

NONREDUCTIVE PHYSICALISM AND MENTAL CAUSATION*

Miguel Amen
jmanen@gmail.com

ABSTRACT

In this work I articulate and defend a problem about the place of the mind in the causation of behaviour. Ask why someone did a certain action and you can see the problem arise, if only you assume certain plausible suppositions about the world. The suppositions are taken to be those of nonreductive materialism.

I think that the argument from exclusion, originally developed by Jaegwon Kim, shows that unless there is overdetermination, the mental cannot be causally relevant in the causation of behaviour. It is my view, however, that a proper understanding of overdetermination shows that the overdetermination move is not available to the nonreductive physicalist. That is, he cannot escape exclusion by claiming that the mental overdetermines the physical in the causation of our actions.

It is argued that neither appeals to economy nor to Bennett's counterfactual test are good ways to decide matters of overdetermination. That should be decided in terms of the ability of a theory to consistently permit such overdetermination, which however is shown not to be the case for nonreductive materialism.

Moreover, in general all realized properties will face this problem - assuming them to be causally relevant will ignite exclusionary claims and in the competition for relevance, physical properties will have a better and more fundamental claim for relevance, threatening once again to relegate realized properties to the category of epiphenomena.

*O texto que agora se publica, dividido por este e pelo próximo número da Revista, corresponde com alterações de pormenor à tese de mestrado apresentada pelo autor no King's College da Universidade de Londres em 2005.

When the nonreductive materialist insists in defending his views against the calls of causal exclusion, I will show that he moves either towards emergentism, which seems to be something that in the end he does not want to hold because it violates completeness, or to type physicalism, which however is unwelcoming because it dilutes the distinctness between the mental and the physical and the calls of multiple realization.

CONTENTS

Introduction

Section I

1.1 Introduction

1.2 Anomalous Monism and Exclusionary Claims

1.3 Causal Properties and Causally Relevant Properties

1.4 Causal Explanation

1.5 Events

1.6 Reductive and Nonreductive Materialism

1.7 Exclusion

Section II

2.1 Introduction

2.2 Overdetermination and Completeness

2.3 More on Overdetermination and Completeness

2.4 Causal Compatibilism and Bennett's Test

2.5 A new way to Understand Overdetermination

Section III

3.1 Introduction

3.2 (A), (B) and (C)

3.3 Emergence

3.4 Numerical Identity of Causal Powers

3.5 (D) and (E)

3.6 Problems for (D) and (E)

3.7 More Problems: a bit of Speculation

Conclusion

INTRODUCTION

I want to articulate and discuss a problem that has been significant in the philosophy of mind for some time. It is a problem about the place

of the mind in the causation of behaviour. Ask why someone did a certain action and you can see the problem arise, if only you assume certain plausible suppositions about the world.

Why did Sophie open the window? Because we know that she wanted to breathe some fresh air and believed that by opening the window she would be able to do it. We explain her action by pointing out her desire to breath fresh air and the belief that a certain course of action, opening the window, would achieve the desired result. This explanation, which is called intentional or mental, explains by citing as causes of behaviour certain aspects of Sophie's mental life at that moment- the content of her thoughts. As such, an intentional explanation is a form of causal explanation, having as causes mental particulars. Furthermore when such an explanation is given it is reasonable to think that we were given a sufficient explanation of behaviour. By that I mean that an intentional explanation can be fully satisfactory without appealing to non-intentional discourse.

However that is not the only causal explanation that can be given of Sophie's action - her action, after all, is an action in a physical world. The opening of the window by Sophie is a complex of such and such movements of her fingers, arms and body. And such movements have a physiological story; one that we assume can be explained *fully* by mention of muscles, bone movements, nerve cells and neurological activity.

When we explain an action, we explain a piece of behaviour, and as we just mentioned, we can explain *that* behaviour, by reference to two sufficient explanations that are, *prima facie*, independent of each other; an intentional explanation and a physiological explanation. The intentional explanation will mention mental properties as relevant causal factors of behaviour while the physiological will mention only physical properties.

Now, if we include in this story certain widely held beliefs about the irreducibility of mental properties to physical properties we get to see that those two causal explanations might bring us more problems than might be apparent. We seem to experience a certain tension between two such irreducible causal features that are used to explain a single explanandum, which results from a certain impression of competition between causes. After all we are talking about causal

explanations, and causal explanations explain by mentioning¹ relevant causal features of the world. This brings us to a certain sense of profligation of causes, which we cannot get away by identification. Now, assuming that the world is not extravagant in this way, we seem to have to choose among the putative causes, the real causes of behaviour. Having into account the causal closure of the physical world, it seems that if we want to avoid overdetermined effects when explaining human behaviour we have to drop the intentional explanation.

We started out assuming that mental explanations were real explanations, no mere pretence or convenient fiction, that the causal factors that were cited as relevant in the causation of behaviour were actual causes, and we very quickly end up with a position where the mental has nothing to cause, because we find out that the causal structure of the world, exactly where mental causation is supposed to matter, is fixed by the physical causal factors of the physiological explanation.

In the first section I will develop this idea; certain distinctions will need to be made concerning the causal efficacy of events and the causal relevance of properties. I will begin by setting up the problem as I think it is most interesting, and this means that it will be not a problem about the causal efficacy of events but about how can mental properties of causally efficacious mental events be relevant for the effects of those events. I will discuss and establish certain ideas of nonreductive physicalism that are important to understand the challenge it faces from the exclusion argument. I try to show that this argument points to a deep tension within the metaphysical commitments of nonreductive physicalism that ultimately should lead us to the view that the mental and other special science properties cannot be seen as causal properties.

In section II I discuss the problem of overdetermination, where it is defended that overdetermination cannot be ruled out based on the principle of economy or based on the counterfactual test, but that it is a matter of consistency within theories or metaphysical systems. The nonreductive physicalist could escape the exclusion argument if he could say that there was overdetermination. But to be able to say this he

¹ Assuming, throughout, causal realism.

has to show that his metaphysical commitments are consistent with the view that both the mental and the physical are causal properties.

In section III it is shown that this project cannot be fully developed. The nonreductive physicalist, in order to give sense to the causal relevance of the mental seems to be torn between emergentism or reductivism both of which are not nonreductive physicalist theories. The conclusion of this chapter is that nonreductive physicalism cannot make sense of overdetermination, and consequently it is not consistent with the view that the mental is causal.

SECTION I

1.1 Introduction

I will start out with a discussion of the problem that besets Davidson's anomalous monism, using it as a dialectic partner in finding and defending the ontological claims that frame the discussion and of establishing its modern form, as I will be addressing it. I start with Davidson not only because of its historical importance but because it illustrates the problem with which we will deal in quite an instructive way. But very soon, particularly after 1.6 I will take as target a version of nonreductive materialism that has as source the argument from multiple realization, and not the anomalous character of mentality.

1.2 Anomalous Monism and Exclusionary Claims

Davidson's anomalous monism was (with multiple realization) extremely important in changing the consensus from type identity theories toward nonreductive physicalism. In it Davidson (1970) defends ontological monism, in the sense that every event is a physical event, and resists the reduction of the mental, based on his defence of the anomalous character of the mental realm, giving it a welcoming (so was the mood) autonomy from the physical world. His position can be seen as resulting from the following three principles:

Mental-physical interaction: Mental events interact causally with physical events.

The nomological character of causality: Causally related events must instantiate a strict law.

The anomalism of the mental: There are no strict laws subsuming mental properties.

Now, if mental events cause physical events, they must instantiate a strict law. However due to the anomalism of the mental, those laws cannot be psychological or psychophysical laws. They must, then, instantiate a physical law. But if mental events instantiate physical laws they must be physical events. This is so because according to Davidson an event is mental or physical if it has a true description that identifies the event as one kind or another. Since a causal efficacious mental event instantiates a physical law, it can be identified by a physical description. So every mental event² is a physical event.

In one important way anomalous monism solves one form of the problem that we were pointing out at the beginning. There we expressed in a vague way that if we assume that both the mental and physical could be cited in causal explanations of behaviour we would get a certain tension between them. One way to express this problem is to inquire about the causal efficacy of mental events. Say a particular mental event *m* is a cause of a certain behavioural event *p* and *p** is a cause cited in a physical explanation of *p*. We might put it this way, following our thoughts in the beginning:

- (1) *m* is a cause of *p*.
- (2) *p** is a cause of *p*.
- (3) *p** and *m* are distinct.
- (4) There is no overdetermination.

Unless we challenge some other premise we seem to be compelled to conclude that *m* is not a cause of *p*, that is, that (1) is false. After all we know that completeness, subsequent to anomalism and the nomological character of causality, must hold and consequently that any physical effect must have a sufficient physical cause. That is why (2) must stay. In such a case *p** must be a cause of *p*.

Nevertheless this way of setting the exclusion problem is not at all a problem for anomalous monism, because here we have a theory that rejects (3) and thus can save the causal efficacy of mental events. If we understand that causation is an extensional relation between events

² If we assume that every mental event is causally related to a physical event or to a mental event that is so related.

however described, then if mental and physical events are identical entities, and if the physical event p^* causes p , then so does m , since m and p^* are the same event. Now, an event dualist still has to deal with this problem, but after Davidson there is a consensus that our mind-body theories should be physicalist in that minimum sense that Davidson endorses, viz., that mental events are physical events, and consequently, for those theories of mind that embrace minimal physicalism, there is no mental causation problem about the causal efficacy of mental events.

Still, soon after Davidson published his theory, a torrent of papers³ was published advancing the proposition that anomalous monism was letting the mental down. The charge made by these critics was that anomalous monism was in a certain sense a version of epiphenomenalism, because, it was said, mental events got into causal relations with physical events, not because of their mental properties but because of their physical properties.

This idea picks up the nomological character of causality to tease up a way in which it makes sense to say that it is because of instantiating a physical law that mental events are said to be in causal interaction with the physical realm. The idea is that independent of the extensionality of the causal relation, it makes sense to look into the events themselves, those that are the causes of behaviour and to selectively focus on those properties in virtue of which those events instantiate a law, and those properties that do not enter into lawful relations.

Central to anomalous monism is the claim that mental properties never get into the kind of laws that *prima facie* are required to ground causality. Consequently it seems to make sense to say that the causal relation is grounded on the physical properties of those events and not on the mental properties that those events have. This makes mental or intentional explanation, explanation couched in terms of content properties like beliefs and desires, causally irrelevant for the causal relation that those mental events enter into. It seems that the causal efficacy of mental events is grounded on the physical causal laws that the physical properties of those events instantiate.

³Some of the critics are: Honderich (1982), Sosa (1984).

One might think, and correctly, that the thought expressed by these critics is awfully wrong, when made having Davidson as a target, because it assumes that there is something about those mental events, their mental properties, that are especially problematic. However, even if one countenances properties, Davidson views on causality, as essentially an extensional phenomenon between particulars, might be said to rule out the view that mental properties or physical properties have anything to say about the causal relation that events get into. The point is that for Davidson, events, if they are efficacious in a particular situation, are efficacious tout court, and not in virtue of their properties.

It is important to note that the main use of anomalous monism here is not to inquire into its deep truth and the scholastic interpretations that it requires, but as a way to express certain difficulties about mental causation that first, historically, were expressed about it but that are, once it is understood their main motivation, in fact quite general about theories that accept token identity but make the distinction between mental and physical properties essential to their view of the world. For such theories, even if the causal efficacy of mental events is granted there is a question about the causal relevance of mental properties of causally efficacious mental events. Even if Davidson can escape the charge I think the lesson is interesting for our purposes.

Moreover, it seems to me that this has been the way that Davidson has been read by most of the critics when they say that it leaves the mental epiphenomenal, albeit wrongly as we have said. Returning to the charge that anomalous monism, *as some of its critic have understood it*, cannot account for the relevance of mental properties, and assuming that we can talk about the mental properties of those events, the important point to make in relation to our problem is that even if mental events are efficacious, there is a problem about mental properties. And the problem is about the *homogeneity* of mental properties vis-à-vis physical properties in their standing to causation. We want that mental properties that we cite in intentional explanations be as causal as the properties that we cite in physical explanations. Mental properties have to be causal properties; otherwise there is a strong sense that the reasons we refer to in the explanation of behaviour are powerless in the causation of our actions.

The problem seems to be that there is something about anomalous monism, as some of its critics have understood it, which makes mental properties unsuitable to be causally relevant properties. Here, the problem is not that there is competition between mental and physical properties, but that mental properties, by their own nature, are especially unsuitable to be causal properties. The problem seems to be that their anomalousness together with the nomological character of causality is thought to imply that *only* strict nomic properties are causal properties. The idea is that strict nomicity (or nomicity simpliciter) is a necessary causal grounding property, and only those properties that are nomic can be causal. Since the mental is not appropriately nomic, it cannot do any pushing qua mental.

This argument depends crucially on the idea that only (strictly) nomic properties are causal. But that does not seem to be what is implied by Davidson's principle of the nomological character of causality. This principle says that causally related events must instantiate a strict law. But that does not mean that there isn't another causation grounding relation X that could do the job. It only means that if there is such causation grounding relation then the events that such grounding relation establish as being in a causal relation, must instantiate a strict law⁴. So if X is such a causal grounding relation and anomalous mental properties satisfy it, then mental properties would be causal properties. Of course those events would also have to instantiate a strict law; a law that would refer to the physical properties of those events.

Then we would have for every action, two causal explanations, one mental and another physical. And so the problem of competition between the causal factors shows itself⁵:

- (1) mental properties are causal properties.
- (2) physical properties are causal properties.
- (3) mental properties are not identical to physical properties.
- (4) physical events are not overdetermined by the causal properties of their causes.

So even if anomalous mental properties are somehow suitable by their own nature to be causal properties, they are however actually

⁴McLaughlin (1989) defends extensively the argument being made.

⁵I will be more precise about this in the beginning of section 1.7

unable to be causally relevant properties because they are preempted of their causal work by the physical properties of efficacious mental events. Notice that (2) is there as a result of completeness: Every physical effect e with a cause c , has a full causal explanation in terms of c 's physical properties.

After Davidson, the main strand of the problem of mental causation, the one we will deal in this work, is the problem of accounting for the causal relevance of mental properties of causal efficacious mental events. Can mental properties as causal properties do their work? They could, easily, if mental properties were identical with physical properties. However the position we want to study has a different view; mental properties of efficacious events are not reducible to the physical properties of those events. In order to see if such a position can somehow keep distinction and maintain that mental properties can do their causal work it's convenient to be clearer about the ontology of nonreductive physicalism. First however lets get a little more precise about what a causally relevant property is.

1.3 Causal Properties and Causally Relevant Properties

The notion of a causally relevant property is fundamental to our project and it's a difficult one, so here I want to go into it with a little more detail and show how it should be understood.

What I have been saying is that a causally relevant property is one that is causal. But what is a causal property? It is not, obviously, a cause, because we have been subscribing to Davidson's view that the causal relation is a relation between particular events. We can start to figure it out, however, by noticing that a causal relation cannot be just a brute fact about the world. That is, if c and e are causally related and c and e^* are not (and the context is the same), then it is reasonably to think that something about c , some property of it P , connect to the way e but not e^* is. If that is the case then we can say that c caused e in virtue of P . This expression "in virtue of" is a way to refer to the connection between P and some property of e . The nature of such connection is what I have been calling a causal grounding relation. This does *not* imply that c qua P causes e qua Q . There is no such causal relation if causal relations relate events. But the extensionality of the causal

relation is perfectly compatible with there being a relation (not a causal relation) between the properties of *c* and *e*, such that because of, in virtue of, instantiating it, *c* and *e* are causally related.

So a property *P* of *c* is causally relevant in the causation of *e* if and only if

- (a) *c* caused *e*⁶ and
- (b) There is a type relation (a causal grounding relation) between a property *P* of *c* and a property *Q* of *e*.

It is because of (b) that we can give a causal explanation of the causal relation. One way to understand the idea of causal properties is in relation to the notion of causal powers. We would say that a property *P* is causal if at least some effects that *c* has are due to the causal powers that *P* endows *c* with. The causal grounding relation specifies and connects those effects with the way the cause is.

We know that physical causal laws are such kind of type relation: once we have such grounding relation between two events we know that they are causally related. But notice that such causal laws only mention physical properties. So for the grounding of a causal relation between events based on their non physical properties we have to go further than physical or strict laws. What kinds of other type relations are sufficient to ground causation in terms of non physical properties is not our concern here, but in the literature one can see attempts that refer to non-strict laws, that is, *ceteris paribus* laws, supervenience relations, or counterfactual dependence. Our problem arises when one such causal grounding relation is given that permit us to say that mental properties are causal. Once that is warranted then exclusion problems start to do their unpleasant work.

Notice that whenever *P* is a mental property, then there will have to be a law that connects physical properties of *c* with physical properties of *e*, on top of the type relation mentioned in (b) that is supposed to ground the causal relation based on the mental properties of the events. This is to say that physical laws are basic laws. They underline every causal relation whatever.

The ontological relation between mental and physical properties that those type relations presuppose is very important to our concerns.

⁶And maybe we should add that *c* and *e* are independent of each other.

But one thing is clear; they will have certain things in common. They will have to understand mental and physical properties as distinct (more about this in section 1.6), as empowering the causal powers of events and objects in virtue of those objects and events having them, and mental properties have to be dependent on physical properties in an appropriate sense. So these general features will permit us to tackle them with a more thick brush than would have been possible otherwise.

1.4 Causal Explanation

There are some issues looming here about causation and causal explanation that I think it is useful to address because many times a discussion is being conducted at different levels of consequence that make them not very productive. Terence Horgan⁷ (2001) tell us that

“In a normal context of psychological explanation it is not appropriate to count the neurophysical realizer as causal properties in addition to the mental properties, whereas in a normal context of neurophysical causal explanation it is not appropriate to count the mental properties as causal properties in addition to neurophysical ones. In either context, such double-counting goes contrary to the contextually operative score in the causal-explanation game.” (p.104)

While there is much true in what Horgan says here, it is not, if we read him right, pertinent for figuring out the causal relevance of mental properties. Because Horgan is either talking of psychological explanation as a form of causal explanation or is not. If he isn't then he is aiming at a point that does not concern us here. But it seems clear from the context of his writing that he aims at causal explanation. But then the question we want to get to is if the properties that are mentioned in the explanation are causal properties or not. Now if we are realists about causal explanation, as we should, then the properties mentioned in the causal explanation have to be the properties in virtue of which the causal relation between the events that have those properties obtain.

But notice the mentioning of the importance of context. Earlier he mentions that

⁷All references in this section refer to Horgan (2001)

“In context, the properties that count as causal properties normally will be the ones that figure in the relevant kind of causal explanation. That is from within an engaged perspective of causal-explanatory inquiry, the properties that qualify as causal will all fall within a contextually eligible range of candidates, as delimited by the current score in the causal-explanation game” (p.103).

There is a clear sense to this. For example, if I want to know why the aeroplane crashed, then it is extremely relevant to know that the engine was faulty. Notice however that the aeroplane would not have fallen if the law of gravitation were appropriately different or nonexistent. Gravity is a clear causal property, and was relevant for the fall. But in the context of explanation, for example, in the report of the investigation team, there might not be a single mention of gravity. That of course wouldn't make it a non-causal property. The reason that there was no mention of gravity had to do with the interest-sensitivity of explanation. Still, if there is mention of a putative causal property in a causal explanation, and that explanation is true, then independently of our concerns, those properties have to be causal properties, properties in virtue of which the causal relation obtains.

Where does this leave us concerning the relation between the properties mentioned in the neurological explanation vis-à-vis those mentioned in the psychological explanation? The point is that to be included in a prima facie causal explanation does not guarantee that those properties are causal properties. And the reason is that the metaphysical nature of the causal relation is prior to our explanation. It is because certain facts of the world obtain that certain of our context-sensitive explanations are often true. But if we have evidence that the metaphysical picture does not corroborate our explanations then we have reason to suspect that our manifest image was not quite right. Curiously, Horgan, as a good nonreductive materialist puts the finger into it, but rejects the proper reading. He says that “[s]ince the higher-order, psychological, patterns and generalization are supervenient upon underlying physical facts and laws, the mental properties that are causal properties at the psychological level have their causal efficacy *via* the causal efficacy of physical causal properties that realize them. The higher-order causal properties of psychology do not generate physical or mechanical forces over and above the physical forces produced by

fundamental physical properties, and they do not intrude upon the causal-explanatory closure of physics *vis-à-vis* physical phenomena as physically described” (p.103).

But then, contrary to the earlier remark by Horgan that psychological explanations do not need to consider the neurological properties, it is appropriate to count them when giving a psychological explanation- after all they might be the ones who are doing the causal work! We count them, to be sure, in the metaphysical sense: it is there that we find the true makers of causal explanations.

In such a case as described the patterns and generalization might give us reason to use them for explanations. I am not so sure that they should be seen as causal if we end up finding out that only physical properties are causal properties. Considering that the supervenience relation is not a causal relation it seems that in such a case explanations that are based on mental properties cannot be strictly seen as causal, not without further argument, and more importantly, not resting on the fact that we might have generalizations and other counterfactual dependencies that make use of the mental discourse. I am not trying to prejudge the issue here; I am however saying that this is an issue for our metaphysics to decide. It might be that our explanatory practice was right all along - but that can only be ascertained by a coherent metaphysical package within nonreductive physicalism.

Horgan says that the “basically mistaken idea is that properties are causal, or not causal. Punkt” (p.104) but contra Horgan we can say that the world is one way or another. When talking about causal relevance and causal properties, we are saying how the world is in certain ways. Explanation is not the issue.

The moral then is that the issue of causal explanation is not a hiding place for nonreductive materialists!

1.5 Events

The view of events that I will assume throughout, will be a version somewhat similar to the Davidsonian, but assuming explicitly that events are bearers of properties, as they should – After all they are particulars. In the same way that we can say that a rock is hard or red, we can say that an event is a firing or an intending or both. And the

reason we can say this truthfully is because the event has the property or properties that make a proposition about the event true. I like this view because it allows one to identify a particular by certain of its properties without knowing all of them or the specific relations that they hold towards each other regarding those kind of particulars (that is, particulars that have the property(ies) that we first used to single them out).

We can contrast this with the view of events as the exemplification of properties. In this case we would have difficulty saying of a particular event that is a firing of x-fibres that it is also an intending, without assuming that the mental and the physical properties are identical. We want to be able to express token identity without compromising ourselves to type identity.

Moreover it makes sense to ask about the causal relations of a certain event, what would have happened if it had different properties than it actually has. It makes sense to ask if the event that was the rock hitting the window would still be a breaking of the window if the rock had a somewhat different momentum. This would alter certain properties of the event that was the hitting, while still making sense to refer to the very same event in a counterfactual situation. Unless of course, the property in question is an essential property of the particular⁸.

1.6 Reductive and Nonreductive Materialism

Token event physicalism by itself only tells us that every mental event is identical with a physical event; that all events that have mental properties have also physical properties. However it doesn't tell us if there are any interesting lawful connections between the mental properties of those events and their physical properties. For all it tells us so far, the same kinds of physical properties are found in different events with an infinite array of extremely different mental properties. That is, we might find out that we have picked out a large number of type identical physical events that share none of their interesting mental properties.

⁸This might be a good reason not to individuate events in terms their causes and effects.

This would be an unhappy state of affairs for a physicalist and is no wonder that Davidson wanted to keep a more tight relation between the mental and the physical by introducing the supervenience relation. The mental is supposed to be supervenient dependent on the physical. Davidson (1970) says,

“Although the position I describe denies there are psychophysical laws, it is consistent with the view that mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respect, or that an object cannot alter in some mental respect without altering in some physical respect.” (p. 214)

In this passage Davidson is taken to have put supervenience in the philosophical discourse. However the way that normally the supervenience thesis is today understood, as strong supervenience, is probably not open to anomalous monism as it seems to imply the existence of laws from the base properties to the supervenient properties. Strong Supervenience is the thesis that:

Necessarily, for any mental property M, if something x has M, then there is a physical property P such that x has P, and necessarily anything that has P has M.

And it seems that a more relaxed form of dependence, weak supervenience, which Davidson seems to endorse is too weak to sustain a physicalist position. Weak Supervenience says that:

Necessarily, for any mental property M, if something x has M, then there is a physical property P such that x has P, and anything that has P has M.

Such a difference between weak and strong supervenience is not detectable in a world where both hold, however once one asks how things might be in other worlds weak supervenience has nothing to say. It only says that certain generalizations hold between such properties in our world. As such it cannot give us the modal strength required to make those generalizations lawful. That modal strength however is available to properties that are related by strong supervenience. Strong supervenience gives us the certainty that in at least physically possible worlds, the same relation that holds between the supervenient and the base properties here will hold there.

There is another reason to prefer strong supervenience to its weaker form. Physicalism should not be seen as a coincidence. And that is what weak supervenience seems to imply. In such a case we could have an exact physical duplicate of our world without any of its mental properties, or with a totally different distribution of them. And that does not seem an acceptable form of physicalism. Strong supervenience however introduces a modal force with at least nomological strength that is much more plausible.

However the supervenience relation should be understood, it has to respect the motivation for nonreductive materialism. One wants to be a good physicalist, by insisting that everything is physical, and a nonreductivist by keeping the mental autonomous. But how and what reasons do we have to substantiate this position? Well, the anomalous character of the mental, as we have seen, is one such reason. Another that I want to consider is the argument from multiple realization. This is important; as it is here that we find a more elaborated and defined position that gives substance to the supervenience relation between the mental and the physical as well as giving us the principal reason, in this work, for the distinction between the mental and the physical. As we will see later it is also a source of some of the problems that the nonreductive materialist faces when accounting for mental causation.

Both reductive and nonreductive physicalism hold that mental events are identical to physical events but disagree about the identities of some of the properties of those events. While reductive physicalists think that every mental property is identical to a physical property, say pain is c-fibres firing, the nonreductive physicalist rejects such identities. He rejects them because he thinks that reduction is not a possible option due to the ontological status of the properties of the different levels of discourse, the mental and the physical. So we have a disagreement about reduction, but what is the disagreement exactly about?

In order to meaningfully speak of reduction one has to identify two levels of theory or description. The aim of the reduction is to show that one theory can be explained by the other. The standard way to make a reduction of one theory to another is to have a set of bridge principles that connect the predicates that figure in both the theories in a way that render it possible to derive all the laws of the reduced theory from those

of the reducing. Now these principles could either be conceptual or empirical depending on the theories. Where it is not possible to establish conceptual principles between the predicates of the theories then what is needed if the reduction is feasible at all are factual connections between the terms of the theories. These empirical connections will have the form of laws, bridge laws, establishing a de facto coextensiveness or express identities between the predicates of the reduced and the reducing theories. A common example is the reduction of thermodynamics to statistical mechanics. This reduction will only be possible if we are able to have bridge principles between the two sets of predicates. Since there is no conceptual connection between them, it has to be shown to exist empirical connections in the form of laws. So a predicate of thermodynamics has to be empirically connected with some predicate or set of predicates that figure in statistical mechanics in the form of laws such as 'heat is such and such average kinetic movement of particles'. Once such set of laws are in place the reduction will follow. We will be able to explain every phenomenon posited by thermodynamics in terms of the laws of statistical mechanics.

Could such bridge laws be in place in order for us to have a reduction of psychology to physics or neuroscience? Well, such laws seem to be what Davidson has called strict, because otherwise identities or coextensiveness would not come out of them. But as we have seen, if the mental is anomalous it cannot enter into such laws. So insofar as we have reason to think that the mental is anomalous, type identities are not forthcoming.

In fact, historically in the philosophy of mind, looking for such empirical identities or coextensiveness between mental and physical properties was the first step in the run from dualism. Smart (1959) and Place (1956) proposed that we should look for mind-brain correlations as an indicator of possible identities. So, my being in pain would hypothetically be identical with my being in a certain physical state, possibly my c-fibres firing. Now, this means that Pain as a type or property is identical with C-fibres stimulation as a type or property. So, 'Pain=C-fibres firing' implies that necessarily any creature that is in pain has C-fibres firing and conversely, necessarily any creature that has C-fibres firing will be in pain. Also we can say that any creature who lacked C-fibres would not be a possible subject of pain. But is this

a plausible picture? Is it plausible that an octopus, a Human and a Martian will all have to have C-fibres firing in order to be in pain?

Putnam (1967) put forward an alternative picture that suggests that it is not. Roughly, his suggestion is that the same type of mental events could be realized by different types of physical events. Say, being in pain might be the firing of C-fibres in Humans but the slow vibration of silicon in Martians. The thought is that pain or any other psychological property is a functional state of the organism and thus is not identical with the neurological properties of the brain while nonetheless dependent on those material process that realize those functional properties. Now this material processes are not individuated by their neurological-chemical structure but by the set of causal relations that they have with the rest of the organism. So, the material realizer of being in pain is whatever in that organism is causally embedded in that system (fit the causal role) such that when the organism suffers material damage in one of his surfaces it (whatever fits that causal role) causes the organism to wince and nurse the damaged area, let's say. So very different neurological-chemical properties can be brought to instantiate the same functional properties in very different creatures, and the fact that they instantiate the same mental properties in a given time will depend not on their having the same neurological-chemical properties but on having each of them a realizer that fits the causal role that is sufficient for pain. Since it is by the causal relations with the rest of the organism that we individuate mental properties on such a view, different organisms that at one time have the same mental properties will be isomorphic in respect to the causal relations of the realizer, however different the neurological properties of the realizer are.

This picture seems clearly more plausible than the type-type (property-property) identity theory of Smart and Place and while it does not show that the former is wrong it tell us that the constrains it puts on the fabric of the world are too strong if we believe that there might be psychological beings in other parts of the universe that while sharing some mental properties with us might have an altogether different biological make-up.

The idea that mental properties are multiple realizable means that mental properties cannot be identical or coextensive with any particular

kind of material properties of a realization and consequently that reduction is not possible.

Now I want to express a deep tension that I find in the nonreductivist position, a tension that I think points to a flaw in its ontological commitments. On the one hand the nonreductive physicalist wants to be a good physicalist and in a sense wants to side with the type physicalist by agreeing with the saying that “the mental is nothing but the physical”. On the other hand he believes that mental properties are a net addition to the world in the sense that the regularities of the mental (or of the special science properties in case of their being multiple realizable, which I will suppose them to be) do not find echo in the physical properties that realize them. Thus the mental discourse and the special sciences have an explanatory power not found in physical theory. It seems clear that this idea is a powerful motivation for the nonreductive materialist thesis that the mental is real. Because they bring an explanatory power not possessed by physical properties, psychological properties are an extra factor in the fabric of the world. I suppose that the best way to understand, in this context, why something brings explanatory power not available to other levels of discourse, where the explanation is seen as causal explanation, means that the explanation picks up the causal powers that explain the phenomena that are not picked up by theories that do not mention such properties. This, importantly, coheres with our initial supposition that intentional explanations are real causal explanations, and the concomitant need to understand mental properties as causal properties.

But then it seems that the nonreductivist is contradicting himself, because in such a case the mental cannot be said to be *nothing but* the physical.

This is a difficult and controversial issue that might be easily challenged. It might be argued that classification does not bring anything new, in the form of a new power or existent, to the world. This seems right to me. And *that*, it might be said, is what the nonreductivist is really saying when he talks about the explanatory potential of the mental. But that cannot be the position of the nonreductivist, as I understand him. Because he thinks that nonreductive properties are still properties in virtue of which we can causally explain, e.g. our actions through our reasons, etc. And mere classification does not introduce

new properties in the world, much less in the robust sense of causal properties. So it has to be that the classification is based on something extra. Does this make nonreductive physicalism a form of emergentism? Later we will try to separate these issues with more detail.

A note about the characterization of the mental and the physical might be useful. By physical properties I will understand those properties that are mentioned in physical science⁹. The mental then, is not physical in this sense, but for the nonreductivist the mental is physical in some sense. Because, as we have seen, for him everything is physical. Because the mental and other special sciences properties are multiple realizable by physical properties they are distinct from them. But because they are dependent on them (and the prominent dependence relation in our discussion will be the realization relation) they are physical in a broad sense. The realization relation is the key to characterizing the mental (or special sciences properties) in this broad sense. Here what we are concerned with is not the mental as eventually characterized by its intentionality or qualia. Here our concern is with its claim to distinctness taken together with its dependence on the physical vis-à-vis its causal relevance regarding mental causation. It is the object of section 3 to show that the notion of realization cannot ground both the distinction between mental properties and physical properties and sustain their dependence in suitable manner. Instead it is shown that the mental either is swallowed up by the physical or is emergent, and as such not adequately dependent on the physical. That is, it is shown that the mental cannot be characterized through the notion of realization to yield a kind of causal property that can be said to be broadly physical.

Though I will be concentrating on mental phenomena, the nonreductivist view of the world is the multi-layered view, in Kim's (1993) words,

“[Of] a hierarchically stratified structure of “levels” or “orders” of entities and their characteristic properties. It is generally thought that there is a bottom level, one consisting of whatever microphysics is going to tell us are the most basic physical particles out of which all

⁹This is not the place to discuss Hempel's dilemma and its implications, as we are giving as much credit to the nonreductivist as possible and then see if he can account for mental causation.

matter is composed (electrons, neutrons, quarks, or whatever). And these objects, whatever they are, are characterized by certain fundamental physical properties and relations (mass, spin, charm, or whatever). As we ascend to higher levels, we find structures that are made up of entities belonging to the lower levels, and, moreover, the entities at any given level are thought to be characterized by a set of properties distinct to that level" (p.337).

So the world is composed of entities that are either fundamental entities given by whatever basic physics postulates or is composed by mereological aggregates of these basic entities. These mereological relations between the different levels of description suggest an inverse pyramidal structure where at the bottom are the entities given by basic physics then at one level up of complex aggregation we have entities of chemistry then up one level those of biology and neuroscience and still in higher levels of complex aggregation the entities of the special sciences such as psychology, sociology and economics.

To fix matters we can say some of the things that the nonreductive physicalist accepts:

- Token physicalism. Every event is a physical event.
- Mental properties are causal properties.
- Physical properties are causal properties. This is due to completeness: Every physical effect *e* with a cause *c*, has a fully causal explanation in terms of *c*'s physical properties.
- Distinctness. Mental properties are not reducible to physical properties
- Dependence. Mental properties are dependent on physical properties by some such relation as supervenience or realization.
- Moreover it seems plausible to think that there is no overdetermination.

1.7 Exclusion

Now that we have a clear account of what the nonreductive materialist position is, we can see in more detail that he cannot account for the causal relevance of mental properties.

The nonreductive physicalist would like that the following propositions could be all true:

- 1) mental properties are causal properties;
- 2) physical properties are causal properties;
- 3) mental properties are not identical to physical properties;
- 4) physical events are not overdetermined by the causal properties of their causes.

However exclusion arises in the following way: say a mental event *m* is the cause of a physical event *p*. Because of (1), that is, assuming mental causation, we know that *m* has some mental property *M* that is causally relevant for *m* to cause *p*. Due to Completeness we know (2) to be the case and that *m* (\Rightarrow *p**; a physical event that is the cause of *p*) has some physical property *P** that is sufficiently causally relevant for *m* to cause *p*. Because (3) tell us that *M* is not identical with *P** nor to any part of *P**¹⁰ and (4) affirms that there is no overdetermination by causal properties we seem to be in need of negating one of the proposition (1) to (4). However, as we have seen (2) to (4) are essential commitments of nonreductive physicalism, consequently we are forced to rule out (1). That is, we are led to the conclusion that mental properties have nothing to cause, because nonoverdetermination demands that we drop one of the causal properties and completeness gives priority to the physical realm.

We have to be careful however because it might be thought that if mental properties are only causally relevant in relation to other mental properties then they might not be in competition with physical properties. The idea is that if *m* causes *m** then there must be two such causal grounding relations, one that grounds the mental properties of those events and one that grounds relevant physical properties of those events. So the completeness of physics is respected; every time a mental event causes another mental event, there is a physical law that is sufficient to ground the causal relation. Nevertheless, since mental properties are supposed to be causal properties there must be a causal grounding relation that relates mental properties of the cause with those of the effect. Put this way there doesn't seem to exist a conflict with physics and still we have a sense in which those mental properties are causal properties.

Yet what is being said is that the causal relation, *m* causing *m**, is grounded by two distinct relations, each of which is sufficient on its own

¹⁰ In case that *P** is a complex physical property.

to ground the causal relation. Now if we have reason to suppose that overdetermination was a problem before, it is not clear why it is not a problem now. It seems that once the mighty engineer had made all causal grounding work based on physical laws then he would have made all that is necessary to ground the causal relation between m and m^* .

But there is a deeper problem with this analysis - it does not go far enough in its understanding of how is it possible, in the nonreductive understanding of things, for m to cause m^* . To see why we have to notice that mental to mental causation is only possible, according to arguments made famous by Kim¹¹, if mental to physical causation is also possible¹². But the latter is quite problematic. It is the explicit way that the operative premises, completeness and overdetermination, do their work.

The first leg of the argument then is that mental to mental causation presupposes mental to physical causation. Suppose that mental property M is related by an appropriate causal grounding relation X with mental property M^* . So, in virtue of X we can say that m and m^* are in a causal relation. Now, since M^* is a mental property it follows by dependence (supervenience or realization) thesis that there is a physical property, say P^* , which realizes M^* . That is, P^* is sufficient for M^* in all the metaphysical similar worlds that are P^* . So if we wanted to know why mental property M^* was instantiated it seems that reference to P^* would be enough. It seems even that we are entitled to say that if P^* were not present then M^* would not be instantiated (provided that no other physical base that was sufficient for M^* was present). We might then ask: if P^* is enough for M^* to be instantiated, why do we have to mention M at all – M 's position in the causal loop is in peril. After all once P^* is there M^* will be there and so P^* justifies M^* without lingering doubts; no need to appeal to anything else, and in particular

¹¹ Kim 89, 93b, 98.

¹² Mental to mental causation should be understood as a way of saying that the causal relation between events m and m^* obtains because there is a causal grounding relation between M of m and M^* of m^* . Similarly mental to physical causation should be understood as saying that the causal relation between m and $m^*(=p)$ obtains because there is a causal grounding relation between M of m and a physical property P^* of $m^*(=p)$. I will however let it stand because the argument flows easier and it is standard in the literature due to the almost universal presence of Kim's version of events.

not to M. This creates a tension because it denies our claim that M is responsible for the instantiation of M*. Now, we cannot insist on an appeal to M by saying that P* and M are each a part of the total causal relevant properties for M*, because ex hypothesis, M is sufficient for M* and by the dependence thesis P* is clearly sufficient for M*.

One possibility to keep M in the causal loop would be to say that we have to accept the conclusion that both M and P* are each sufficient for M*.

This would lead to the thought that M* is overdetermined. Not in a standard way because the supervenience relation cannot be understood that way, but in a partial causal way, by M, and by way of metaphysical necessitation, by P*. The thought would be that if either one of M or P* were not present the other would be enough to secure the instantiation of M*. However on a closer look we can see that this is not the case. For this to be a true case of overdetermination we would have to say, following the standard way to think about overdetermination, that in the counterfactual situation, M* would still have been instantiated if its physical base were not present because M would have caused it, but this is obviously wrong since by the dependence thesis no mental property can be instantiated without its physical base. So there seems to be no way for M to be causally relevant for the instantiation of M* without it also being causally relevant to the instantiation of P*. That is, there has to be a causal grounding relation between M and P*.

This is what Kim calls downward causation. Causation from higher-level to lower-level. This argument seems to show that mental to mental causation is possible only if mental to physical causation is also possible.

This is a principle that Kim seems to think plausible on its own in a quite general way. Moreover he thinks that if someone fails to see the tension between M and P*, cannot fail to see the plausibility of this principle. He says in *Mind in a Physical World*¹³ that

“To cause a supervenient property to be instantiated, you must cause its base property (or one of its base properties) to be instantiated.” (p.42)

And again in *The Nonreductivist Troubles with Mental Causation*¹⁴ we are told about the causal realization principle

¹³Kim (1998)

¹⁴Kim (1993b)

“If a given instance of S occurs by being realized by Q, then any cause of this instance of S must be a cause of this instance of Q (and of course any cause of this instance of Q is a cause of this instance of S.” (p. 352)

Kim is quite right on this. A supervenient property cannot be responsible for the instantiation of another supervenient property without being responsible for the instantiation of one the base properties of the effect. In a sense there is no way to get to a supervenient property without messing with its subvenient base. My headache will not go away if there is no change in its neurological base, consequently I have to act on the neurological base to be free of the pain.

Now we get to the second leg of the argument; the incoherence of downward causation. Suppose a mental event *m* causes a physical event *p*. Was it in virtue of *m*'s mental property *M*, say being a headache, that *m* caused *p*, the event of you reaching for an aspirin? Now, *M* itself being a mental property must be dependent on a physical property *P* (this refers to the nonreductive physicalist commitments, but the completeness of physics is also a good reason to demand the existence of such property). So *P* must be a causally relevant property¹⁵ of *m* and it is right to say that *m* caused *p* in virtue of being *P*. So we get a situation where there are two causally relevant properties of *m*; both *P* and *M* seem each on its own sufficient to ground the causal relation between *m* and *p*. The worry now is that *P* will pre-empt *M* as the causally relevant property of *m*, leaving *M* dangling by the side with no causal work of its own left to do.

¹⁵ After all if there is a causal grounding relation that covers *M* and *P** as implied by the first leg of the argument (*P** is a property of *p*, in virtue of which *m* and *p* are causally related), there must be a law that covers *P* and *P**, since *P* is sufficient for *M* and *M* is sufficient for *P**. Kim (1998) says “We must now compare *M* and *P* in regard to their causal status with respect to *P**. When we reflect on this point, I believe, we begin to see reasons for taking *P* as preempting the claim of *M* as a cause of *P**. If you take causation as grounded in nomological sufficiency, *P* qualifies as a cause of *P**, for, since *P* is sufficient for *M* and *M* is sufficient for *P**, *P* is sufficient for *P**. If you choose to understand causation in terms of counterfactuals, again there is good reason to think that *P* qualifies: If *P* hadn't occurred *M* would not have occurred (we may assume, without prejudice that no alternative physical base of *M* would have been available on this occasion), and given that if *M* had not occurred *P** would not have occurred, we may reasonably conclude that if *P* had not occurred, *P** would not have either.”(p.43)

Why is there anything to worry about? Why can't both M and P be causally relevant properties of m? Well, to say that both are causally relevant properties of m's causing p, is to say that it was in virtue of both M and P that the causal relation obtained. Now, the issue is not about causal efficacy, like in the case where two distinct events overdetermine an effect. For example, two rocks hitting a window, where each of which is sufficient by itself for the breaking, are said to overdetermine the breaking because we believe that there is a doubling of the necessary causal powers needed for the breaking. In the mental/physical case, however there is token event identity, and so the efficacy of the mental event is not at issue, and it seems not to make sense to say that because we have two distinct properties that can ground the causal relation, somehow, the event has a doubling of the causal powers¹⁶. The problem is that if one property of m by itself can ground the causal relation (ex hypothesi) then we are entitled to ask what is the other property doing, since no more is needed for the causal relation to obtain. If M is there to ground the causal relation, what causal work is there left for P to do, and obviously we can put it the other way around by asking about the causal work that the mental property does when there is a physical property that can ground the causal relation. However now, enters the principle of completeness and we know that it must be the mental property that is left without work to do. We know that the instantiation of M is not a necessary causal factor for p, but due to the completeness of physics, some physical property is necessary, and in the present case the candidate for such a property is P.

A word about overdetermination and the work it is doing in the argument. The situation described seems to be that there are two sets of laws (or more precisely grounding relations), one of which connects M instances with P* instances (the relevant physical property of p) and another that connects P instances with P* instances. Now, this is a very tense situation because if in this case we could reach the conclusion that there is true overdetermination, then it would follow trivially that every case of mental causation is overdetermined. This is taken sometimes to be bad news for the nonreductivist as there is nothing else in the universe that reveals such extent of massive overdetermination. It is

¹⁶This however will be discussed with some detail in the 3rd section of this work.

clear that sometimes events happen that are overdetermined, as in the much-mentioned case of the man who dies because he was simultaneously struck by lightning and was shot at with the bullet crossing his heart. The case is of overdetermination since each cause was sufficient by itself and independent of each other for the effect. However it is clear that there is a case for saying that overdetermination, as that of the man killed by two distinct causes, is a product of coincidence and not that every case of a man's death is in such a way overdetermined. But in the case of mental causation we are talking about natural laws¹⁷ and saying that there is overdetermination would entail that those mental events would have been covered by laws twice, laws that are not reducible to each other since their properties are not so reducible.

Now there are strong reasons for thinking that such account is highly implausible on the grounds of the principle of economy. Why believe in such an account if one set of laws is perfectly sufficient. Maybe if there were no other ways of accounting for mental causation this possibility would be worth a second look. As always in such cases where the principle of economy is called for, its force is inversely proportional to the theoretical need of more complexity. So if we somehow could show that in fact positing such complexity entails even more difficulties we would be on good ground to suppose it unnecessary.

My stratagem in relation to overdetermination, or what might be said to be a better description of the case- redundancy- is to show that taking the overdetermination route to save mental causation, leads, on one hand to an impossible metaphysics or to a sense that the mental is not really there as a distinct entity. This will be the burden of section II and III.

Aside from various considerations of parsimony, Kim (1998) has charged nonreductive materialism with inconsistency if the overdetermination option is taken. He has pointed out, as will soon see, that if the nonreductivist insists on overdetermination then he can be charged of the violation of the principle of the causal closure of physics. Since in the case of true overdetermination the following counterfactual

¹⁷Once again it is more precise to speak of causal grounding relations.

seems to be true- that even if P (and no other physical property that could realize M was present) were not present M would have been sufficient to ground the causal relation. This would be a very strong and simple way to rule out the possibility of overdetermination. However as we will see there is much to say about this argument and we will go into it in more detail in the next section.

The Exclusion problem, by starting out with principle (1) assumes that mental events and their properties are suitable, by their very nature, to cause and be causally relevant to other mental or physical events. Whatever difficulties of reconciling anomalous mental properties with lawful physical properties, or how essentially external content properties might be causes, the problem of causal exclusion is there in the end of the road to bring to a halt their efforts. Even if the mental surpasses those difficulties it has to confront the possibility that there is nothing for it to cause- the mental is preempted by the physical.

I think that causal exclusion is a sound argument that shows that non-physical properties cannot in general be seen as causal. I will in the remaining of this paper defend it against various objections. I will be especially interested to show that when forcing the issue, the nonreductive materialist tends to move to a position that is best seen as a form of emergentism. But there completeness is violated. So he will try to give a metaphysical relation between the mental and the physical, such as realization, that justifies supervenience and tries to secure mental causation. It will be seen that there is no such magical relation.

Now lets move to a consideration of certain issues that surround overdetermination.

Section II

2.1 Introduction

The main points to be made in this section are: Kim's charge that taking the overdetermination route leads to a violation of completeness is incorrect; that Bennett's counterfactual test of overdetermination does not give us a correct understanding of overdetermination. Moreover a proper way to understand overdetermination is developed.

I want to consider some objections that have been raised against the exclusion problem that are based on the idea that the mental and the physical share a certain kind of intimacy that disallows the notion of overdetermination being set against them. That is, the objector to exclusion can say: there is nothing wrong with saying that mental properties are causal properties. It might seem, at first sight, that due to completeness we are countenancing overdetermined effects. After all in each case of mental causation there is a physical explanation that mentions a physical property that not only can do the job of grounding the causal relation but that is *necessary* for the obtaining of the causal relation. However the intimacy between the mental and the physical makes them not at all like most other cases of overdetermination, taking, so they claim, the spite out of it.

In its general form these arguments go like this:

Suppose that physical properties necessitate mental properties. If, for example mental properties strongly supervene on physical properties, where the supervenience relation is given a metaphysical reading, then such necessitating relation would hold. And that would mean that if M strongly supervenes on P, then whenever an object had P it would have M. Of course there would probably be many other creatures or systems where P would not be necessary for that creature or system to have M, because of multiple realization, but the point is that provided that P then it would be metaphysically necessary for M to belong to the object.

But why would such intimacy be good news for the nonreductive physicalist? Well, the point of such intimacy can be deployed in two related ways. First it could be said that if the mental and the physical are so close, then the charge of overdetermination is missing the mark because the closeness of the mental and the physical makes it fail our conception of overdetermination. Such conception, it is argued, depends in part on the deployment of certain counterfactuals, which are then seen to be false or vacuously true in case of such intimacy obtaining. Secondly, the notion that the overdetermination option is precluded for the nonreductive physicalist because it would lead to a violation of the principle of the completeness of physics, an argument that we have seen Kim employing in the last section, can be shown to rest in a mistaken

assessment of the counterfactuals. However the nonreductivist commitments are coherent, if only the worlds where we assess the counterfactual are worlds where strong supervenience holds. I will begin with the latter.

2.2 Overdetermination and Completeness. Looking for the right worlds

We mentioned before that if the nonreductive physicalist makes the overdetermination move, whereby he accepts that there are two distinct properties of the cause that are on their own sufficient to ground¹⁸ the causal relation, then it might be argued that he thereby is failing to be bound by his own rules- as he seems to be accepting the violation of the completeness of physics. Kim (1998) says that this

“Approach may come into conflict with the physical causal closure. For consider a world in which the physical cause does not occur and which in other respects is as much like our world as possible. The overdetermination approach says that in such a world, the mental cause causes a physical event – namely that the principle of causal closure of the physical domain no longer holds.” (p. 45)

The idea is that in the case of overdetermination, we would be in fact saying that the causal relation between m and p^* is grounded in two distinct and sufficient causal grounding relations, one that depends on mental property M and the other that depends on its base property P . Since in such a case each of them is sufficient by itself to ground the causal relation we would then be able to find a close world that would have a mental property M such that M would be enough to ground the causal relation between m and p^* , but such that in that world P would not be instantiated in m . So to causally explain p^* in that world we would have to mention M with necessity, and this is a clear violation of the principle of closure that, recall, says that the only properties that are necessary to explain a physical effect are physical properties. What one

¹⁸Remember that if a property F of c is causally relevant in the causation of e then there must exist a causal grounding relation between F and a property G of e . In the case we are considering there must exist two such causal grounding relations; one that related the mental and the other the physical.

is saying is that the counterfactual ¹⁹ $(M \& \sim P) \Box \rightarrow P^*$ ²⁰ is non-vacuously true. One is saying that there is a close world in which an object has M, does not have P and still behaves appropriately. But why think that the antecedent of the counterfactual is non-vacuous and why in such a case the consequent obtains and how come there is violation of completeness?

Now, in the nearby worlds where m is $M \& \sim P$, would m still cause p^* ? Presumably, since M by hypothesis is type connected with P^* of p^* in such a way as to ground the causal relation, it would seem that the fact that P is not available would not make much of a difference. After all the connection between M and P^* is independent of the physical law that exists in the actual world between P and P^* . It seems that even without P m would still have caused p^* . But then, to causally explain p^* one would have to refer, necessarily, to the fact that m is M , contradicting completeness. Kim (1998) says that

“I do not think we can accept this consequence: that a minimal counterfactual supposition like that can lead to a major change in the world.” (p.45)

Since the nonreductive physicalist would not be inclined at all to drop completeness, it would seem that he would have to give up the overdetermination move. If such point could stand it would be a swift and elegant way to push the nonreductivist out of scene. However things do not go so smoothly and I think that the nonreductivist can stand on his feet quite well.

Just consider that when looking for possible worlds where m is $M \& \sim P$ we have to keep in mind that we are looking for worlds that are close to the actual world. It is irrelevant to our quest what happens in worlds that are $M \& \sim P$ but are so different from ours that most of our physics is false there, for example.

¹⁹ Here I will follow much of the discussion in assessing these counterfactuals in the Lewis-Stalnaker semantics that says that $X \Box \rightarrow Y$ is true at a world w if, in all the worlds closest to w where X is true, Y is also true. Here closest is to be assessed in terms of similarity of natural laws and matching of facts.

²⁰ Of course, the correct way would be to write $(Mm \& \sim Pm) \Box \rightarrow P^*p^*$. But I will write as in the text as it is more convenient. P^* is the property with whom M and P are ex hypothesi type connected in such causal grounding relation.

This points to a simple idea: that the worlds that we should consider, are worlds that are $M \& \sim P$ but are similar enough to ours in that they respect completeness. We want to assess what would happen if a certain fact were different, but we want to keep all the other things as equal as possible. And if completeness is such an important principle for our understanding of science, such that the nonreductivist would not be able to motivate his position without it, then it seems plausible to look for worlds that respect it when assessing those counterfactuals. But then in those worlds there would have to be a physical property P' of m that would be mentioned in a basic law that would connect it with P^* of p^* , such that it would ground the causal relation.

But would this argument save the overdetermination move for the nonreductivist? I don't think it is likely. Though it will soon enough be seen that there is a way out it is not this way. Because now we have to ask if supervenience holds in those worlds. Presumably there will be some where it holds and some where it doesn't. But in those worlds where supervenience fails are worlds where the causal relevance of mental properties is even more difficult to grasp. In such a case we could have two individuals in different worlds where they could be physically as alike as you please and behave as similarly as you like while having totally dissimilar mental states. In fact they could even fail to have mentality. But in such a case it seems not to make sense to say that it was in virtue of such mental states that the same action issued from those individuals. Such explanation seems however to make sense by referring to the similar physical states. So the nonreductivist will want to assess the counterfactual in worlds that not only respect completeness but respect supervenience as well.

This move however might seem strange. After all certain factual differences might only be possible if certain laws or metaphysical constraints are different too. And so by keeping our evaluation of counterfactual in such tight condition we might be simply pointing out worlds that are logically possible but not nomologically or metaphysically possible. And it is reasonable to think that merely logically possible worlds are irrelevant for the assessment of the counterfactuals. In fact it might be thought that we already have given content to such a complain. That by supposing that $(M \& \sim P) \Box \rightarrow P^*$ is non-vacuously true we were in conflict with the completeness of

physics. But to be an objection to nonreductive materialism, this would have to say that all worlds that are like this world in respect of nonreductive materialism are inconsistent. But this does not seem to be the case at all.

After all, all close worlds *W* in respect to nonreductive materialism (that is assuming that nonreductive materialism is the reference world) that are *M* & \sim *P* are nonetheless worlds where the mental supervenes on the physical. That such worlds are *M* and not *P* does not contradict or in any way enter in conflict with supervenience provided that in any *w* of *W* *M* has a subvenient physical property *P'*, such that whenever an object has *P'* it has *M*. But then there will be no problem to account for completeness²¹, because if *P'* is metaphysically sufficient for *M* and *M* is causally sufficient for *P** then *P'* is sufficient for *P**²². In such a case there wouldn't be a violation of the completeness of physics.

Our preliminary conclusion then is that the nonreductive physicalist can accept the overdetermination move without thereby committing himself to an implausible rejection of completeness. In what follows I will support this conclusion, but by way of a better understanding of the uses of counterfactuals in causal contexts.

2.3 More on overdetermination and completeness. Reasserting the overdetermination move.

It seems to me that the way we assessed counterfactuals was not quite right, and that a proper understanding will make clearer the reasons for thinking that the nonreductive physicalist can make the overdetermination move without violating completeness.

I think that when one asks what would causally happen if *m* were *M* but not *P* one is asking what would have happened if *M* were present but nothing relevantly similar to *P* was there.

Compare:

If the rock had not hit the window, it wouldn't have shattered.

²¹ Crisp, T. M. & Warfield, T. A. (2001) make a similar point.

²² As we have seen before.

Now to assess this counterfactual we look for a world where the rock didn't hit the window but was in every other respect as similar to our world as possible. But now suppose that the reason a rock was thrown was that the house was on fire and people on the street thought it was a good idea to break the window to wake up possible sleepers inside. Then in the nearby worlds if that specific rock wasn't thrown another would have been. It could be that if that specific rock were removed in the closest world, the thrower would have picked up another suitable rock. The point is that he was looking for a rock and not for the specific rock that was actually thrown. And so the window would have broken anyway. This is bad news. It seems that unless we suppose that nothing was thrown that is sufficiently similar like the rock that was actually thrown, it is impossible to make use of such seemingly familiar counterfactuals. Moreover certainly such conception would be quite disastrous to counterfactual theories of causation.

Analogies only go so far and the point of this one is to bring out a feature of the assessment of counterfactual *in* causal contexts. Karen Bennett (2003) puts it quite starkly when she says that in such cases when one has to imagine something gone “you simply snip it away as though you had a metaphysical hole-puncher” (p.15 of the online pdf version). Of course in many contexts such reading is implausible if not downright incoherent. So if I inquire quite generally if there might be pain without c-fibres, it does not make sense to rule out all the physical bases of pain. Because what I am looking for is exactly if something else, some physical base could give rise to pain. So there is context sensitivity in the way to assess these counterfactuals. In a context of overdetermination, this context sensitivity is about the need to know what would happen if only one thing and not the other had occurred. So we have to look for worlds where a whole class of thing is metaphysically deleted in order to make sense of the idea that only *one other* thing can do the causing.

If we take this lesson to the supervenience case then it seems that to evaluate the counterfactual $(M \& \sim P) \square \rightarrow P^*$ we have not only to suppose P gone but we have to suppose that no other base of M is present in m. So in fact what we have to evaluate is the following counterfactual $(M \& \sim VP) \square \rightarrow P^*$ where VP is the disjunction of all the base properties of $(M \& \sim VP) \square \rightarrow$ says that none of them is instantiated in m.

It seems clear that a world in which $(M \& \sim VP) \square \rightarrow P^*$ is non-vacuously true is a world where there is a violation of completeness. But why should the nonreductivist worry about this? After all the “minimal counterfactual supposition” that Kim talks about means nothing less than the negation of supervenience. The only world in which the antecedent is true is a world where the mental floats unsupervised by the physical. But if the nonreductive physicalist is right in that our world is such as he describes, then the worlds in which the counterfactual is non-vacuously true are metaphysically impossible worlds. And why should he care about those?

So it is not even possible to inquire within nonreductive materialism what would causally have happened if the mental had occurred without the physical. Since this idea in an overdetermination context presupposes that the mental happened without any of its competition partners and nonreductive physicalism does not permit the violation of supervenience.

For someone who finds this reading of the counterfactual too strong he might fall back on the previous argument where it is shown that the nonreductivist can hold himself against the threat of inconsistency. However I think the last argument is an improvement and it shows that the reason that the nonreductivist can hold himself against the threat is that the only worlds where completeness fails are worlds where nonreductive materialism no longer holds. The counterfactual is only non-vacuously true in metaphysical impossible worlds.

But then it might seem that if they can claim that every case of mental causation is a case where the causal relation is doubly grounded the nonreductive materialist might be able to save a role for mentality after all. Of course they would be countenancing something that Kim and many others have thought to be especially bad; but, presumably, it would not be a charge of inconsistency that they would have to deal with as Kim (2003) as acknowledged.

Before moving on to consider the second argument, causal compatibilism, let's see where we are in the dialectic. One way to escape exclusion might be to bite the bullet and say that in spite of concerns regarding economy, one should embrace overdetermination – after all, in such a case, there is no inconsistency. Crisp and Warfield

(2001) even think that the nonreductivist should be quite happy to embrace overdetermination because, independently of the issue of supervenience, distinctness and completeness is enough to show that every case of mental causation is a case of overdetermination. After all if we suppose that M can ground the causal relation and we take account of completeness, then we know that there must be a physical property of m that is sufficient to ground the causal relation. Now, distinctness prevents the identification of mental and physical properties, and so the only seemingly option is to countenance overdetermination. So the nonreductive materialist “if they wish to hold on to the thesis that the mental qua mental interacts causally with the physical, then they’re committed to accepting overdetermination” (p.314).

So, it might seem that prevented from accusing the nonreductivist of inconsistency the exclusionist only resort to try to outmanoeuvre him is to stick to his guns and affirming loudly that overdetermination is not a good way to build a world. But then, his position seems much weaker and difficult to sustain in face of the nonreductivist mentioning the pervasiveness of mental causation discourse. It is very difficult to allocate points if the principle of economy is the only judge. There would be an impasse.

2.4 Causal Compatibilism and Bennett’s Test

Now however we can see that the second argument that I mentioned above can be thought to tip the balance in favour of the nonreductivist. It is the claim that what I have been calling the overdetermination move is not after all a move into overdetermination. That is, by accepting that the mental and the physical can ground causal claims when a mental event causes a physical event, we are not thereby claiming that the world is a prolific place in virtue of that. Whatever is the case it is not the bad case that we find in standard cases of overdetermination. This points to a position called “causal compatibilism” that is defended by Karen Bennett (2003). She wants to say that an effect can be caused by more than one sufficient causal property and yet not be overdetermined. She wants to do this while preserving the genuine causal relevance of the mental and its

distinctness from the physical. So she is saying that we can have two sufficient and distinct causal grounding relations for a certain causal relation, but they do not overdetermine.

That is, if it could be defended that the *form* of overdetermination that we find in the mental/physical case is not that bad, or that the two causal properties in spite of being sufficient and distinct do not overdetermine, then we would not have a problem of bad engineering.

But take a look at what our intuitive notion of overdetermination says. It seems, intuitively, that if we have an effect caused by two independent and sufficient causes, the effect is overdetermined. That is what the standard cases of overdetermination are saying. Consider the firing squad case or the fire in the house caused by a malfunctioning electric wire and a simultaneously strike of lightning - here we have two (at least) independent causal histories, each of which is sufficient by itself to cause the effect. So the causal compatibilist if she wants to be successful in her enterprise has to say why the mental/physical case is not in the same boat with the standard cases of overdetermination.

Of course there is a trivial way that they are different. After all in the standard case we have at least two different causal chains causally responsible for an effect whereas in the mental causation case we have only one causal chain. But the issue is about the redundancy of the causal properties. In that sense a mental property seems to be as redundant to the causation of behaviour.

In order to break the analogy between mental causation and standard overdetermination Bennett comes up with a necessary test for overdetermination that aims to include the standard examples while excluding, in an appropriate reading, the mental/physical case. I will present the test for the property case and for the event case as well, because there are some things that I want to discuss later to motivate a response to Bennett that is easier to understand if we first discuss it in the event case. The Test is given as follows²³:

In the events case:

p^* is overdetermined by m and p only if

(OI) if m had happened without p , p^* would still have happened:

$(m \ \& \ \sim p) \ \square \rightarrow p^*$, and

²³Bennett (2003), p.8 of the online pdf version.

(O2) If p had happened without m , p^* would still have happened:
 $(p \ \& \ \sim m) \ \square \rightarrow p^*$.

In the property case:

p^* is overdetermined by Mc and Pc only if

(O1p) $(Mc \ \& \ \sim Pc) \ \square \rightarrow p^*$, and

(O2p) $(Pc \ \& \ \sim Mc) \ \square \rightarrow p^*$.

Bennett justifies the test by pointing out that it captures the “reasoning we engage in when we want to distinguish cases of genuine overdetermination from cases of joint causation, or from cases in which one of the putative causes is not really a cause at all”²⁴. I will soon put this very claim into doubt, but first let’s see what can be done with the test if it were true.

She wants to use it to break the analogy of the mental/physical case with the standard cases of overdetermination. She wants to say that the mental/physical is in an appropriate sense not at all like the firing squad variety of overdetermination. A promising way of breaking the analogy is by defending that there is a tight connection between the mental/physical case that is not found in the standard varieties. And an appropriate connection between the two might affect the evaluation of the counterfactuals that make up the test.

Noticing that there are certain conditions for the application of the test Bennett thinks that for overdetermination to obtain both counterfactuals have to be non-vacuously true: For example both counterfactuals cannot be vacuous. If m and p were identical then both of the counterfactuals would come out vacuously true, but no one would say that p^* was overdetermined. Similarly, if one of the counterfactuals was false then there was a case for saying that only one of the “causes” really is causally effective.

This condition suggests two ways for the mental/physical case to fail the test - showing that one of the counterfactuals is false or that it is vacuous.

Both for tactical and strategic reasons Bennett concentrates her efforts (O2)/(O2p). On one hand she is not sure that the reasoning she applies to (O2)/(O2p) would work on (O1)/(O1p), and on the other she thinks it is strategically advisable for the compatibilist to suggest that

²⁴ Bennett (2003), p.8 of the online pdf version.

somehow the physical needs the mental, than the reverse, because the risk of taking the mental as epiphenomenal or only derivable efficacious is always present and tempting, even for the compatibilist.

One way to claim that (O2)/(O2p) is vacuous is to say that it is metaphysically impossible for p to happen without m or that c cannot be P without also being M . This will be the case if p necessitates m or in the property case if P necessitates M . Now in the event case we know that there is no claim to overdetermination because of event identity. In such a case both counterfactuals are vacuously true, which answers directly to our intuitions that in such a case there is no issue of overdetermination. In the property case if one takes strong supervenience with a modal strength of metaphysical necessity then there would never be a metaphysical possible world where P and not M , making the counterfactual $(Pc \ \& \ \sim Mc) \ \Box \rightarrow p^*$ vacuously true.

So we would have enough reason to think that in such a case the mental and the physical do not pass the test and consequently are not overdetermining. There are some intricacies of Bennett's argument that I am not considering here fully, like her claim that P as a physical property that normally is given in exclusionary claims is probably not necessarily sufficient for M . The point that Bennett is making is that the properties that normally are subjected to exclusionary problems are not as they stand sufficient to bring about the mental. That is clear in the functionalist case, where he has to pack into the physical property a reference to the causal relations it enters into without which the second order mental property is not realized. Nonetheless she thinks that there are such physical properties that necessitate the mental

“Even though the physical property that we initially fixed on does not necessitate the mental one, there presumably is a richer physical property that *does*.” (Bennett (2003), p.20 of the online pdf version)

I take this to be a point about circumstances. A physical property realizes a mental one only in certain circumstances and where those circumstances obtain then P necessitates M . Striking a match is sufficient to make fire *in* the circumstances. If there isn't oxygen present then there will be no fire. Strictly speaking, it might be said that most if not all of the causes that we cite in explanation are actually insufficient for their effects because they always need those circumstances to obtain. A way to circumvent this difficulty is to pack

in the cause, not only the usual suspects but all the extra factors that are needed for the effect to occur, i.e., whatever it might be said to constitute the circumstance. In a somewhat similar fashion, Bennett thinks we should build on P, by bringing in whatever it takes to make the necessitation claim to hold. However it doesn't seem to me that such strictures are necessary. We do not have to amend the way we have been talking about base or realizing properties and their realized or supervenient properties, provided that we keep in mind that the causes we cite are efficacious only inside a web of existing causal relations. I take it then that there is reason to accept that (O2p) is vacuously true and consequently the mental/physical case fails Bennett's test.

But more can be said about the test, especially about the other counterfactual. As we have already seen in the previous discussion about the claim of the violation of completeness, there are reasons to suppose that (O1p) $(Mc \ \& \ \sim Pc) \ \Box \rightarrow \ p^*$ is always vacuously true if metaphysical supervenience is upheld. And so it is seen that mental causation is not like the standard cases of overdetermination. Supposedly that should have been clear all along; after all the reason that is responsible for it to fail the test is the metaphysical dependence of the mental on the physical.

Such existential dependence is not commonly found in standard varieties of overdetermination. But is such existential dependence an issue when it comes to the idea of redundancy and of bad metaphysical construction? I don't think so and will try to motivate the claim in what follows. It seems to me that Bennett is right that we use something like the test when counterfactually thinking, but that has to do with the fact that it can have practical utility without being metaphysically deep for our concerns, and consequently fails when the going gets tough.

Bennett's test handles very well standard cases of overdetermination. Two shooters independently of each other kill a man by hitting his heart simultaneously. Or take the case of two rocks that break a window at the same time. Both cases pass the test; if one shooter hadn't shot the other would still have made the kill; if one rock wouldn't had been thrown then the other would have broken the window. Now, in these examples we are supposing that each one of the causes was sufficient in the circumstances for the effect; the death or the breaking.

Now what does it mean, in a context of overdetermination, to say that they were each sufficient for the effect in the circumstances? It has to mean at least that the other cause cannot be part of the circumstance of this cause. That is, if c_1 and c_2 are said to overdetermine e , and are each sufficient in the circumstances for the cause, then the circumstances that permit c_1 to cause e cannot include c_2 , and vice versa. This is obvious quite apart from the claim that otherwise c_1 and c_2 wouldn't pass the test. After all we wouldn't say that the striking of the match and the oxygen are competing for the causation of the fire.

In the mental/physical case, we also suppose that the mental and the physical are sufficient in the circumstances for the effect e . Now Bennett intends the test as a way of breaking the distinction between an appropriate reading of the mental/physical case and the standard examples just mentioned – her test marks the distinction. Of course one might quibble here and, as was mentioned before, point out that since there are two sufficient causes, then there is a problem. But the response that she intends for this quibble is that the mental/physical case is different in such a way that even if we still say that they are overdetermining, it is not in a bad kind of way. Why is that? I suppose it is because a very tight connection, as she proposed, somehow prevents us thinking that there is doubling of causal factors in the causal relation.

Let suppose, with plausibility, that this is what is particularly objectionable about overdetermination. Now there are two objections that need to be mentioned, only the last one will I pursue here with detail. The first one is that in such a case, where there is no doubling of causal factors, it is not at all clear, having in account completeness, how exactly are we conceiving the initial starting point (to establishing exclusion) that the mental is *causally* sufficient for the effect and is distinct from the physical. There is a lot to explore here and it will pursue this issue later in section 3. The second comes in the form of counter-examples to the test.

(Continua)