in the sense that if something instantiates any mental property M at t , there is a physical base property P such that the thing has P at t , and necessarily anything with P at a time has M at that time.

For example, if a person has the mental property of pain, it must be the case that that person has some physical property, perhaps a complex neural property, such that necessarily

whenever anyone has this physical property, they have pain. The principle brings mental phenomena within the scope of the physical, so that the physical determines the mental and if supervenience fails, that is, if the mental domain were unanchored in the physical domain, then mental causation to the physical would obviously breach the causal sufficiency of the physical. Without invoking it we say nothing more about the relationship between the mental and physical than the claim that all objects with a color have a shape says about the relationship between colors and shapes. The mind-body supervenience is an attractive option for many philosophers because it seems to provide a way of protecting the autonomy of the mental without lapsing into substance dualism all over again. The supervenience principle captures a commitment common to all positions on the nature of the mentality that are physicalists. And it should be a shared minimum commitment for all physicalists.

In the literature there are many arguments both in favour and in objection to the causal exclusion principle. For the most part it has not been warmly received in the philosophy of mind, even though Kim's arguments for it have been extremely influential. Many find it highly counterintuitive. Kim, on the other hand, has claimed that it is *prima facie* obvious and "virtually an analytic truth" (2005, p. 51). His (2005) reasoning is that from principle 2) *Irreducibility*, we have $M \neq P$ and because of the causal exclusion principle we must eliminate either M or P as P*'s causes.⁵ By causal closure, principle 4) (also called the *Physical Completeness* principle), we have to choose P over M for being the cause of P*. Note, however, that the exclusion itself is neutral with respect to the mental-physical competition; it does not favour one over the other, but merely states that not both can stay. In fact, principle 3) and 4) together is compatible with the epiphenomenalist conclusion. In any case, following the reasoning here, we must agree that M is the one that must be excluded and P retained if we believe in a physically complete world. Besides that Kim sees it as an obvious principle, there is motivation for it by appeal to a general scientific value which states that we should get by with as little or as few entities as possible, and excluding additional causes if we already have a sufficient cause. Then there is the argument that overdetermination is a rare coincidence when considering causation, and to postulate that there are wide spread coincidences is just too implausible to even consider. Perhaps this is not

⁵ Let's assume here that we are not dealing with a case of causal overdetermination. We will return to the issue of overdetermination.

problematic enough for the compatibilist who may argue that it is not a matter of coincidences at all. Therefore, the compatibilist solution, which either find overdetermination unproblematic or simply that there is no overdetermination, must presuppose supervenience and a weaker principle 3) than I will argue that physicalism requires. So why is supervenience a minimal commitment for physicalism? Let's take a closer look at how the principle plays a part in our problem.

2.2 Supervenience and Physicalism

Physicalism is a general claim about the nature of the world; it is the metaphysical thesis that everything is physical and there is nothing over and above it, hence Armstrong writes: "What supervenes is no addition of being [...] The supervenient is ontologically nothing more than its base" (Armstrong, 1997, p. 12-13). For everything that is real, even things that are complex properties, are such that those properties must supervene or be entirely entailed by physical processes. Physicalism means that everything that exists is due solely to the interactions of matter and energy. In order to capture the core commitment of physicalism, David Lewis in his (1983) has the intuition that a minimum understanding should be that no two pictures can be identical in the arrangement of dots and yet differ in their global properties:

"A dot-matrix picture has global properties - it is symmetrical, it is cluttered, and whatnot - and yet all there is to the picture is dots and non-dots at each point of the matrix. The global properties are nothing but patterns in the dots. They supervene: no two pictures could differ in their global properties without differing, somewhere, in whether there is or isn't a dot. [...] The idea is simple and easy: we have supervenience when there could be no difference of one sort without differences of another sort" (Lewis 1983, p. 18)

If we were to ask about what a phenomenon or concept is, such as a lightning strike or feeling hunger, then there are different levels of explanations; the physical and the phenomenal. A lightning strike is a flash of light that occurs between a cloud and the earth, or we could say that it is an electric discharge between the atmosphere and an object. The first is a higher level explanation and the latter a lower level explanation. Hunger could be explained at a higher level as the sensation of a desire to eat, to feeling hungry. Or, a lower level explanation would be to say that hunger is the complex process that involves the

gastrointestinal tract and hormones, such as ghrelin that sends signals to the brain through vagal nerve fibers. If we could recreate an exact physical duplicate of me and feed all the exact same signals from my gastrointestinal tracts to my nervous system, I would experience the exact same thing as the original me did, feeling hungry. Every such phenomenon is said to supervene on the material properties that compose it. Which brings us back to supervenience, and the fact that the principle itself is not an explanatory theory; as Kim rightly highlights “it merely states a pattern of property covariation between the mental and the physical and points to the existence of a dependency relation between the two.” (Kim, 1998, p 14) But not why there is such a relation.

The fact that the supervenience can be shared by many diverse positions on the mind-body problem, from reductive physicalism to dualist emergentism, shows that it is not a metaphysical relation and we must look elsewhere for metaphysical grounding. An account of the mind-body relation incorporating supervenience must specify the dependence relation between the properties that grounds supervenience. And since both the reductionist and the non-reductionist claim to be physicalist, it should be the physicalist intuition that guides the relation. This means that if physicalism is true at our world, then no other world can be physically identical to it without being identical to it in all respects. With these principles in place and the implications of overdetermination and the exclusion problem defined, we shall consider the solution Bennett has to offer for our problem.

3 The Counterfactual Compatibilist Solution

Compatibilism, or rather *counterfactual* compatibilism argued by Karen Bennett (2003) is a position which includes the truth of *Mental Efficacy*, *Mental Irreducibility* and *Physical Causal Closure*. The fourth principle, that no effect has more than one sufficient cause, is denied, and according to it the truth of the former three principles does not result in improbable overdetermination. Thus, Bennett argues that if the truth of the three principles does not result in the bad kind of overdetermination, the proponents of the exclusion problem would no longer reach the conclusion that either of them must be given up. And as such we will be able to have both sufficiently efficacious mental properties and a physically complete

domain. According to Bennett, counterfactual compatibilism is a physicalist position by the fact that mental properties supervene on physical properties. They supervene in the sense that whenever a physical subvenient property occurs, so does the mental property that supervenes on it. For example, if the physical property of being c-fibers firing is a subvenient of the mental property of being in pain, then, whenever the former occurs, so does the latter. The kind of necessity that, on compatibilism, is involved in supervenience is *metaphysical* (Bennett, 2008, p.5). So it is no surprise that a necessary metaphysical relationship must exist between the mental and physical properties and is at the heart of the compatibilist thesis. In order for Bennett's counterfactual model of causation to get off the ground she will need to contest overdetermination as traditionally understood, so let's take a look at overdetermination before we discuss her solution further.

3.1 Overdetermination

In cases of overdetermination more than a single sufficient cause bring about the same effect at the same time. Many believe that either event is a sufficient cause of the effect, the other event(s) is(are) unnecessary as a cause for the effect, or else it would be an overdetermined effect. To use the textbook example, two gunmen shoot Mr. X at the same time, so that the death is overdetermined by bullet *a* firing and bullet *b* firing. The following counterfactual test establishes this fact:

(Test1) Had bullet *a* fired without bullet *b* firing, the death *c* would have occurred: $(a \ \& \ \sim b) \ \square \rightarrow c$.

(Test2) Had bullet *b* fired without bullet *a* firing, the death *c* would have occurred: $(b \ \& \ \sim a) \ \square \rightarrow c$.

In this case the test would run as follows: if both of these counterfactuals are true, Mr. X's death is overdetermined by two causes that are individually sufficient as a cause of Mr. X's death. It also indicates that each bullet firing is unnecessary to bring about *e*. Bullet *a* firing would suffice to cause the death, since we can remove the other bullet firing, leaving only bullet *a* firing, and the death still occurs. Similarly, bullet *b* firing suffices to cause Mr. X's death because we can remove bullet *a* firing, leaving only bullet *b* firing, and the death still occurs. And to the latter point, bullet *a* firing is not needed to cause Mr. X's death because we can get rid of bullet *a* firing and the death still occurs. Neither is bullet *b* firing needed as

a cause of Mr. X's death because likewise, we can remove the bullet *b* firing and the death occurs. Traditionally it has been argued that his death is thus *overly* caused. If either of the gunmen were to fire without the other, he would still die.

The test from the counterfactual account for overdetermination may likewise be applied to the mental and physical by substituting bullet *a* for *m*(M), mental events or properties, and bullet *b* for *p*(P), physical events or properties, and the death or the caused event, for the effect *e*. And as is the case with the gunmen with regard to their sufficiency, so is the case with the mental and physical. Following Bennett's necessity claim, it takes the following form:

T1: Had *m* occurred without *p*, *e* would still have occurred ($m \ \& \ \sim p$) $\square \rightarrow e$.

T2: Had *p* occurred without *m*, *e* would still have occurred: ($p \ \& \ \sim m$) $\square \rightarrow e$.

The idea behind the test is to determine cases that are overdetermined. So, if either of the two causes *m* and *p* are both capable of causing their effect *e* without the other, then, the scenario containing **T1** and **T2** passes the test and comes out as true. If the scenario does not pass the test it would not pass as a case of genuine overdetermination. In other words, if both of these counterfactuals are true, then the effect *e* is overdetermined, since the individual sufficiency of each cause renders the other cause individually unnecessary. Many argue that overdetermination is a rare coincidence. After all, bushfires infrequently ignite by the simultaneous occurrence of a dropped cigarette bud and a lightning strike by pyrocumulonimbus clouds (Moore 2017, Engelhardt 2015, Kim 1998, 2005, and Roche 2014) To suggest that overdetermination is not problematic therefore seems at first sight to have the odd consequence that I would still have gone to the fridge to eat even if I hadn't been hungry because my cortical neurons would still have been firing; and that I would still have gone to the fridge even if my cortex hadn't been firing because I would still have been hungry. And mental causation is ubiquitous, so overdetermination would then too, if true of mental causation, be ubiquitous and these massive amounts of coincidence would be unacceptable. The overdetermination in the case of the mental would be extremely widespread, happening each and every time we move our bodies. Genuine cases of overdetermined effects are, they say, very rare. Naturally, then suggesting that overdetermination is unproblematic seems the wrong model for mental causation. After all,

overdetermination implies that even if one cause had been absent, the result would still have occurred because of the other cause. It just seems wrong to say that I would still have walked to the fridge even if I hadn't felt hungry (because my neurons were firing), or that I would still have gone to the fridge even if my neurons hadn't been firing (because I felt hungry). To postulate that there are systematic coincidences is also just not something we should consider for a serious account of mental causation. The counterfactual compatibilist agrees with this type of criticism. Rather, Bennett argues that it is not a matter of coincidences:

“But why should the sheer extent of the overdetermination make it any less troublesome? The only answer I can see is that its pervasiveness would give us a reason to think that it is not a coincidence. [...] The difference, the compatibilist will say, is that there is an important tight relation between the mental and the physical.” (Bennett, 2003, p. 475)

To defend physicalism a common strategy is to show how mentality depends in some intimate way on physical features. The physicalist position can be characterized as Van Gulick (1992) put is: “[i]n every instance in which a property applies to the world of space and time it does so in virtue of physical properties that apply to the world of space and time” (ibid. p. 164). This is much the same as Bennetts characterization:

“The test opens the door to the idea that a *tighter* connection between the two causes *would* help defuse the threat of overdetermination – and would do so in a slightly different way than we have yet seen. If one of the causes *guarantees* the existence of the other, there is no issue about skipping over some worlds to get to one where the antecedent of the relevant overdetermination counterfactual holds. There are no further worlds to skip to. To put the point more formally: if one of the causes necessitates the other, if it is at least metaphysically impossible for the one to occur without the other, then one of the overdetermination counterfactuals will come out vacuous. And there is something to be said for the idea that the vacuity of one of them means the effect is not overdetermined. [...] For one thing, the idea that it is metaphysically necessary that one of the causes occurs whenever the other does gives some content to the often-heard idea that despite not being identical, the mental and the physical causes are not exactly *distinct*, either. And it also means that there is a sense in which one of the overdetermination counterfactuals is not quite up for discussion –you cannot quite ask what would happen if one occurred without the other if it just can't occur without the other” (Bennett, 2003, p. 479-480).

The compatibilist insists that overdetermination is not so bad, because she can point out a difference in an important tight relationship between the mental and the physical, which does not hold between the two shootings in the traditional example of the firing squad. Theodor Sider likewise objects to the coincidence claim and argues as follows:

“Imagine a paranoid who thinks that every time someone is shot, there are two causally independent shooters. He is crazy, but why? One reason [...] is that it would be a coincidence that all these sharpshooters just happen to fire at the same places at the same times. This great regularity would need an explanation, and none could be given. [...] But this is all wrong: it is no coincidence that baseballs and their parts, or mental and physical events, are correlated, given the necessary truths governing these correlations. It is necessary that appropriately arranged atoms compose a baseball, and that physical properties instantiated in appropriate circumstances result in the instantiation of an appropriate supervenient mental property.” (Sider, 2003, pp. 722-723)

Both Bennett and Sider are making a claim here about the *necessary relationship* that holds between the physical and the mental. Sider uses the analogy of the necessary arrangement of the atoms that compose a baseball and its parts to argue that both are sufficient and independent clauses to have the effect *e*, say the window shattering. In the next section we will be discussing the issue of the necessary relationship deeper, so we will return to these arguments. Similarly, Michael Roche (2014) argues that the exclusion argument can only succeed if, as Kim claims, the case is not about genuine causal overdetermination. Kim has consistently argued that the exclusion principle, somewhat different than principle 4), is this:

Kim's Exclusion Principle: No effect has more than one sufficient cause at any given time, “*unless it is a genuine case of causal overdetermination.*” (Kim, 2005, p. 42 added emphasis)

Roche's overdetermination challenge is that one can resist Kim's conclusion by denying the principle claim above, by the following:

No overdetermination: E is not causally overdetermined by M and P at *t*.

One may maintain that the effects of the mental are always genuinely causally overdetermined.

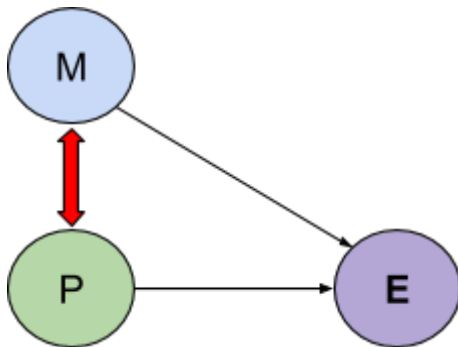


Figure 1. E has two sufficient causes, M and P, at a given time, t.

In the situation depicted in figure 1 above some effect, *E*, seemingly has two distinct sufficient causes at a given time. *Kim's Exclusion principle* permits an event to have more than one sufficient cause at a given time only if that event is genuinely causally overdetermined, in other words that *E* would have happened if *M* didn't and vice versa if *P* didn't happen. Accordingly, if *M* and *P* do not overdetermine *E* at *t*, one of them must be excluded or reduced to the other. But the non-reductionist has the option to simply reject *No Overdetermination* and maintain that it is a case of genuine overdetermination; and live with ubiquitous overdetermination, at least in the domain of the mental. Now, one might object and wish to defend *No Overdetermination* by pointing to the fact that two or more events or properties can only casually overdetermine an effect if the causes are independent (not just distinct) of one another. But even though as per physicalism, *M* supervenes on, and depends on *P*, the problem is that they would lack ontological distinctness. I believe that there is a core difference between reductive and non-reductive physicalism and by the principle of charity we should attempt to avoid drawing the distinction in a way that implies that either side of the debate has committed an obvious error, is radically mistaken about the nature of its own position, or is defending an obviously inconsistent view. It isn't simply so that the non-reductive physicalist sometimes just lost track of their basic commitments, or that the reductionist is just too stubborn to see that many physical phenomena cannot be casually explained by fundamental physics. However, it is a question of where our commitments are. So, again the basic argument is whether there are two causal and distinct properties in our physical world. The question about the nature and existence of properties are nearly as old as

philosophy itself, thus naturally we cannot devote as much of this essay as would be desirable to discuss the many issues tied to this problem. The gap between the non-reductive physicalist and the reductive physicalist is widened by what captures physicalism. And this will prove to be a crucial difference for Bennett's account.

But as far as overdetermination is considered, the compatibilist accepts that the effects of mental causes are always overdetermined, "just not in a bad way" and she claims that the notion is, or at least should be completely unproblematic, as long as she can break the analogy between the standard textbook example of overdetermination (the firing squad, or so-called bad overdetermination), with the mental/physical case. And as long as she can successfully do that, it matters not if we think of overdetermination as bad or simply denying that there is any overdetermination. It is merely a terminological issue.

3.2 Vacuity and Falsity

To establish that there are *permissible* overdetermination cases and so escape the exclusion problem, the compatibilist has to show that at least one of the counterfactuals, **T1** and **T2**, is either false or vacuous. What she really needs is to successfully deny that both are non vacuously true. Bennett must establish a necessary condition on overdetermination if she wants to show that the distinct mental and physical causes may themselves be strongly counterfactually dependent. So she might solve the exclusion problem by showing that the tests for overdetermination comes out as either false or vacuous; and if so then she will have established that cases of mental causation are not cases of genuine overdetermination (such as *Kim's Exclusion principle*); and hence not on par with cases such as the firing-squad example. The puzzle of course is to do so while maintaining the causal sufficiency of both mental and physical properties and maintaining that even though those mental and physical properties are not identical their coexistence does not violate physical causal closure. As we have previously discussed, if physical subvenient properties by themselves are capable of doing all the causal work of their supervenient mental property, then the mental properties do not seem to be needed and therefore seem to be dispensable. And if she claims that *e* would not happen if *p* occurred without *m*, and *p* does not need *m*'s help, the question remains why such causes should always be so counterfactually dependent, if they are ontologically distinct.

First, Bennett needs to rule out that it is not **T1** that is false or vacuous. The main reason for most compatibilists to be compatibilist in the first place is that they don't want to identify mental properties with physical ones, because *M* is multiply realizable and could be possessed in a variety of physical ways. Compatibilists have previously argued for the truth of **T1** (Yablo 1992; Lepore and Loewer 1987). The argument has been that if an event *m* had occurred but event *p* not occurred there would be some other physical event, say *p** that would have occurred and it would have been sufficient as the cause of *e* (see figure 2).

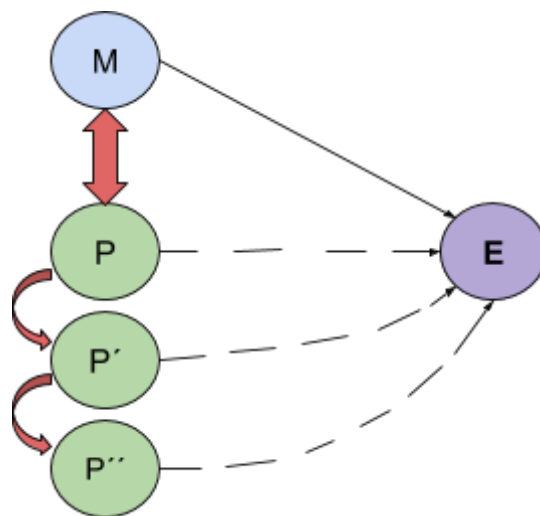


Figure 2
The wrong vacuity strategy—*M* as multiply realizable and modally different than *P*, means *M* could be possessed in a variety of physical ways, and compatible with epiphenomenalism.

The same goes for the property case. The problem with this argument is that it is compatible with the mental being epiphenomenal. Bennett better have another reason. How is it possible that *m*'s (and *M*'s) occur without *p* (and *P*)? The answer for Bennett lies in the assumption that physicalism is contingent. She argues:

“The claim that it is impossible for property *M* to occur without any relevantly *P*-like property basically amounts to the claim that physicalism is necessarily true. Yet most people think it is only contingent [...] Though there may not be any souls here, there are worlds in which there are, and in those worlds things can have *M* without having any physical properties at all. *So property M can indeed be instantiated without any physical property.*” (Bennett 2003:483-484, emphasis added)

But at our world *M* is always instantiated together with *P*. So, what would be the necessary relationship between the two? Bennett doesn't say. I will, however, argue that physicalism

requires necessitation which Bennett has not satisfied. But first, let's see Bennett's reason why p does not need m 's help to be causal.

3.3 Does p Need m 's Help?

Typically the non-reductionists justify their commitment to physicalism with the claims that (i) mental properties are physically realized, and (ii) that the mental supervenes on the physical. These are both claims that are essential to physicalism, and both are claims that Bennett utilises in order to remain a physicalist. Since I want to argue that Bennett cannot have her cake and eat it too, meaning she cannot hold on to the idea that mental properties are irreducible while also claim to be committed to physicalism, I have to show how (i) and (ii) are not coherent with the claim that mental properties are not physical properties.

First, consider that both counterfactuals are nonvacuously true: “[...] I believe she *can* deny that both of them are nonvacuously true” (Bennett, 2003, p. 480). The reason is that she wants to make sure that the bond between m and p is essential. If m and p are so tightly connected, which it needs to be in order for her to avoid *bad* overdetermination, then first we need to ask the point of the counterfactual to being with. Why ask what would happen if one occurs without the other, if “*it just can't occur without the other?*” In a sense, the one using the test must agree that there is some fundamental difference between m and p , such that we could at least imagine one occurring without the other. After all, the distinctness claim is a crucial part of the compatibilist position, as Bennett writes: “without it, she would not face the exclusion problem, and would not need to be a compatibilist in the first place” (ibid. p. 486). And her claim is that physicalism is contingent, but true at our world would answer this concern, but note that only if physicalism is contingent.

In essence what I am suggesting here is that the proper moral to be drawn for Bennett's account might just not be that two sufficient and distinct causes yield no bad overdetermination, but instead that the non-vacuous truth of the counterfactuals is not necessary for overdetermination. Nevertheless, she digs her heels in and insists that her conclusion and her reasons for thinking that the tight relation between the mental and the physical makes a difference to the kind of overdetermination involved in firing squad cases.

Naturally the followup question is this: if her point is that the properties are extremely tightly connected, indeed they are almost inseparable, is she still faithful to the compatibilist/property dualist intuition and protective of principle 2), the mental and physical being distinct and irreducible? In what, then, lies their distinctness? We will be concerned with this question in 4.

Returning to the test, this leaves **T2** as the counterfactual which the compatibilist wants to deem vacuous and/or false. By suggesting that if p happened without m the effect e would not have occurred, she is arguing that the physical would not cause whatever it would without the mental. As I, and many others, argue, the physical cannot *need* m 's help if it is causally sufficient for e , which is stated in principle 4). This would violate the completeness of the physical domain. But how principle 3) is stated so far follows Bennett's reading. I want to argue that principle 3) in particular, taken conjointly with 4), should satisfy a much more robust principle. However, this kind of robust principle, which I will defend in the next section, blocks the compatibilist and Bennett's account trivially, and renders it a non starter for a physicalist.

Continuing still a little while longer with our reading of the physical closure principle, Bennett is confident that her way, as an alternative understanding, will not give rise to the idea that p needs m 's help. Therefore it would not be a threat to p 's causal sufficiency. To remind us of the solution for this problem, in Bennett's own words are that "the conditions that must hold for p to bring about e —physical conditions, note—are *basically the same as the conditions in which p necessitates m* . So if p were to occur without m , those conditions would

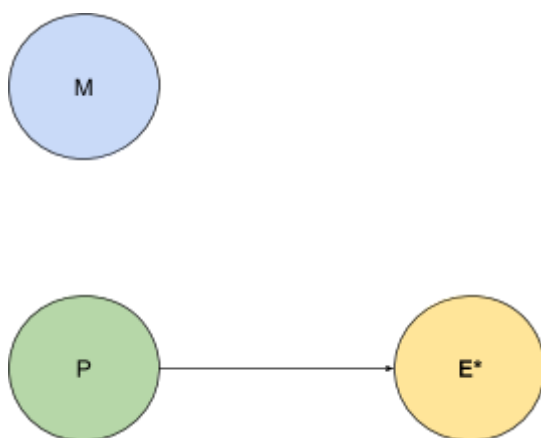


Figure 3

The Falsity strategy— P without M would not cause E but some other effect, perhaps E^*

not hold—and p would not, or at least might not, cause e . And that does not mean that p does not *actually* cause e .” (ibid., p. 488-489)

Bennett believes that the worlds where P would occur without M are different enough from the actual world that there is no reason to think that e would occur there at all, but rather E^* or some other effect, as long as it is not E . And that is why for Bennett the counterfactual **T2** is false since the physical cause can indeed happen without the mental one but it would be untrue that it caused E . The example Bennett uses is this: While it is true that the pattern of her neural activity could occur in a petri dish, it is also true that if it did it would not be a world in which we could expect it to cause whatever it may cause now that it occurred in her brain, such as her raising her hand. But, as she argues, this is not the case; so the physical property or event that does not necessitate M , is not causally sufficient for E . And, according to Bennett, it does not undermine P 's putative causal sufficiency for E either (ibid., p. 481).

However, a mere counterfactual connection between the causes is not enough to defuse overdetermination. Bennett suggests that “a *tighter* connection between the two causes *would* help defuse the threat of overdetermination, [namely that] if one of the causes *guarantees* the existence of the other, there is no issue about skipping some worlds to get to the one where the antecedent of the relevant overdetermination counterfactual holds” (ibid., p. 479). But why would the fact that such a relation holds between m and p defuse the difficulties of overdetermination? They differ in the respect that mental and physical properties are in a relationship in the way that the two gunmen from the classical example are not. Mental and physical properties are *dependent* in the sense that they, due to supervenience, are co-occurring. Recall supervenience tells us that whenever a subvenient physical property occurs so does a supervenient mental property. Bennett informs us that “the physical cause constitutes, realizes, or determines the mental causes; perhaps the mental cause simply supervenes with metaphysical necessity on the physical one” (Bennett, 2007, pp. 327-328). This cannot be said of the two gunmen in the firing-squad example. Each gunman could easily occur without the other since there is no tight relation holding between them. She continues: “[...] if it is at least metaphysically impossible for one to occur without the other, then one of the overdetermination counterfactuals will come out vacuous.” (Bennett, 2003, p. 479). And as we already rehearsed, if the counterfactual test is vacuous or false, it means that

the effect is not overdetermined. It could then be said that genuine overdetermination involves independent causes. And this genuine overdetermination, Bennett argues, is in no way conflicting with compatibilism since “the compatibilist could in principle accept that the effects of mental causes are always overdetermined, just not in a bad way—the overdetermination is perfectly acceptable, unsurprising, and unproblematic” (Bennett, 2003, p. 474).

So, permissible overdetermination does indeed seem compatible with sufficient mental and sufficient physical causes while taken together (yet distinct!) would cause the same effects. Now, what Bennett wants to have shown is this: with a tight relationship between the two different properties, we no longer need to be concerned about this causing overdetermination in the traditional sense, and this relationship ensures overdetermination that is not problematic. But is counterfactual compatibilism compatible with physicalism, if we have a robust understanding of principle 3)? And is the supervenience principle enough to secure the physicalist commitments for the counterfactual compatibilist? We shall now turn to the reasons why I believe the exclusion problem still seems intractable for the counterfactual compatibilist.

4 Not Compatible with Physicalism

For Bennett’s account of counterfactual compatibilism to remain a physicalist account, she has had to maintain two essential principles. Firstly, that mental properties supervene on physical properties and that this is enough to ground the mental in the physical; and second, to secure the third principle (*physical closure*), which says that the physical effects are accounted for by physical causes, or that the physical is complete and does not need anything “extra” to account for causation. I will argue that this is where Bennett starts to lose points as a physicalist. Or, at least where her view diverges from being the rigorous account she promised it would deliver. So let’s start with a closer look at the latter principle regarding physical closure and work our way to the first, the supervenience principle, and see how these create trouble for the counterfactual compatibilist and the nonreductive physicalist in general.

4.1 A Robust Principle of Physical Closure

In this section I hope to show why we need a more robust principle of physical causal closure; and that Bennett's reading and the conclusions she derives from such a reading leads to a weakening of the principle that was supposed to satisfy the physical intuition. The intuition that physical effects are exhausted by physical causes, or at least ontological physical properties. Dwayne Moore has argued along these lines in "Causal Exclusion and Physical Causal Completeness" (2019). He says that "either neomorous nonreductive physicalist solutions fail on account of the fact that they do not satisfy a robustly defined principle of physical causal completeness, or there is an accelerating trend of solving the causal exclusion problem by suitably relaxing the principle of physical causal completeness" (Moore, 2019, p. 479). What does a robust account of principle 3) look like, and why isn't the version Bennett and other compatibilists have defended enough?

Physical causal closure says that every physical occurrence has a physical cause. Bennett calls it *Completeness*, which essentially says the exact same thing: "Every physical occurrence has a sufficient physical cause" (Bennett, 2007, p. 325). To reject this principle is to claim that physics is causally incomplete, and we need to reach beyond the physical in order to give a full explanation of a physical occurrence. This is just not a direction a physicalist can take. David Papineau has argued that there are empirical reasons to accept it (Papineau 2002, 2001). But how does this principle exclude mental causes if overdetermination is not so bad afterall? The truth is, it doesn't. Without exclusion, and if we understand closure to mean just that it is enough for us to include a $p(P)$ for every $p^*(P^*)$ that occurs in our account, then there seems to be no problem for the compatibilist. So how can this principle be strengthened, and perhaps most importantly, why should it be?

First of, it is worth reflecting on Kim's understanding of the principle. He has often included the view that causal closure also implies a causally closed physical domain. He argues that if we were to choose M over P as P*'s cause, the closure principle would kick in, forcing us to posit a physical cause for P* (Kim, 2005, p. 43). But if we get rid of the *Exclusion* principle, we don't have to choose between M and P, Bennett argues. However, Kim has a further analysis of what the physical completeness (closure) encompasses. He writes:

“It is the causal closure of the physical world that excludes the mental cause, enabling the physical cause to prevail. If the situation with causal closure were the reverse, so that it was the mental domain, not the physical domain, that was casually closed, the mental cause would have prevailed over its physical competitor. I suppose this could happen under some forms of Idealism; one would then worry about the “problem” of physical causation” (Kim, 2005, pp.43-44)

Kim frequently refers to principle three as including this strict understanding of closure, the closure of the physical domain, that physical causality is kept within a complete physical domain. And this of course excludes Bennett’s solution from the get-go. If, in order to be a physicalist we must accept that causality is kept exclusively in a closed physical domain, without any possibility of any “extra” influences, compatibilism would obviously disqualify as a physicalist account per definition. So, what reasons does Kim have to assume such a strong understanding of the principle? According to himself, not many. As a more direct way to rule out overdetermination, he suggests that we might want to consider adopting “a stronger form of physical causal closure” (ibid. p. 50). It would ensure that no nonphysical event can be a cause of a physical event. This strong closure would not only stop overdetermination in its tracks, but it would also allow us to dispense with principle 4). We don't need the principle since, the stronger closure in conjunction with *Mental Irreducibility* makes M ineligible as a cause of P*. However, Kim still feels we have reasons to not trade 3) and 4) for the stronger new principle just suggested. The reason is that he believes there is a philosophical gain in staying within the weaker closure premise. Starting our argument with mind-body causation already ruled out is apt to provoke the complaint that the argument begs the question, he says (ibid. 51). And that would just be a mistake and would not really gain any real progress for anybody.

So, let’s consider a related condition for establishing physical causal completeness, suggested by Moore (2019). Moore argues that principle 3) has numerous important nuances, but we will mainly focus on one of the conditions which he argues are necessary for establishing physical causal completeness. He calls it the “Absolutely Sufficient Physical Cause Condition” and it states:

Sufficient Condition: “All behavioural effects have some absolutely sufficient physical cause, where an absolutely sufficient physical cause is a minimal set of

individually necessary causes that are jointly sufficient for the behavioural effect, and the minimal set of individually necessary causes is entirely composed of physical causes” (Moore, 2019, p. 482)

There are two main motivations for *Sufficient Condition*. The first one is pretty intuitive, and something that both Kim (1993) and Moore (2019) have argued. Given that we have a sufficient physical cause p that is complete, p is all the causation we need for p^* . Some p is the minimal set of causes that are individually necessary and jointly sufficient to bring about p^* . Since the sufficient physical cause p is *physical* incontestably, what constitutes p is only physical causes; and the minimal set p *ipso facto* does not contain m as a cause (assuming of course that m is not something physical). There is just no need for us to depart from the physical world to include m as part of the minimal set of causes required to bring about p^* , if the complex physical cause p is sufficient to cause p^* . This pretty much amounts to a tautology and Kim’s concern again, that we exclude the compatibilist solution per definition. And, even though there might be a strong intuition for this understanding I don’t think we can convince the compatibilist to change her mind. However, I do see another possible way that we might want to understand this problem. I will return to it at the end of this section. But first, consider the second motivation which perhaps would be more convincing.

The second motivation is from neuroscience and the likelihood that the science’s advancement in understanding the cause of behavioural effects will most likely be some set of entirely neural processes that will be sufficient and contain no ‘gaps’ that would require supplementation by some non-physical cause. Andrew Melnyk, has in his (2015) “The Scientific Evidence For Materialism About Pains” argued rather forcefully along these lines, and found that “[t]he empirical supervenience of pain on the neural is shown [...] to favor the hypothesis that pains are, in a sense that is made precise, purely material” (ibid., p. 1).

The benefit of compatibilism and Bennett’s account is of course that it appears to ensure that p^* has a sufficient physical cause p (securing *Physical Causal Closure*) while it also makes it impossible to exclude the distinct mental cause m (securing *Mental*

Irreducibility) from causing p^* (securing *Mental efficacy*). But if we agree that *Sufficient Condition* is a necessary condition for robust physical closure, then compatibilism, by violating *Sufficient condition*, weakens principle 3) and as such it might not cut it as a robust enough account for the physicalist.

The argument as discussed is a quite straightforward way of dismissing Bennett's and the compatibilist conclusion: since were m a metaphysically necessary cause for p^* , then m is an individually necessary cause of p^* ; and so it must be included in the minimal set of individually necessary causes that are jointly sufficient for p^* . Now, since m is included within this set, it is not the case that the sufficient cause is entirely physical. And p^* does not have an *absolutely* sufficient physical cause; thus violating *Sufficient Condition*.

Note Papineau (2009)'s articulation of the requirement on sufficiency:

“Now consider the requirement that the physical cause be ‘sufficient’. This is needed to ensure that it causes the physical effect by itself, and not solely in virtue of its conjunction with some sui generis non-physical cause. Imagine, for example, that some neuronal activity is caused by the conjunction of some chemical state and some sui generis mental cause. Then that neuronal activity would have a physical cause (the chemical state), but this cause would not have sufficed on its own, in the absence of the sui generis mental factor. This is clearly less than we need for a philosophically significant closure thesis. To make sure we have the right kind of closure thesis, *we thus need to require that every physical effect have a physical cause that suffices on its own*” (Papineau, 2009, p. 59, emphasis added)

According to the principle of *Physical Causal Closure*, every physical occurrence has a sufficient physical cause, however, according to Papineau it must be sufficient *on its own*. As such, p is an individually sufficient physical cause of p^* . So, either it is confusing to have a test which separates the properties the way counterfactual compatibilism does, and the compatibilist has forgotten about principle 1) and they are not fully distinct properties, events or entities; or they might just violate the conditions for physicalism. Thus:

- A. If physicalism is true, then the mental obtain solely in virtue of physical phenomena.
- B. If **T2** is false without the mental being reducible to the physical, then the mental does not obtain solely in virtue of physical phenomena.

- C. The counterfactual compatibilist maintains that **T2** is false (or vacuous) and that M is an irreducible cause.
- D. Therefore, counterfactual compatibilism is not compatible with physicalism.

But even if we may have strong intuitive reasons to have a more robust reading of the Closure principle, which would exclude Bennett's account from being compatible with physicalism, as shown by A-D, we still haven't gained much more ground than Kim in his attempt to strengthen it. Bennett (2008) comments on a similar objection in a footnote as follows:

"[...] 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)" (Bennett, 2008, p. 1).

The completeness of physics doesn't itself say anything about non-physical things. It is purely a doctrine about the structure of the physical realm. It says that, if you start with some physical effect, then you will never have to leave the realm of the physical to find a fully sufficient cause for that effect. It seems that if we want to get from the completeness of physics itself to the physicalist conclusion that everything is physical, we need an argument. David Papineau provides what I think is an argument very similar to the one I try to make above, and it is this: "if the completeness of physics is right, and all physical effects are due to physical causes, then anything that has a physical effect must itself be physical. Or, to put it the other way round, if the completeness of physics is right, then there is no room left for anything non-physical to make a difference to physical effects, so anything that does make such a difference must itself be physical" (Papineau, 2000, p 5).

However, let's just assume that we just have to accept that we cannot satisfy the compatibilist with this type of argument. I still think we can learn a few things that are important from these fruitless attempts to strengthen the completeness of the physical. I think what the *Sufficient Condition* might be hinting at is that the idea that there might be two different kinds of explanations for the physical behavioural effect, p^* . One is that it is caused by M and the other that it is caused by P. These two kinds of explanations are on a different descriptive level. In this sense then we might say that both are permissible. The M explanation is permissible because it has an important personal and social function. However, regarding the

question of what things actually exist (ontologically), there is only the explanation that P causes p^* . And this explanation has to be sufficient in describing the world. The M explanation on the other hand is just a way of speaking. It would be compatible with the P explanation in the sense that it is sensible to think that M and P are causal in a particular context without violating the natural sciences, but it is not compatible in the context that we allow the M explanation in our account of how the world really is. So, the difference between the M and P explanations is that we need P in order to describe the world as it is, and M is a socially and culturally developed way of speaking. M is not a property that p^* really has but rather a particular concept. Thus, the two kinds of explanations using M and P respectively are not equally fundamental explanations. P is ontologically more fundamental, and in this sense *Sufficient Condition* would exclude the mental as a causal factor if it is not physical.⁶

So far we have come to see the idea that is roughly this: a strong dependence between m and p accounts for their relationship which ensures m 's place in the physical world, and in particular a reliance on supervenience as securing the dependence relation. We could ask for a stronger closure principle, which states that no other cases are possible, but dismissing the compatibilist's account as not plausible by changing our understanding of the Closure principle might not be the most convincing way to flesh out the problems with the compatibilist account, at least it won't be for the compatibilist. So, I suggest we turn to its reliance on supervenience and the role the principle plays in the account. This raises the question of why m and p should always be so counterfactually dependent, perhaps therefore we must ask whether supervenience is enough or not.

4.2 Why Supervenience Is Not Enough

The compatibilist maintains the distinctness of $m(M)$ and $p(P)$ because of their modal difference: "The modal difference ensures their distinctness by Leibniz's Law" (Bennett, 2003, p. 483); and because of multiple realization: "the main reason most people refuse to identify mental properties with physical ones is multiple realization –it certainly seems as though mental property M could be possessed in a variety of physical ways" (ibid.). Because of a tight relationship that is *close enough* she can defuse overdetermination. The physicalist intuition tells us that there ought to be good physicalist reasons why the duplicate of an

⁶ I owe this point to Gloria Mähringer.

ordinary rock will not have mentality, but my duplicate will. It is not enough for physicalism that there is a physical realization of the mental properties or events. We must also assume *something about the way* in which mental properties correlate with physical properties. This is the appeal to supervenience. Supervenience assures that physical properties determine all the mental properties that are had. But does it establish that an asymmetrical necessity between $m(M)$ and $p(P)$? Is it enough to ensure that a subvenient property or event is more fundamental than the supervenient property or event?

So far we have defined the supervenience thesis called strong supervenience. Whether this principle captures an essential component of physicalism depends on whether physicalism is a contingent thesis or not. Many of the supervenience claims (such as the local variety) may need some modification to render physicalism contingent. Our supervenience above ("*Strong Supervenience*") may well allow for physicalism to be contingent, enough to satisfy a minimal commitment to physicalism. But while supervenience is a necessary component for physicalism, it cannot satisfy the physicalist intuition by itself. The reason is, as Robert Francescotti (2014) rightly points out that "the supervenience of the mental on the physical is perfectly compatible with mental properties being instantiated in a wholly non-physical fashion" (ibid., p. 34). In fact, it is compatible not only with property dualism, but also with substance dualism. The following example by Francescotti will hopefully make the point obvious:

"Suppose that any bearer of mental properties is comprised of an immaterial soul, existing in some non-physical realm, in addition to a physical body. Suppose also that all immaterial souls are dependent on the operations of physical bodies in such a way that any variation in soul properties occurs only with a variation in physical properties of the body. Then any physical duplicates of this world will be mental duplicates of this world. [...] This scenario is compatible even with Kim's strong supervenience. Suppose that mental properties are instantiated only in immaterial souls. Suppose also, that necessarily any soul x that has an immaterial soul has a physical body on which that soul is dependent, and dependent in such a way that each mental property of the soul is necessitated by some physical property of the body. In this case, necessarily, for any mental property M of individual x , there is some physical property P , such that

necessarily for any individual *y*, if *y* has *P*, then *y* has *M*. So the mental strongly supervenes on the physical in this case despite the substance dualism” (ibid., p. 35).

If only one thing could be certain about physicalism, certainly it would be that it excludes substance dualism. Physicalism simply does not allow substance dualism to be true. Physicalism, or the idea that physical facts fix the mental facts, entail that physical duplicates of our world have all the actual mental episodes as well. Note though that this allows for our actual world to have mental ‘extras’. However, with the exception that these mental additions don’t interfere with the physical laws that obtain, as per physical completeness, also implicit in principle 3). In this way the physicalist view that all mentality is exclusively a function of the physical is kept. What we should have learnt by now from Descartes and Elisabeth of Bohemia’s letter correspondents is that as long as the spirits, soul, mental episodes, or whatever we want to call it, does not interfere with the operations of physical laws upon physical substances, they may very well coexist alongside each other, but they cannot interfere. And what we have already seen is that supervenience is also compatible with substance dualism, thus it alone cannot capture adequately the content of physicalism. We must formulate supervenience in such a way that we may avoid substance dualism. Francescotti recommends that “we simply need to conjoin a supervenience thesis with some constraints on the *composition* of mental items” (ibid. p. 35). This is precisely what is implied in our third principle as discussed in 4.1.

Principle 3) *Physical Causal Closure*, states that every physical occurrence has a sufficient physical cause. Built into the principle is that everything concrete is exhausted by basic physical objects (Hellman & Thompson 1975). Essentially, the third principle implies the constraint that each mental property (or mental particular) in a physicalist framework is a physical item or else it may be decomposed into parts that are physical items. But the third principle is also by itself not enough for physicalism, since it is compatible with the possibility of there being a world, indistinguishable from our world in every physical respect but has a radically different distribution of mental properties (perhaps even entirely devoid of mentality). Now, we see that in order to capture physicalism we must conjoin the third principle and supervenience.

The reason **T1** cannot be vacuous in Bennett's account is that she believes property M can be instantiated without any physical property. Remember her argument that "the claim that it is impossible for property M to occur without any relevant P-like property basically amounts to the claim that physicalism is necessarily true. [...] Though there may not be any souls here, there are worlds in which there are, and in those worlds things can have M without having any physical properties at all. So, property M can indeed be instantiated without any physical property" (Bennett, 2003, p. 484). She also argues against the so-called 'upward' necessitation relation. "This particular pattern of neural firings could occur in a petri dish" she writes (ibid., p. 484). "The instantiation of the property *being a C-fiber firing* does not guarantee the instantiation of the property *being in pain*; again, C-fiber firings can occur in petri dishes" (ibid., p. 485).

The point is that the counterfactual **T2** can be false if, as Bennett puts it, "barring a metaphysical miracle, *m*'s being gone" would involve such changes that would change what *p* causes. (Bennett p. 488) Now, recall that Bennett insists that **T2** being false does not imply or mean that *p* in anyway *needs m*'s help to bring about *e*. The alternative she presents is that the conditions which must hold for *p* to make *e* happen, are the same as those conditions in which *p* necessitates *m*. Imagine a world *w*, where a complex physical process takes place that involves your gastrointestinal tract and various hormones, in particular ghrelin. Signals are sent to your brain through your vagal nerve fibers. Usually, in our world this is followed by a sensation or your phenomenal experience of feeling hungry and that you head to the fridge looking for something to eat. In this imagined world, however, your hunger sensation just doesn't arise and so you won't go looking for food. In other words, zombie-worlds are just not a possibility. Zombie-worlds would allow the exact same process that occurred at *w*, but you would still go to the fridge looking for food, even without having experienced the sensation of hunger. Indeed it seems to be a powerful response to the exclusion problem.

It means that the kind of necessitation between *p* and *m* that Bennett is advocating is such that, there cannot be any worlds where it is "possible to strip the mental off the world" in such a way as in zombie-worlds. But the more serious issue for Bennett is that she has not, as she says she has, managed to evade the force of the exclusion principle and remained a physicalist. The grave problem seems to be as Francescotti puts it:

“[...] mental properties not being identical with physical properties prevents the physical facts from necessitating the mental facts in the way that physicalism requires.” (Francescotti, 2014, p. 42)

It seems Bennett’s use of supervenience, and the tight connection between *m* and *p*, is such that the psychophysical laws are not purely physical laws. Both vacuity and the falsity claims show that there is a difference between what *p* without *m* and *p* with *m* can or does cause. The actual world would be different without *m* :

“The claim is rather that if the mental cause had not happened, that just *constitutively involves* various changes in the world that change, or at least may well change, what *p* causes. [...] *m*’s being gone partially just *is* these other changes.” (Bennett, 2003, p. 488)

If it were the case that there are non-physical properties irreducible to physical ones we would need, in addition to physical laws, psychophysical laws to connect the physical properties with non-physical properties. Francescotti claims that “[i]f a physical duplicate of the actual world with radically different mentality were possible, then clearly the mental facts that actually obtain would be a function of more than just the physical facts, contrary to physicalism. [...]if fixing the mental facts requires psychophysical laws, then fixing the physical facts alone is not sufficient to fix the mental facts” (Francescotti, 2014, p. 37)

This suggests that if the physical laws allowed worlds that had the exact same distribution of physical properties but had a different distribution of mental properties, there would be a sense in which the mental facts are at least in part conditioned by something other than just physical facts. The covariance that seems to be required for a tight relationship, that makes it the case that *M* is instantiated whenever *P* is for any property *M* and any property *P* on which *M* supervenes, are the actual physical laws that obtain. But perhaps there is a further need to explain why these laws themselves are true. Given that under physicalism the mental facts are solely a function of the physical facts, the requirement should be that the mental facts are determined by the physical facts in such a way that the distribution of physical properties and the physical laws together would make it impossible for the mental facts to be absent. Now, it is Bennett’s position that the same requirements hold for the non reductive physicalist view. So, it is impossible for the mental facts that obtain to fail to do so as long as the physical facts

are as they are in our actual world. It is a trivial fact that, were the mental properties identical to physical properties there should be this kind of necessitation. But according to the compatibilist mental properties are not physical, then what reasons do we have to expect the physical facts to determine the mental facts? Bennett has given us very little in her account in way of answering this necessitation. Really, all she has said is that “the conditions that must hold for p to bring about e —physical conditions, note—are *basically the same as the conditions in which p necessitates m* ” (Bennett, 2003, p. 488-489)

What then grounds the mental properties? In section 2 we discussed the option to inject supervenience to secure the relationship that must hold between the mental and physical for it to be a physicalist account. Indeed it is a minimal requirement. Bennett emphasizes the tight relation: “we would have to say that it is impossible for anything to have the neural property without also having the desire property” (p. 485). Why would e not occur if p occurred without m ? According to Bennett because the conditions that must hold for p to bring about e would be different. Were p to occur without m , then the conditions which are necessary to bring about e would not hold – and p would not cause e . (ibid., p. 488-499). So, is supervenience enough to establish the asymmetrical necessitation between m and p ?

As we have examined, for the counterfactual account to be different from the traditional firing squad example and avoid bad overdetermination is to claim that overdetermination is not a coincidence. But the fact that it is not a coincidence does not automatically make overdetermination permissible. We could imagine a world in which everything is just the same as we now experience, except in this world there is a meddling, Malebranchian god. This god steps in and causes effects, even all those effects that already have causes that are sufficient enough to bring about those effects. This would be a world in which everything would have two distinct causes, one supernatural and one natural. And they would be sufficient causes that wouldn't be coincidental. But, surely this is not what is going on in the physical/mental case. Then it would be no different from the textbook example for overdetermination. Even the compatibilist agrees that something more needs to be doing the work: “What is doing the work?”, she asks, and continues, “[t]he difference, the compatibilist will say, is that there is an important tight relation between the mental and physical that just does not hold between the two shootings” (ibid., p. 475).

However, this necessitation relation fails to give us the independent reasons we need for establishing that a relation of relative fundamentality holds. What I mean is that the supervenience is not enough to guarantee that the supervening property or entity is nothing over and above the subvening entity, as the example of the meddling god clearly showed. Jessical Wilson (2012) clarifies the supervenience base formulation of physicalism and writes that its “aim [is] to characterize the relation between strictly physical entities and other entities in modal correlational terms that are strong enough to guarantee “nothing over and aboveness” of the latter vis-a-vis the former, while being abstract enough to accommodate the irreducibility of the latter to the former” (Wilson, 2012, p. 9). Thus, if such a relationship fails to hold then what reasons do we have to assume that our world is not just the world with the meddling god? In such a world, the mental properties would supervene with metaphysical necessity on physical properties just the same, only the former (the mental) would clearly be an ontological addition to the latter. And as Wilson (2012) rightly points out, this case shows that asymmetrical necessitation is compatible with “over and aboveness”. Even in cases of a tight relation, asymmetrical necessitation fails as a criterion of non-fundamentality. Consider the example from Wilson that shows why it is insufficient even with a close relationship. She writes:

“Consider, for example, a theistic metaphysics according to which we live and move and have our being in – that is, are grounded in – God. On such a view, I am non-fundamental, but I asymmetrically necessitate God: my existence entails God’s existence, but assuming that God has a choice about who or what to create, God’s existence does not entail my existence” (ibid. pp. 9-10)

What this example shows is that, not only is asymmetrical necessitation insufficient for non-fundamentality even in the cases with a tight relationship, but also such necessitation is not necessary. In order for permissible overdetermination, that is to justify just why we can allow for two sufficient causes to be the cause of the effect *e* while not having to conclude that mental efficacy is blocked (and staying within physicalism), it is necessary to guarantee “nothing over and aboveness”. But this is fully compatible with the supervenience principle. In other words, what the examples show is that it is not enough for Bennett to secure supervenience

and also argue that the mental and physical properties and events are on the same level with the same causal powers. Remember, the game that the compatibilist wants to play is to deny overdetermination while holding fixed the full-fledged causal efficacy of the mental. The only problem is that she cannot do so without grounding the mental in the physical. The compatibilist account thus includes over and above. Again, given that we have good reasons to deny that Bennett has anchored counterfactual compatibilism with a supervenience relation to physicalism, she has not established how supervenient properties are grounded in physicalism. And her account has a dualist leaning rather than being a materialist approach.

4.3 Bennett's Reply to Objections

In her 2008 publication, Bennett has given a reply to some of her critics and answered some objections that are similar to the issues I raise in this essay. Her (2008) article argues that the solution to the exclusion problem offered in (2003) is not available to the property dualist, since "*compatibilism requires physicalism*" (2008, p. 18). I suspect therefore that she would strongly disagree with my conclusion in the previous section. But has our analysis of her account shown that indeed the property dualist solution is available also for the substance dualist? I think we have extensively argued for it. Let's repeat why.

Supervenience (or *Strong Supervenience* as we have defined it) is a minimum requirement for physicalism. Everything must have the relation to supervene with metaphysical necessity upon the physical. Most kinds of physicalist should accept this principle, even the compatibilist who aims to be physicalist agrees to it. But Bennett rejects that physicalism should require something more. She states the following:

"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” (Bennett, 2008, p. 18-19).

Others have also argued that we need an explanation, or grounding of the mental. Andrew Melnyk, for example has argued that metaphysically necessary supervenience does not guarantee that realization holds. Both Melnyk and Frank Jackson have argued that the metaphysically necessary supervenience claim is consistent with dualism (Melnyk 2003, p. 58; Jackson 2006, p. 243). And Bennett considers to this line of criticism and writes:

“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” (Bennett, 2008, p. 21).

However, in 4.2 we showed just how it is compatible with ontologically distinct properties as sufficiently casual for an effect. Clearly, nothing over and above is necessary for there to be such a notion as “permissible overdetermination”. But as Wilson (2012) has shown, over and abovness is compatible with Supervenience; and that is why Bennett’s counterfactual compatibilist solution is not sufficient for physicalism, as per my conclusion in 4.2.

5 Conclusion

In the philosophy of mind, the Causal Exclusion problem has been a puzzle about just how mental phenomena fits into the physical world. It composes the problem of how we can account for the mental being efficacious while maintaining the principle that the physical causes are complete; that is to say that all physical effects have sufficient physical causes. In this essay we have considered the forceful account by Karen Bennett, which argues that overetermination (there is more than one sufficient cause for an effect) might not be such a bad thing after all, if we just have the correct understanding of what kind of overdetermination the mental and the physical cause. The great advantage of a Compatibilist view is that it really does seem to preserve distinctness between the properties or events, while satisfying physicalist intuitions that every physical effect has a sufficient physical cause. Bennett's strategy is to change how we understand the mental and the physical overdetermining the effect. Using counterfactual statements to show that a physical property or event occurring without the mental property or event, the event, e , would not have occurred, and thus arguing that they are not coincidental causes, but that there must be a tight relation between the distinct properties or events. The tight relation is a metaphysical necessitation relation that holds between the two. And while p is sufficient to cause e , m is also a sufficient and distinct causal factor for e . However, what I hope to have shown is that if we take a closer look at the principles that are basic physicalist commitments, they are not all met by the counterfactual compatibilist. First of, we have strong reasons to have a stricter, more robust understanding of the physical causal closure. One that really says that the physical is wholly and completely sufficient to cause all the physical effects. I believe a materialist has the strong intuition that our principle which is supposed to secure physicalism cannot be taken to be so weak that it would allow for other ontologically distinct causes, in addition to sufficient physical causes. This would amount to violating a robust understanding of physicalism. Nevertheless, to suggest that she must have another reading of the exclusion problem might not be a convincing enough argument for the compatibilist. However, we also know that the principle by itself is not enough for physicalism, since it is compatible with the possibility of there being a world, indistinguishable from our world in every physical respect but has a radically different distribution of mental properties. So, in order to capture

physicalism, the Compatibilist must conjoin the principle and supervenience. But injecting supervenience alone to her account, I argue, is not enough to secure an asymmetrical necessitation between the mental and the physical. The necessitation relation fails to give us the reasons we need in order to establish that the supervening property, M, is nothing over and above P. As long as this necessitation remains unexplained in the counterfactual compatibilist account, it is hard to see just how it has managed to keep M and P at the same level while also explaining M's causal role in our physically causally complete world. That's why, at least for now, the causal exclusion problem is still an intractable problem for the counterfactual compatibilist.

References

- Armstrong, David. 1997. *A world of states of Affairs*. Cambridge: Cambridge University Press, xiii, 285.
- Bennett, Karen. 2003. *Why the Exclusion Problem seems Intractable, and How, Just Maybe to Tract it*, *Noûs* 37: 471–471.
- Bennett, Karen. 2007. *Mental Causation*. *Philosophy Compass* 2/2:316-337.
- Chalmers, David. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- Davidson, David. 1970. *Mental Events*. Reprinted (1980) in *Essays on Actions and Events*. Oxford: Clarendon Press, 207-225.
- Engelhardt, J. 2015. *What is the Exclusion Problem?* *Pacific Philosophical Quarterly*, 96, pp. 205-232.
- Francescotti, Robert. 2014. *Physicalism and the Mind*. SpringerBriefs in Philosophy.
- Francescotti, Robert. 1996. *The Non-Reductionist's Troubles with Supervenience*. *Philosophical studies* 89: 105-124.
- Horgan, T. 1997. *Kim on Mental Causation and Causal Exclusion*. *Nous Supplement: Philosophical Perspectives* 11: 165–184.
- Haug, Matthew C. 2009. *Two Kinds of Completeness and the Uses (and Abuses) of Exclusion Principles*.
- Haug, Matthew C. 2011. *On the distinction between reductive and nonreductive physicalism*.
- Hellman, Geoffrey Paul & Thompson, Frank Wilson. 1975. "Physicalism: Ontology, determination and reduction" *Journal of Philosophy* 72 (October):551-64.
- Jackson, Frank. 2006. "On ensuring that physicalism is not a dual attribute theory in sheep's clothing." *Philosophical Studies* 131: 227-249.

- Kim, Jaegwon. 1989. *Mechanism, Purpose, and Explanatory Exclusion*. *Philosophical Perspectives* 3: 77–108.
- Kim, Jaegwon. 1993. *Supervenience and Mind*. Cambridge: Cambridge university Press.
- Kim, Jaegwon. 2007. *Causation and Mental Causation*. In *Contemporary Debates in Philosophy of Mind*, edited by B. P. McLaughlin, and J. D. Cohen. oxford: Blackwell.
- Kim, Jaegwon. 1998. *Mind in a Physical World*. Cambridge: MIT Press.
- Kim, Jaegwon. 2005. *Physicalism, or Something Near Enough*. Princeton: Princeton university Press.
- Kim, Jaegwon. 2006. *Emergence: Core ideas and issues*. *Synthese* 151: 547–559.
- Kim, Jaegwon. 2009. *Mental Causation*. In *Oxford Handbook of Philosophy of Mind*, edited by B. McLaughlin, A. Beckermann, and s. Walter, 29–52. oxford: oxford university Press.
- Kripke, 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Lewis, David. 1973. *Causation*. *The Journal of Philosophy* 70: 556–567.
- Lewis, D. 1979. *Counterfactual Dependence and Time's Arrow*. *Noûs* 13 (4): 455–476.
- Lewis, David. 1986. *On the Plurality of Worlds*. oxford: Blackwell.
- Lewis, David. 1999. *New Work for a Theory of Universals*. In his *Papers in Metaphysics and Epistemology*, 8–55. Cambridge: Cambridge University Press.
- Lewis, David. 2000. *Causation as Influence*. *The Journal of Philosophy* 97: 182–197.
- Melnyk, Andrew. 2003. “A Physicalist Manifesto: Thoroughly Modern Materialism.” Cambridge: Cambridge University Press.
- Melnyk, Andrew. 2008. “In Defense of a Realization Formulation of Physicalism.” *Topoi* 37, 483–493 (2018).

- Melnyk, Andrew. 2015. "The Scientific Evidence For Materialism About Pains", in Steven M. Miller (ed.) *The Constitution of Phenomenal Consciousness: Toward a Science and Theory* (John Benjamins Publishing Co., 2015), pp. 310-329.
- Moore, Dwayne. 2017. *Mental Causation, Compatibilism and Counterfactuals*. Canadian Journal of Philosophy, 47, 1, pp. 20-42.
- Moore, Dwayne. 2019. *Causal Exclusion and Physical Causal Completeness*. dialectica Vol. 73, No. 4, pp. 479-505.
- Sider, T. 2003. *What's so bad about overdetermination?* Philosophy and Phenomenological Research 67: 719–726.
- Papineau, David. 2001. *The rise of Physicalism*. C. Gillett and B. Loewewr, eds, Physicalism and its Discontents, Cambridge: Cambridge University Press, pp. 3-36.
- Papineau, David. 2009. *The Causal Closure of the Physical and Naturalism*. The Oxford Handbook of Philosophy of Mind.
- Pereboom, Derk, & Kornblith, Hilary. 1991. *The Metaphysics of Irreducibility*. Philosophical Studies 63:125-145.
- Roche, Michael. 2014. *Causal Overdetermination and Kim's Exclusion Argument*. Philosophia 42:809-826
- Van Gulick. 1992. *Nonreductive Materialism and the nature of Intertheoretical constraints*. Essays on the Prospects of Nonreductive Materialism. Berlin: Walter de Gruyter, pp. 157-178
- Yablo, Stephen. 1992. *Mental Causation*. The Philosophical Review 101: 245–280.
- Wilson, Jessica. 2012. *Fundamental Determinables*. Philosophers Imprint Volume 12, NO.4.