Nonreductive Physicalism and Mental Causation

By

João Miguel Rodrigues Oliveira Amen

Mphil Thesis
Department of Philosophy
King's college London

March 2005

Supervisors:

Prof. David Papineau
Prof. Jim Hopkins

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.

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.

Table of Contents

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 mentioning[1] 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

[1] Assuming, throughout, causal realism.

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 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 event[2] 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 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

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

physicalism, there is no mental causation problem about the causal efficacy of mental events.

Still, soon after Davidson published his theory, a torrent of papers[3] 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.

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,

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

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 law[4]. 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 itself[5]:

(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 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

[4] McLaughlin (1989) defends extensively the argument being made.
[5] I will be more through about this in the beginning of section 1.7

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 <u>not</u> 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

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

(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. 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 Horgan[7] (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

> 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)

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

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 out 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

> Since 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 <u>via</u> 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 <u>vis-à-vis</u> 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 particular[8].

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.

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

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

> 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 event 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 he 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 science[9]. 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

[9] This 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.

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 metal 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 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 pyramidical 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 (=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.

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

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 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 Kim[11], if mental to physical causation is also possible[12]. 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

[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.

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 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 it 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

> 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 property[15] of m and it is right to say that m

[13] Kim (1998)
[14] Kim (1993b)
[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

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.

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 powers[16]. 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*

---

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)

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

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 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 laws[17] 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 way 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

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

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 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 ground[18] 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

[18] Remember 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.

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 is saying is that the counterfactual[19] (M&~P) $\rightarrow$ 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 & ~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 scene. However things do not go so smoothly and I think that the nonreductivist can stand on his feet quite well.

---

[19] Here I will follow much of the discussion in assessing these counterfactuals in the Lewis-Stalnaker semantics that says that X $\rightarrow$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 term of similarity of natural laws and matching of facts.

[20] Of course, the correct way would be to write (Mm&~Pm) $\rightarrow$ 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.

Just consider that when looking for possible worlds where m is M&~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&~P but are so different from ours that most of our physics is false there, for example.
This points to a simple idea: that the worlds that we should consider, are worlds that are M&~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 a 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&~P) $\rightarrow$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&~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 completeness[21], 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:

[21] Crisp, T. M. & Warfield, T. A. (2001) make a similar point.
[22] As we have seen before.

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

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 $(M\&\sim P) \rightarrow P^*$ 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 $(M\&\sim\vee P) \rightarrow P^*$ where $\vee P$ is the disjunction of all the base properties of M and $\sim\vee P$ says that none of them is instantiated in m.

It seems clear that a world in which (M&~$\vee$P) $\rightarrow$ 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 follows[23]:

In the events case:

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

p* is overdetermined by m and p only if

(O1) if m had happened without p, p* would still have happened: (m & ~p) $\rightarrow$ p*, and

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

In the property case:

p* is overdetermined by Mc and Pc only if

(O1p) (Mc & ~Pc) $\rightarrow$ p*, and

(O2p) (Pc & ~Mc) $\rightarrow$ 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 came 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.

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

Both for tactical and strategic reasons Bennett concentrates her efforts (O2)/(O2p). On one the 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 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 & ~Mc) $\rightarrow$ 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) $(Mc \ \& \ \sim Pc) \rightarrow p^*$ 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 c1 and c2 are said to overdetermine e, and are each sufficient in the circumstances for the cause, then the circumstances that permit c1 to cause e cannot include c2, and vice versa. This is obvious quite apart from the claim that otherwise c1 and c2 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 is 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 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.

2.5 A new way to Understand Overdetermination

I want now to consider what the correct way to understand overdetermination could tell us about the exclusion problem. But what is a better understanding of overdetermination? I think the following example, to be understood as a counterexample to Bennett's test, can lead the way.

Two rocks are thrown and break a window. It is clear that the event is overdetermined; it's a standard case of overdetermination if anything is, and it passes the test; both counterfactuals came out true.
Now, suppose that we interpose, between the place where the rocks were thrown and the original window, a second window with such a resistance to impact, that if the two rocks are thrown they still break the original window (and each has more than the required energy to break the original window independently, that is, each is still *causally sufficient* for the breaking), but if only one is thrown then the original window fails to break. It might be that the breaking of the interposed window makes the momentum of the single rock to change in such a way that when it hits the original window it doesn't have the necessary energy for the breaking. It seems to me that this example, while being one of overdetermination, fails Bennett's test. After all, if one of the rocks were thrown without the other, then it would fail to break the window. In such a case, both counterfactuals would be false.

This example depends on a factual matter of this world. But the idea can be expanded to nomological worlds with a different example. Suppose that the world is such that when one elementary particle of kind z hits certain types of atoms A, it will not make a difference to A, but that when 2 or more hit A it causes A to disintegrate. Moreover suppose that the reason that A disintegrates is because of certain impact between z and A's core. Now the reason that z alone cannot disintegrate A is because it cannot pass A's exterior protecting field (two or more z-particles are needed). But once inside the field it has the energy to disintegrate A on its own. Now this case seems to be another example of overdetermination that does not pass the test. If only one

particle hits A , A will not disintegrate. And this extends to all nomological possible worlds.

Because, in these examples, both counterfactuals come out false, should we think that events c1 and c2 are necessary to cause e? Are c1 and c2 partial causes of the effect e? That is the reading that Bennett would probably make of this fact, and it seems a plausible one, if we have a view of overdetermination that implies that both of the causes have to be sufficient[25] for the effect e. That is a view that is implied in the counterfactual test and is probably universally accepted. Moreover there seems to be some features of this example that might be picked up to argue that this is an example of joint causation.
After all, for the causal chain that begins at c1 to break the original window it needs the help of c2, and vice versa. And to need help in a causal context is to be part of the circumstances of the cause; c2 has to be one of the circumstances of c1 and vice versa. But as we have seen, when talking of sufficient causes for a certain effect in an overdetermination context the other cause cannot be part of the circumstances of each other. And in this example I want to claim that c2 is not part of the circumstances of c1 and vice versa. So one has to accept that they are not causally sufficient for e. (if one focuses on c1 and c2.)

However, I still think that it is clear that the case of the two rocks with interposed window and the case of z and A, as it happened in the *actual* world is a case of overdetermination. I will go on later to argue that the fact that c1 and c2 might not be sufficient causes is not important. So forgetting for now, this difference, does the example fail to pass the test? There are certain ways that are wrong ways to assess counterfactuals, as when we make use of backtracking counterfactuals.

Is this case an example of backtracking? Backtracking happens when one does a reasoning of the following form: "if c1 had not happened, that must have been because x happened, and if x had happened, c2 would have happened in such a way that it would have failed to cause e" (Bennett (2003),p.9 of the online pdf version). To illustrate this point take the following counterfactual situation about a firing squad: "

[25] As we will see later on, they have to be sufficient, but this needs qualification: they have to be sufficient at the moment of *impact.*

suppose that the first gunman is quite serious about his work, and would only fail to fire his gun if some terribly traumatic event occurred just before he was to do so- the sudden collapse of a beloved commanding officer, for example. But that kind of event would leave the second gunman shaken up as well, and would throw off her aim. Consequently, it looks as though the victim would not have died if the second gunman had fired without the first – the second gunman would have missed." (Bennett (2003),p.9 of the online pdf version)

Ruling out backtracking is ruling out counterfactual dependence of the past on the present. The sudden collapse of the beloved one is counterfactually dependent on the fact that the first gunman failed to fire his gun. But that just seems wrong in causal contexts; after all the past is fixed and accordingly whatever happens in the present cannot alter it. Moreover the reasoning takes the following form:

If ($\sim c1 \rightarrow X$ and $X \rightarrow (c2 \rightarrow \sim e)$[26]) then $(c2 \,\&\, \sim c1) \rightarrow \sim e$.

That is, we go back and forth between the present and past, to find out what would have happened in a certain counterfactual situation that makes reference to the future. But what will happen, and in particular the finding out if a certain effect will occur, should be ascertain by the present state of things; it shouldn't depend on a counterfactual that fixes the past depending on certain affairs of the present and then assesses the future based on which facts about the present were changed by the fact that certain aspects of the present fixed the past.

The case of the two rocks with interposing window is quite different. There is no backtracking, since we are not going into the past to inquire why one of the "causes" didn't actually happen. We just assume, rightly, it gone. However, it is clear from this example that the other "cause" doesn't quite make it to cause e; it alone when enters into contact with the original window does not have the necessary energy to break it. Moreover the fact that e does not happen has nothing to do with facts anterior to any of the causes c1 and c2, but with the causal structure of the world. A similar reasoning applies to the z and A case.

[26] Or instead of $(c2 \rightarrow \sim e)$ it is possible that "c2 wouldn't have caused e" is better way to express it.

But are we concentrating on the wrong events? Should the relevant events for the evaluation of the counterfactuals be d1 and d2, the breaking of the interposing window by c1 and c2 respectively? It seems to me that the move is unwarranted and unprincipled, but still there is no difference whatsoever in the evaluation of the counterfactuals in such a case. If d1 happens without d2 it will not break the original window and vice versa for d1. And if both of them occur, then the window will be broken with much more causal *oomph* than it's needed.

The only other objection that I notice is to say that d1 when it happens without d2 is a totally different event because it has considerable less momentum. But this has to depend on a very fragile view of events that I don't share. It seems that when we are looking for the counterpart event d1 in a world in which c2 does not happened, our relevant event is the breaking of the interposing windows caused by the throwing of the rock c1. We are still referring to the same event even if in that world the counterpart of d1 has less momentum. It is the same because it is still the breaking of the interposing window by c1. (Moreover, it seems that accepting such view of events would make impossible to talk about possible worlds in causal contexts)

There is an objection that needs to be addressed at this point that might have been for a while in the mind of some readers. It is the thought that one cannot get rid of a sense in which each of the causes is needed for the other in order for the breaking to occur. I think that this is a legitimate worry that needs to be addressed here and placed in the context of this thesis. After all, if c1 didn't occur then c2 wouldn't have been able to break the window. The point is that one cannot get rid of a sense in which c1 and c2 are each necessary, and thus each part of the total causal factors needed for the breaking. What is being said is that because of the interposed window, or in the case of atom A, because of the surrounding field, the event e can only occur if all of those factors obtain. And so it might be argued that our example is missing the point because it is after all at best an example of joint causation.

I think that this objection can be met. It seems to me that such an argument would make it impossible to speak of overdetermination at all. Since if we go back enough we can always find that there is a connection between the causes that prima facie we thought were overdetermining. The causal world, after all, is extremely

interconnected, and there is reason to suppose that the two causes that we focus on as being overdetermining might depend on a causal factor antecedent that relate causally both histories. But in normal parlance, and in particular when one countenances cases of overdetermination, one is talking of synchronic causal circumstances and their relevance to the effect, and when in the actual world, each rock hits the window they are not synchronic circumstances of each other!
In the case of the oxygen and the striking of the match, we know that each is a synchronic circumstance the obtaining of which is needed in order for the lightning to occur. Consequently it wouldn't be correct to suppose that causal competition happens, but for events that are not synchronic circumstances of each other the problem of causal competition raises its head once again.

I think, however, that it is important to make a note to those of you who remain unconvinced and still think that this is at best a case of joint causation because even then the main point in this thesis will stand. Because for this to be a genuine case of joint causation the mental has to make a causal contribution to the causal relation in each case of mental causation. However as we will see in section 3, this contribution will have to come either from emergent properties or from reduced properties, in which case one wouldn't obtain an acceptable solution to the problems that the nonreductive materialist faces.
Still, I register this possible disputation about the interpretation of overdetermination and my tentative contribution to its definition, however since it does not affect the main points that follow, I will not mention it again. I will then continue in my attempt at a better understanding of overdetermination in the hope of changing some minds.

Now I said that even though c1 and c2 weren't sufficient for e they still were overdetermining e. I think that it is here that some intuitions about overdetermination show their face and tell a somewhat different story that Bennett would want us to believe. I have a very strong intuition that tells me that the reason we say that the breaking of the original window was overdetermined was because there were two causal histories that when entering into causal "contact" with the window, each one of them had the necessary causal power for the breaking. And here we can express the sufficiency condition of each one of them because none needs the other as a causal

background condition. This sufficiency condition is expressed, in relation to the actual world, as it happened, and in the last instance, just before or in contact.

Let me express better this idea of overdetermination. I will suppose that it makes sense to say that in the causal history from an event like c1 or c2 to another like e, there is one last event of such causal history that is the one that makes causal "contact" with e, lets call it an ultimate event. Now, in my view, to say that c1 and c2 overdetermine e, is to say that they cause u1 and u2, respectively the last event of the causal chain that goes from c1 to e and from c2 to e. And that each of u1 and u2 had the necessary causal means to break the window on its own. Notice this idea will fit perfectly into the normal case of overdetermination, and also in the case of the two rocks with interposing window.

We can express those conditions for overdetermination. An event e is overdetermined by c1 and c2 only if

(1) c1 is an ultimate event or is a cause of an ultimate event u1
(2) c2 is an ultimate event or is a cause of an ultimate event u2
(3) c1 and c2 are actual causes of e
(4) c1≠c2 and u1≠u2
(5) u1 is a sufficient cause of e
(6) u2 is a sufficient cause of e

Here we can see that both the two rocks with interposing window and the A and z case both meet the requirements of overdetermining causes. Notice that condition (3) rules out possible cases of preemption and (5) and (6) thwart the possibility of joint causation.

We can also see why such overdetermination might be thought to be bad, since it keeps certain features of the standard cases that we saw before, when it is obtained in lawful or metaphysical terms in such a way that all cases of a certain kind of events

meet the condition 1 to 6. After all we get a strong sense that e has more than it is needed, and this more is there with at least nomological necessity in certain cases.

Now to apply all this to the case of overdetermination (or redundancy) by causal relevant properties, lets imagine that e' is an ultimate event, and U1 and U2 are properties of e'. Let P1 and P2 be properties of c such that they give two distinct causal grounding relations between P1 and P1* and P2 and P2*, where P1* and P2* are properties of e. Moreover since e' is an ultimate event lets suppose that P1 and U1, P2 and U2, U1 and P1* and U2 and P2* establish all causal grounding relations between those events. U1 and U2 are ultimate causal grounding properties, that is, the properties of ultimate events that are in such causal grounding relation with properties of the event caused.

Now we can express those conditions for overdetermination in the property case. An event e is overdetermined by the properties P1 and P2 of its cause c only if

(1) P1 is an ultimate causal relevant property or is in a causal grounding relation with an ultimate property U1
(2) P2 is an ultimate causal relevant property or is in a causal grounding relation with an ultimate property U2
(3) P1 and P2 are actual causal relevant properties of e
(4) P1≠P2 and U1≠U2
(5) U1 is a causal grounding property of the causal relation between e' and e
(6) U2 is a causal grounding property of the causal relation between e' and e

I think that this notion of overdetermination captures all the kinds that are normally talked about, but goes deeper by picking on what are in reality the specificity of overdetermination making the notion more explicitly and more easy to understand. As such we can see that certain causal events that might be otherwise difficult to assess in relation to issues of overdetermination become easier to assess because we now know better. And we know better because we know that the issue of overdetermination *is an issue of matters that have to do with the ontology of the situation and what in terms of causation is going on in the actual world*. As such if there is anything wrong with

overdetermination it has to do with the picture that it presents us with. Is it a coherent one? Can we make sense of such ontology, where two different things are related somehow, say by some metaphysical form of dependency, and still have different causal powers as to establish overdetermination? Of course, and this is the point, it depends on how thing are in the case under consideration. The present point is that issues of overdetermination should be assessed in terms of the ontological commitments of a theory. Can it make sense of the picture given above? I will argue latter that in the mental/physical case it does not seems likely that a coherent picture of overdetermination can be made to save the nonreductive materialist from the exclusion argument. That is, the ontological commitments of nonreductive physicalism do not permit us to say that both properties are actual causal properties of an event.

Does this way of thinking about overdetermination contradict somehow the kind of counterfactual reasoning that Bennett says that we normally make in such cases? Well, it sure does even if instead of concentrating on the usual kind of events like c1 and c2 we concentrate on ultimate events. After all the same reasoning about necessitation claims that would fail to pass Bennett's test are also easily applicable to ultimate events.

Is there any deep lesson to be learned for this failure? Not at all, because it is usual that the common sense understanding of things have to be understood anew or in a slightly difference way when more is demanded of it than is usually found in common practice. After all, counterfactual thinking fits the bill in most of the cases. It is a practical way of doing it, without being metaphysically deep. And when we get metaphysical it sometimes fails.

But it has implications for our understanding of exclusion arguments. A trivial one is that it refutes Bennett thinking. It is irrelevant to questions of overdetermination, and the worries that it brings with it, to know what would have happen in a counterfactual situation. The metaphysics of the situation, the metaphysics that our theories imply, should be our guide and should be enough.

It seems to me that it is right to say that the objector cannot defend causal compatibilism by diminishing the issue of overdetermination. By all means it seems, at first sight, that in such a case, as the mental/physical one, there is

overdetermination. However I think that the superficial look can be shown to be misleading. In the next section we will see that in the metaphysics of the nonreductive physicalist, the relation between mind and body is quite unstable and in light of mental causation difficult to make sense of. It seems to me that he cannot account for overdetermination.

Section III

3.1 Introduction

In section I we have seen that nonreductive physicalism faces the problem of exclusion. This problem threatens to preempt the mental of any causal relevance in the causation of behaviour. But the form of the argument gives a clue as how the nonreductivist might escape and save the causal potency of the mental – countenancing overdetermination. In such a case he would elude the argument made at the end of section I. In section II we have seen that he can do this only if his metaphysical commitments are consistent with overdetermination. In this section it is shown that he cannot take the overdetermination route to escape exclusion. Countenancing overdetermination brings his position close to either emergentism or reductionism in a way that ends up being inconsistent with his own position.

I think the uneasiness that is frequently seen in discussions of mental causation regarding the issue of overdetermination is in fact best seen as a tension, mentioned in section I and defended further in section II, within the ontological claims of nonreductive physicalism. On the one hand it is committed to the distinctness of mental and physical properties and their causal relevance - raising the prospects of overdetermination. On the other hand it finds the prospects of a mental realm having an ontological significance on its own, which seems to be a consequence of the postulation of mental causal properties, quite puzzling and threatening to a good physicalist. Anxious for his position not to be confused to any form of significant dualism, he will immediately point out not only that mental properties are properties of physical objects and events, but also that the mental is metaphysically dependent on the physical. But then such intimacy between the mental and the physical is seen as a puzzling feature of an overdetermination claim.

We can now say what the puzzling feature is, or where it resides, following our thoughts in section II. What the tension is pointing to, or why the overdetermination claim brings such complains, is that it is a symptom of something going wrong at the ontological level.

It is curious that when one follows the literature that attempts to solve the problem of mental causation one finds it common the complacent assumption that the position they try to save, namely nonreductive physicalism, is well defined and that what has to be done in order to solve the problem of mental causation is just to tinker within the position with the assumptions it makes and find a way to accommodate the causal relevance of mental properties. However the troubles for the nonreductive physicalist go deeper than mental causation. The trouble of accounting for mental causation is just a symptom of more fundamental and constitutive difficulties that besiege the nonreductivist position. The trouble with overdetermination brings those difficulties to the fore.

The problem, I think, lies in that the notion of dependence, in particular the fundamental notion of realization, is made on the base of an intuition and is developed in the context of an ambition - not very deep ontological concepts, to say the least. The intuition is based on the multiple realization argument and the ambition is the desire that psychology could be investigated independently of its physical substrata. This ambition is expressed starkly in the hope of the computational theory of mind that the appropriate implementation of a computer program would have genuine mental properties just as we do, making the study of the human mind, moreover, the study of psychology, independent of the study of the human brain and as such quite a general and independent science on its own. I think that this view colours the interpretation that the nonreductivist makes of the multiple realization argument and that leads to the development of his position in a way that is incoherent or implausible. Of course, the nonreductivist would say that the reason he thinks that psychology could be an autonomous science is just a consequence of the multiple realization argument.

However, it seems to me that this is a misguided reading; nonreductive physicalism is a position that is inherently incomplete and I think incompletable since cannot account for the notion of realization. He can advance some as we will see, but it does not respect all his commitments, either by being inconsistent or highly implausible. And the reason is due to its ontological framework that is directly dependent on its interpretation of the multiple realization argument. The lack of consideration to ontological details, that spring from the ambition that psychology can be studied

independently, is the source of the trouble. Once we consider the details, they do not add up to the intentions.

The notion of realization by grounding the metaphysical relation between the physical and the mental give us the target to investigate if the claim of overdetermination can be sustained. If the notion of realization is coherent with the nonreductivist position regarding the causal relevance of both mental and physical properties, then premise (4) of the exclusion arguments given in section I, which rule out overdetermination, could not be maintained, bringing down this form of arguments against nonreductive physicalism. It will be seen in this section that the nonreductivist cannot claim the overdetermination move, because his metaphysical commitments do not allow it. In what follows I intend to show this in detail.

We saw that the worry about overdetermination was particularly troubling in the case of downward causation. Lets take a look at the ontological details.

So we have something like this: m causes m* in virtue of M and in virtue of P, both properties of m that can with sufficiency ground each on its own the causal relation between m and m*. Downward causation is the claim that M can ground the causal relation only if type connects appropriately with P* of m*. Now the type connections are supposedly distinct; after all the type connection between P and P* is given by a physical law, and whatever type connection exist between M and P* cannot be such a law. M is a bona fide mental property after all.

Now, as we have seen in section I it is plausible to hold that causal grounding relations specify the causal powers that the properties endow the object or events with. It might even be right to say that it is because of those causal powers that the causal relation obtains and because of those powers that the laws or other type relations are what they are. We do not have however to take sides regarding this matter. It might be instead that those properties *plus* those causal-grounding relations endow the objects with those causal powers. Either way, once we have the grounding relation we know that those objects or events have certain causal powers in virtue of the properties that the object or event has. Its seems then reasonable to suppose that different properties, even supposing that they are more than their causal powers, have

to endow different causal powers to the objects possessing them, even if due to certain circumstances those powers go without manifestation.

Now let the causal powers of the mental property[27] M be Mcp1, Mcp2, Mcp3, etc and let the causal powers of the physical property be Pcp1, Pcp2 etc. Now to say that both M and P overdetermine some of the effects of m, specifically the capacity of m interacting causally with m*, is to say that each causal property on its own can ground the causal relation, which means that the causal powers that each of the properties bestows on m are sufficient on their own to determine the causal relations that m has with m*. But this cannot be the whole story. That the causal powers of M can determine m's ability to causally interact with m* cannot rule out the ability of those of P to do their job. Not without giving rise to a conflict with completeness.

So every time we have mental causation the picture is that both sets of causal powers, the causal powers that m has in virtue of being M and the set of causal powers that m has in virtue of being P get to do their causal work. Moreover since this is a case of mental causation they are sufficient on their own for effect m* to be P*.

But we know, and the nonreductive physicalist will make sure we remember, the mental is dependent on the physical. This dependence tells us that m is M because m is P. And that if m were not appropriately P-like, m would not be M. This implies that if m were not P or P-like then m would not have the causal powers that we supposed M to endow m with. It is here that we want to know about the relation between M and P and their respective causal powers. Since such dependence relation is given more fully in terms of the notion of realization I will follow those proposals, however we can see the options open to us can be expressed in more general terms. Either (A) the mental and the physical have a totally different set of causal powers, that is, the set of causal powers of the mental and the physical do not intersect or (B) they have some causal powers in common, but the physical and the mental have powers that go beyond what they have in common, or (C) the mental while sharing some causal powers with the physical surpasses it by having causal powers that are not found in the physical property, or (D) the physical while sharing some causal powers with the mental surpasses it by having causal powers that are not found in the mental property, or (E) they have the same causal powers.

[27] Of course this means the causal powers that the property bestows on the objects in virtue of the object having the property.

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

I will begin by analysing (A), (B) and (C) since they have a problematic feature in common: they all presuppose that there are causal powers that the mental has that are not found in the physical realizing property.

Randolph Clarke (1999) suggests a way of formulating the realization relation that gives substance to (C) and since we will find that position problematic on account of this feature that it shares with (A) and (B) it is a good place to start. Consider, says Clarke, that when a mental event m causes another mental event m*, as when pain causes the desire that the pain goes away, there will be some physical realization base PH* for M* in humans another PA* for aliens etc for all kinds of systems and creatures, where M* is the desire that pain goes away. Now take in consideration that this means, due to downward causation, that mental property M has the causal power to interact causally with one set of physical properties PH* in humans another in PA* aliens etc for all the different systems.

After all M's causal powers, or the causal powers that M endows the events or objects that have it with are always the same. How could the very same property at one time or in an object endow it with a set of causal powers and at another time or in a different object endow it with a different one? It does not seem plausible. Either view of properties and their causal powers, seems to support such analysis. If one thinks that properties have the causal powers they have, primitively, due to their nature, a view that I favour, then the same property will always endow the objects or events with the same set of causal powers. Or if you think that properties have the causal powers they have due in part to their categorical nature plus the laws of nature then it seems that the very same property will always endow their objects or events with the same set of causal powers in nomological possible worlds.

Not to confuse this with the view that all of those causal powers are active every time an object possesses a property. A vase does not break because it possesses the property of fragility. This property can endow a vase with a set of causal powers that remain without expression till certain conditions are fulfilled. Only then their causal work is done.

This means that M is such as to be able to establish a causal grounding relation with each of M*'s heterogeneous physical bases. M has such set of causal powers that are

sufficient to interact with each of M*'s realization bases. However it is not plausible to suppose that this causal power is found on any of M's realization bases. Take PH the physical property that realizes M in humans. Now this physical property will have the power to causally interact with the physical base PH* of M* (the desire that the pain go away) in humans. But it is highly unlikely that PH has powers to causally interact with anything like PA*, that is, the physical realizer of M* in aliens. Clarke's proposal then seems to be based on the idea that the causal powers of the mental exceed those of their physical bases.

This view has some strange consequences. It seems to imply that when a human has the mental property of being in pain then, it has the power not only to causally interact with PH* but also PA* and any other realizing property of M*. But how can this be? PH, M's realizing physical property in humans is some complex structure of neurons, and this structure, realizes M in humans because, lets say, it is causally embedded in the human organism in a way that occupies the causal role of pain. One of the consequences of this causal role will be the bringing about PH*, that is, the realization base of the desire that the pain goes away in Humans. But there seems to be no way to make sense that PH can also bring about PA*. Just consider that PH is embedded in a system composed by neuro-physical properties, carbon based, while the alien system might be extremely different from the human, so different as being chemically incompatible. So M has some causal powers that none of its particular realizers has, though in the disjunction of the realizers one can find every type of causal powers that M has. It seems very difficult to make sense of this hypothesis.

Clarke(1999) says that

> When an object possess a property that carries with it the power to cause, in certain circumstances, a certain effect, that object may lack the power to cause, in the present circumstances, that effect.(p.301)

So we might say that while a human has M, due to the circumstances M can only bring about PH* and not PA* or any other alien base. But compare this with the example given above, about the vase and its fragility. If the vase is fragile and it is in a very safe and stable position it will not break, due to these circumstances. But the vase has, due to having the property of fragility the causal power to break even if it does not manifest such potential. But the question then is how can a human, by having

M, have the potential to bring about PA and other physical effects that are extraneous to its chemistry, even if they go without expression.

Clarke suggests a way to make sense of this idea – to understand how M has the powers to bring about all the different realizers of M*, we have to suppose that in any realization of M in a particular organism, say a human, all the other realizers have to be present somehow in that organism, but that somehow they are in such a way embedded in the organism that the causal powers idiosyncratic to other realizers except the human are somehow inactive. This, at first sight, might seem to be a way to understanding why M has more causal powers than the powers of the realizing property.

There is however some implausibility to this suggestion to say the least, and one difficulty that jumps to the eye is that it is not clear that PH is realizing M in this last case, if the other realizers are needed to justify M's causal powers. But there are more serious problems. To say PH realizes M and M has causal powers that cannot be found in PH, implies that some other facts about humans have to be true in order to understand the realization relation. How can PH realize M and M have powers that are not found in PH? This seems not to makes sense. After all the realization relation is a relation that attempts to specify a metaphysical dependence of the mental on the physical. But how can physical primacy be defended if the notion that supposedly gives its ontological foundation cannot account for such important fact as that of causation.

3.3 Emergence

The point is that if somehow the only way to give an account of mental causation within the nonreductive physicalist framework were to accept this picture, or some improved one but still claiming that the causal powers of the mental go beyond those of the realizing property, then the only way to make sense of the notion of realization would be to appeal to emergent properties of physical systems.

However the positing of emergent properties with its set of distinct causal powers brings lots of problems for the nonreductive physicalist.

That these features of mental properties are taken as definitive of emergence can be seen in the following passage by Tim Crane (2001). He writes that

> An emergent property, on this conception, is one that has causal powers that are distinct from the causal powers of the lower level properties on which it supervenes. If you give a list of an object's causal powers, listing only the causal powers of the lower-level properties of the objects, then you will not have given a complete list of the object's powers. (p.216)

This is exactly the view that we have been analysing. But then one cannot see how to make sense of the completeness of physics, that the nonreductive physicalist, as a physicalist, is committed to.

Of course the problem of the completeness of physics would not be an issue to a champion of emergentism, as he would happily admit its violation. By the presence of emergent properties, certain systems enter in causal relations that cannot be accounted for in terms of the causal powers of physical properties. For the emergentist the mental (and other special sciences properties that share this feature) is something "over and above the physical". Moreover the emergentist can provide a natural justification for the explanatory potential of the special sciences - they work as well as they do because they do pick up the properties and causal powers that are proper of their specific domain of explanation. No lower level type of explanation or base properties could do the explaining they do since they do not possess the causal powers that are required to appeal in order to have a successful explanation.

But wait, maybe we are going a bit too fast. Maybe there is a physicalist way of meeting the challenge of accounting for type distinct mental causal powers that certain objects are said to have in virtue of having a mental property. Suppose a creature or system S in virtue of having a mental property M has a certain causal power, say Ω, that none of its possible physical realizers have[28]. By having this power, lets suppose that there will be a possible situation where this power will manifest itself in S in a physical way. Now in such a case, in order to account for the physical effects of S does completeness need be breached?

[28] This is going a little bit farther than Clarke's account in that in his account M's causal powers are accounted for by the union of all the causal powers of all the physical realizers. However the same reasoning developed here applies to Clarke's account also, if only we take Ω to be type identical to some causal power of some physical base of M, but not of this particular realization.

Suppose someone says no, because all that is happening is accounted for by massive overdetermination. That is, all that happens in the physical world can be explained physically but it's sometimes duplicated by distinct mental powers, not only numerically distinct, but type distinct also from the physical. Now, due to the claim of massive overdetermination, all effects would have physical causes, and consequently there would always be a fully physical explanation for every physical effect.

Notice however that what we are conceiving is that the mental power Ω endows S with a power type distinct from any existent in its putative physical realization. Now how can we suppose that S has Ω and simultaneously not attributing to S's possessing of Ω certain effects that would otherwise be absent from S's manifestations? It seems that S's possessing of a causal power type distinct from the causal powers that are due to its physical properties, implies that to explain S's physical manifestations we have to appeal to Ω, thus violating completeness.

Now since accounts (A), (B) and (C) share this feature, they have to be rejected by the nonreductive physicalist. He cannot allow for the mental to have causal powers that are not type identical to causal powers possessed by physical properties. That is, the realized property cannot have any causal powers that are type distinct from the causal powers of the realizing property.

3.4 Numerical Identity of Causal Powers

But this raises the possibility that while being type identical they are numerically distinct. So, let P be a physical realizing property of M in S. Now suppose that P endow S with causal powers CP1, CP2, CP3 and CP4, and that M endows S with causal powers CP1, CP2 and CP3 and these latter causal powers are numerically distinct from those that P endows. Lets suppose that M can ground causal relations between events involving S. Since M does not have causal powers type distinct from P, those causal relations can be subsumed in physical laws. So there does not seem to be problems with completeness.

Moreover physical events are overdetermined by the causal properties of their causes. Contradicting claim (4) of the exclusion arguments that says that there is no overdetermination. But as I argued in section II, the issue of overdetermination has to be decided in terms of the ontological claim of the position. If a proposal gives

ontological substance to a claim of overdetermination, then there is not further issue to settle.

But can a nonreductive physicalist make sense of such proposal? I don't know any philosopher who defends such a thing. And the reason for that seems to be that it does not make much sense in a physicalist framework. After all the mental is supposed to depend with metaphysical necessity on the physical, and this dependence is in part used to express the causal priority of the physical domain. Is the fact that the mental duplicates some of the causal powers of its physical realization base an assertion of the priority of the physical and as such taken to be in the right physicalist spirit?

Take fragility and the physical causal structure that realizes it. On this proposal, fragility duplicates some of the causal powers of this structure, presumably the powers that manifest certain effects that are characteristic of fragility. Now, what does it mean for a vase that is fragile, to have this doubling of causal powers? Is it more fragile than it would be, if per impossible, it had the same structure and not the fragile property? Thus in the actual world it would break more easily that it would break if it had the same structure and not the realized property?

I can't make sense of this idea as it seems a form of emergence, but if we suppose that the doubling of causal powers that individuate fragility are not operative then why do we suppose them to be there at all? Moreover since these causal powers are taken to be type identical to the causal powers of realizing properties, on which they depend, it is only natural to suppose that the causal powers of the mental are numerically identical to the causal powers of the realizing properties. This seems also to be the position of most philosophers that reject (A), (B) and (C) and opt for (D) or (E).

So in the rest of this discussion I will take it that the nonreductive physicalist denies both the possibility of the mental having causal powers that are not found, or surpass those, of the realizing property, and that there is no doubling of causal powers. So the following condition has to be respected: Each causal power of a mental or special science property is identical, in the sense of numerically identical, with a causal power possessed by its realizing physical property.

3.5 (D) and (E)

If (A), (B) and (C) cannot give substance to the notion of realization, it might seems however that (D) is a lot more promising. By sharing all causal powers of the mental with the causal powers of the physical, while letting the physical having powers that surpass those of the mental, it gives a clear sense to the primacy of the physical, and takes the magic out of the notion of realization, that seemed to be the only way out in those other cases.

What is in need is a proper characterization of the realization relation that gives sense to this demand. Sydney Shoemaker (2001) gives one such understanding of the notion of realization that tries to make sense of the difference in causal powers between realizer and realized while simultaneously giving an account of mental causation.

Shoemaker thinks that distinct properties confer to a system different sets of causal powers. This view of properties as individuated by their sets of causal powers is consistent with the view that we have expressed about the way to understand the reality of mental properties. To be real is to have causal powers, which means to make a difference in the world. Such difference might be expressed by the ability of mental properties to ground causal relations.

Now if a certain property P confers to S a set of causal powers, say, CP1, CP2, CP3 and CP4 then it seems that a property, say M, individuated by the fact that would confer to the same system causal powers CP1, CP2 and CP3 is in fact a different property since it doesn't endow S with exactly the same set causal powers.

How should we understand the relation between P and M? Notice that when a system has P it will also have the causal powers that individuate M. Can we then say that both P and M are present in the system? Shoemaker thinks that we should understand properties that have this feature in terms of the determinable-determinate relation. So if a system has some determinate property, say it is red, then it has the determinable property of being coloured. The idea is that red has a certain set of causal powers that it confers to the system and there is an appropriate subset of those causal powers that constitutes or amounts to the system being coloured. So when P is instantiated in a system, it means that M is also instantiated in that system since on this reading

property M will be a determinable of P, and when a determinate is present so is the determinable.

So in system S, P realizes M because the causal powers given to S by M are a proper subset of those given by P. Since in each case a mental property will have a subset of the causal powers of the physical realizing property there is no causal competition between them and as such the issue of exclusion does not even arise.

Now this seems to give an especially interesting reading to multiple realization. Multiple realization tell us that due to the probability of extreme heterogeneity of the physical properties that realize mental properties, we expect to find for such diverse realizers very different sets of causal powers. But since all those realizers no matter how heterogeneous, realize the same mental property they have to have in common at least the subset of causal powers that are sufficient for the causal role that individuates the mental property that is realized.

Take pain and two of its realization bases. Say C-fibres firing and the vibration of silicon. When C-fibres fire in humans they will feel pain and have all sorts of pain behaviour and the same for aliens upon the instantiation of one of their physical bases. Whatever the set of causal powers that C-fibres and Silicon vibration endows to the organisms that have one of such realizing properties it has to confer the causal powers to do the things that specify the causal role of pain. Supposing, as always, that the pain they share is the same (otherwise there wouldn't be any prospect of an autonomous general scientific psychology) then those realization bases have to have the causal powers to lead to avoiding behaviour, nursing behaviour etc. if such behaviour is part of the causal role of pain. That is, they have to share the same causal powers that constitutes the causal role of this mental property. However they also have other causal powers that aren't shared. When a scientist opens up a human skull and looks at C-fibres they will look grey to him and when he studies the brain of an alien he might find out that it is translucent and emits a buzzing sound. However it might be that none of the causal powers that are responsible for these phenomena are relevant for those properties to be the realizers that they are.

While accounts (A), (B) and (C) had difficulties because they made the notion of realization mysterious, this one, by keeping the causal powers of the mental within the bonds of the physical promises to offer a proper physical characterization of

realization. The mental here has no causal powers that are not part of the causal powers of a bona fide physical property.

But there are problems. The crucial question is why we should understand M as an extra feature of creature or system S. Why is M a mental property and not a physical property? Since all the causal powers of M are a subset of the causal powers of a realizing physical property P, why should we conceive the existence of any other causal property besides P? If S is P then S has all the causal powers that P endows S with. Let this causal powers, be CP1, CP2, CP3 and CP4. Now what this notion of realization says it that by having these causal powers S also has M, since M's causal powers are, lets suppose, CP1, CP2 and CP3 (a proper subset of the causal powers of P). So M shares the selfsame causal powers CP1, CP2 and CP3 with P. But then why should we suppose that when a creature S has P it also has M as a distinct causal property? Since there is no doubling of causal powers, CP1, CP2 and CP3 are causal powers of S that are present in S once and not twice over. So, once S has P why do we need to suppose that there is also a mental property present? The present point is that M as a realized property seems in risk of collapsing from a broadly physical property into the kind of physical properties that prima facie realizes it. In the remaining of this paper I want to show the plausibility of this point in detail.

I might seem that a defence of M can be found in some considerations offered by Shoemaker that are in the same spirit to some of what we said in section I concerning the special sciences and cross-classification. The view[29] might be expressed this way: A set of causal powers is such that there is a property that confers such causal powers to an individual, just in case this set of causal powers is such that there are laws that pick this property and causal powers and unifies them in a theory M. This theory then, consist of laws that specify how these properties together operate to bring about, in certain circumstances, the various effects that they are capable of bringing about.

So P being a genuine physical property figures in all kinds of laws that describe how the causal powers CP1, CP2, CP3 and CP4 operate together to bring about all the

[29] Here I am merely saying what I think is the spirit of a justification of taking M as a real property, and not actually specifying anybody's take on it, though close to what Fodor(1974) says in "Special Sciences".

kinds of effects that P is able to bring about. But *supposedly* these laws will not be able to express the theory that concerns the causal powers of M. For that we need theory M. Moreover, following multiple realization, M can be realized by many different physical properties, PH, PA etc. Now since these properties are different and sometimes said to be wildly heterogeneous from the physics point of view, there are no similarities to be stated in physical terms. No point of view within physics that can unify these realizations. But they nonetheless can be unified from the point of view of M, the mental point of view. And this unification points to the fact that all the realizing properties of M, have to have in common, in spite of their physical heterogeneity, the set CP1, CP2 and CP3, the causal powers of M. Fodor (1974) says,

> Any pair of entities, however different their physical structure, must nevertheless converge in indefinitely many of their properties. Why should there not be, among those convergent properties, some whose lawful interrelations support the generalizations of the special sciences? Why, in short, should not the natural kind predicates of the special sciences cross-classify the physical natural kinds? (p.134)

Indeed why not. After all we started out with the assumption that there are causal grounding mental properties. And these causal grounding relations as there was said to exist between mental properties (and within the special sciences properties) were supposed to specify the causal powers of the mental. It seems natural then to think of theory M as consisting of sets of those causal grounding relations, and M having the causal powers CP1, CP2 and CP3. According to such a view we know that there are special sciences properties because of such generalizations, because there are such causal grounding relations type connecting special sciences properties.

But the fact that CP1, CP2 and CP3 are also causal powers of property P seems to imply that those causal powers are physical causal powers. Contrary to claims (A), (B) and (C), where some powers of the mental had no *grounding* in the physical, here they are trivially grounded by being causal powers that an object has in virtue of being P. Is this a weakness or a strength of this position? A nonreductive physicalist, like Fodor (1974), would claim that such view is a strength, because

> The point of reduction is not primarily to find some natural kind predicate of physics co-extensive with each natural kind predicate of a reduced science. It is rather, to explicate the physical mechanisms whereby events conform to the laws of the special sciences. (p.131)

Such a picture would be somewhat like this: There would be a causal grounding relation LM that would specify, or establish a type connection between two mental properties, M1 and M2. In S these mental properties would have as realizing properties P1 and P2, lets suppose. Now due to LM we know that m1 is causally related to m2. But M1's and M2's causal powers are a subset of the causal powers of P1 and P2 respectively. However since physics is primary we know that there must be a causal law LP that relates these events. This law will be one that relates P1 and P2, giving a coherent picture that explains how there are "physical mechanisms whereby events conform to the laws of the special sciences". This conformity can be seen by the fact that the physical law is compatible, in this case, with whatever causal interactions are specified by the grounding relation that subsumes the mental property. After all such grounding relations specify the causal powers of the mental and those are a proper subset of the causal powers of the physical realizing property.

Now, in every instance of mental causation or of special sciences causation, there would have to be some story as told but the point is that the causal grounding relation LM would have an application, that is, would be able to ground causal relations between mental events or special science events, *even* in cases where LP would not be applicable. In fact every time one changes systems that still have M one would probably find out that the physical laws would be distinct, due to the possibility of the heterogeneity of realization bases, but that LM would still be valid. Here one sees that LM gives a unification that those different physical laws could never give. Or so it seems.

3.6 Problems for (D) and (E)

This view of the relation between special sciences' properties and physical properties and their standing to causation seems quite compelling and seems to be the reason

why nonreductive physicalism has been so appealing to a lot of philosophers. But I don't think it quite works.

If we take a look at this account and ask why do we think that the grounding relations that type connect mental properties are not physical causal laws it seems to me difficult to give any positive reason. In fact I think that the reason the account given seems so seductive is because we are focusing on the wrong kind of things when we begin to reason. We begin by focusing on the regularities of the special sciences and psychological type relations. In the beginning of this section I said that there was an intuition, the multiple realization argument, and an ambition, the independence of psychology and other special sciences vis-à-vis the physical sciences that coloured the interpretation of these issues in a way that would not let us be ontologically serious. It seems to me that this is evident here. When one is presented with a view of realization where the causal powers of the mental are said to be a proper subset of those of the physical one does not question the ontological implications of the view and reason from them as a starting point but instead reads into it the ambition of our previous assumptions.

But looking at the metaphysics of the realization relation I think the following picture is much more plausible. Since every realization of M or any other special science property has in common certain causal powers, say CP1, CP2 and CP3 in the case of M, then why should we not believe that in fact what a creature has in common when it realizes any of those special sciences properties is some physical property P*? After all those causal powers are good physical causal powers, powers possessed by P! P* would be a physical property that endows S with exactly those causal powers, CP1, CP2 and CP3 that were first identified as being the causal powers of M. P* would be the realizer of M in all systems that have M. In what follows I will try to motivate this reading.

Of course not any set of causal powers points to a physical property at work, but if those causal powers are found together and unified by type relations then the best explanation for such a case is that we happened to hit on a physical property. We, after all, are talking about causal powers that are a subset of those of a physical property. How come these causal powers do not point to a single physical property that is common to all the realizations of M?

The fact that there are causal grounding relations that specify the causal powers CP1, CP2 and CP3 seems to argue for M has we have seen. It is because of those causal powers that it makes sense to refer to the presence of the mental property. M is real on account of having those causal powers. But those powers are in S on account of P – P has causal powers CP1, CP2, CP3 and CP4. Now the difficulty is that it does not seem to make much ontological sense to suppose that there are two properties that are said to be causal, since two such properties would seem to require two sets of causal powers where one is *not* a subset of the other. If one is a subset of the other then that is *best* seen as pointing to the presence of one thing in the object, and not two. But in this case it is not possible to see the mental as able to ground causal relations, unless it is reducible to some physical property P*. That is, the supposition that we always have taken for granted, that the mental is causal, only makes sense if the mental is reducible to physical properties (or if it is emergent).

The crucial assumption in the last paragraph is that if one property P endows an object with a set of causal powers it does not make much sense to suppose that the same object has another property M that is also a causal property where the causal powers of that property M are numerically identical to a proper subset of the powers of the physical property P. Of course this needs a little qualification: If the physical realizer of M, say P, is a complex property with physical constituents, say constituted by P' and P*, then P will have as constituent P* that, lets say, has CP1, CP2 and CP3 and P' has CP4. So P endows S with those causal powers because its constituents have those powers. This is fine. But if we want to claim in the mental case that M is a property that is not a constituent of any physical property, but is such that endows to an object certain causal powers that are numerically identical with some proper part of the causal powers of its realizer than we have a problem.

Take two realizers of M, PH and PA. So there are laws that subsume PH and PA respectively such that each of them is said to endow certain causal powers to the creatures that have them. Among them are CP1, CP2 and CP3 the causal powers in virtue of which each of the realizers are said to realize M. Now the reason, according to such a view, that M is a significant entity is that there are grounding relations that subsume M. Such relations specify how the causal powers of M, operate. But those powers are identical to some of the powers specified by the physical laws that subsumed PH and PA. So how came these grounding relations are not physical laws,

thus making M reducible to a physical property? Once we try to understand this problem from bottom-up, instead of top-down, it seems that this is the picture that is required to make sense of such a view. It gives us the best explanation.

That is, if we assume that a set of causal powers points to a property in case there are grounding relations that subsume such property, then it would seem that there would have to be a physical property in common in every case of the realization of a mental property. And the reason is that the causal powers of the mental are a subset of the causal powers of the physical. Only on the implausible assumption that there are causal powers of physical properties that could never figure in physical causal laws could we negate such a reading. But then those physical laws (grounding relations) would describe exactly those powers that we take to be the powers of M, pointing to a physical property at work. But if this were plausible then such a physical property would be coextensive with the putative mental property. The making of reduction. That means that it is reasonable to think such notion of realization cannot account for the delicate tension that is required to keep distinction and dependence between mental and physical properties while accounting for mental causation.

I think that these considerations are enough to settle the issue regarding causal powers of the mental that are numerically identical to a subset of the causal powers of the physical. I will however go into a more speculative mood and try to show another possible reason to be suspect of either (D) and (E).

3.7 More Problems: a bit of Speculation

I think there are more reasons to believe that such account fails. A good way to see the problem is to focus on the realizers. PH and PA are physical realizers of a certain mental property M in humans and in Aliens respectively. They are said to be heterogeneous properties from the physical point of view; nonetheless these properties realize M. In such a case M's causal powers CP1, CP2 and CP3 are a subset of the causal powers of PH and PA. The following question brings to the fore my worries : How come PH and PA, being heterogeneous, can endow an object the causal powers that are said to be distinctive of certain mental property? There is a straight answer to this: presumably because they realize M! And by realizing M they have to have those causal powers. But there is something funny here that needs to be settled. How come

realizers that are heterogeneous from the physical point of view can have and bring about the same physical manifestations? Note that this happens to all physical realizations of M, no matter how heterogeneous they are.

I think that this issue can be settled within Shoemaker account of property individuation: that properties are individuated by their causal powers. If one takes the view, that causal laws depend on the nature of natural properties, as is the natural reading of such an account, then the causal powers of those properties are to find a specification on those laws. And so if there are regularities to be found that depend on causal powers CP1, CP2 and CP3, or the causal powers of any prima facie special science property, then it is compelling to think that this is so because of the existence of some physical property P* that is individuated by CP1, CP2 and CP3.

That there are such regularities can be ascertained because of the existence, in every case of mental causation, of physical mechanisms that ensure that the right effects came along. Even extremely heterogeneous physical systems, if they realize mental property M, have to bring about those causal effects that are distinctive of M. That means that those systems have to have the causal powers CP1, CP2 and CP3. But then, if we individuate causal properties in causal terms it seems that those systems have to have some physical property in virtue of which those systems bring about those same effects. Physical systems that can, hypothetically, be so different from each other can implement those causal powers physically only if they have in common something physical. To be able to bring about those effects in physical ways means they have to share physical properties.

But what about the view that objects have the causal powers they have due both to the properties plus the laws of nature? Is it not coherent to suppose that two distinct properties, even wildly heterogeneous physical properties, could be subsumed under laws in a way that those objects would have a set of causal powers in common?

There are two distinct puzzles here, which need explication. One seems to be asking for an explanation of why certain objects have came to possess certain properties with certain causal powers. Why do certain objects have heterogeneous properties that

share a proper subset of their causal powers, in virtue of which there are realizers of the same property?

David Papineau (1993) raises the question about functionalism

> If there is nothing physically in common among the realizations of a given mental state, then there is no possibility of any uniform explanation of why they all give rise to a common physical result. And that's what I find puzzling. (p.2 of chapter 2 of the online version)

He goes on to argue that in cases where there is no other explanation then "the notion of a reduction is precisely the notion of an account which shows that nothing incredible is happening at the physical level"(p.7 of chapter 2 of the online version). What is the other explanation that would preclude the need of reduction? Whenever there is purpose or design in the objects under consideration. If we can have an explanation of why certain objects behave the way they do that points to a selection of the properties that can do whatever is necessary then there is one sense where the puzzle is dissolved. Why do they behave that way? Because they have properties that were selected exactly because they could do the job.

Suppose we want to build eyes and we need material to make a cornea. We need something that lets the light in without much distortion. So if someone asks why all those eyes, in spite of being physically heterogeneous have the power to let the light in, we can give them a straight answer. Because, in spite of the heterogeneity of the material, we chose only as building blocks whatever we could find that lets the light in. Papineau takes it that such explanation remove the puzzle of why heterogeneous properties can nevertheless realize the same properties. But lacking such explanation, he thinks that the best explanation for such phenomenon is in the existence of certain property P* that all those realizations share. Leading to reduction.

But there is another puzzle. A puzzle that applies to both cases; where there is a teleological explanation and to those cases where there isn't one. Take all those materials that were used to build up a cornea. All those distinct materials are able to do the job as well as they do because they were selected by their ability of letting the light in in a suitable manner. But, once again, ask: If those materials are heterogeneous how come they are able to bring about the same effects? How came those materials share so many causal powers? Having a teleological explanation does

not even give us a clue to a possible answer. Pace Papineau, I find that it is here that it would be incredible that there was no explanation as to why heterogeneous properties have certain causal powers in common but do not have a physical property in common that is responsible for it. It would seem that it is a necessary requirement of a physically acceptable world that commonalities of causal powers point to commonalities of physical constitution. I think the reason Papineau does not accept the second puzzle has to do with the common view that assert that it is a brute fact of our world that different things can bring about similar effects. The idea is that it is a brute fact that wildly heterogeneous objects, can bring about similar effects. That is, distinct properties could be subsumed by laws in such a way that they could endow objects with the same set of causal powers. Nevertheless I am not persuaded.

The reason is that I find Ned Block's (1997) Disney Principle highly intuitive and plausible

> The laws of nature impose constraints on ways of making something that satisfies a certain description.(p.120)

This principle would make it credible to think that there are only so many ways that the world can be made that lead to the same results. One consequence of this would be that the notion that a realizer can be wildly heterogeneous from the physical point of view is extremely implausible. But it seems to me that this principle can be seen to be much stronger than the uses Block makes of it. One way to motivate this principle might be to ask what happens at the lowest level of causation. Take two distinct fundamental particles that in an accelerator are made to crash into a third fundamental particle. Do they cause the same effects? I would assume that rejecting the Disney Principle would amount to the claim that they could cause exactly the same effects. But how could that be? It might be a brute fact of the world, but I take it that such idea goes directly against the principle of economy. A particle that had exactly the same effects of another would seem not to be required.

But take the following supposition; that both properties had the same effects in certain types of situation but would differ in another's. This would give space to make sense of their distinctness. Take one of the situations where they would differ and another where they would have the same effects. How to explain this? There is nowhere else to go to explain the difference since we are talking about the fundamental entities of

the world. The only way would be to refer to the laws of nature as a brute fact to explain such a difference and I take it that that is strange. Compare this to a situation where one of the particles were a complex entity with a component of the same type of the other. In such a case we could perhaps see how they could behave identically in certain situations. And how they could differ. This would be explained by their different structure. But in the case at hand, where we consider them basic properties there is no way to account for the difference, except by referring to the brute fact that there are some laws that tells us that such particles behave in such a way. This would be much more implausible than the former explanation.

It would seem to me, that even those that think that objects have certain causal powers due not only to the properties they have but also to the laws of nature, should accept that once the laws of nature are fixed, then properties could be partly individuated by their causal powers. Once the laws of nature are fixed there should be no issue about the properties of an object if they had the same causal powers. That case should prompt us to think that the same properties are at work. But this brings us close to accept that commonalities of causal powers point to commonalities of physical constitution.

My claim here is not that these considerations are definitive but that they are highly plausible. But if we take them seriously, then the same argument that we applied to the causal view of properties is good enough for the present case. That the same causal powers, CP1, CP2 and CP3 point to the presence of a physical property P*. And if so, reduction would be in the offing.

In light of such notion of realization it would seem that the reasons that lead us to do psychology are purely epistemic reasons.

Notice that whatever laws one would get from the property individuated by those causal powers it seems that they will be strict laws, or at any rate they will have the same precision and level of strictness of the physical law that subsumes the physical property that has those causal powers as a subset. So while the nonreductive physicalist seems to regard mental properties and other special sciences properties as subsumed under ceteris paribus generalizations, on this proposal, they seems to enjoy the same level of precision of physical laws. And the reason for this seems to be that

such proposal suggests, contrary to the intention of Shoemaker, that there is a physical property that unifies what we would have thought to be the province of the special sciences.

About (E), the view that the mental and the physical have the same set of causal powers, we can see that it suffers from all the problems that were faced by (D) but moreover falls prey to the view that natural properties are individuated by their causal powers. Here the nonreductive physicalist would have difficulties to even state the existence of mental properties, since they seem to be one and the same with the physical.

It seems then that those views that assume the numerical identity of the causal powers of the mental with those of the physical cannot give substance to the mental. It seem that the mental is, as John Heil (1999) puts it, "swallowed up by its realizes"(p194).

Conclusion

It is interesting to note that proposals (A), (B) and (C) are proposals that understand the mental as emergent. Clarke's proposal, that the causal powers of mental properties exceed those of the physical realizers seems either to lead to emergent properties that risk violating completeness, something that the nonreductive physicalist does not want and simultaneously it seems that the realizing relation is always somewhat incomplete since it is difficult to make sense of a property that is realized by another, that depends on it and ends up with causal powers not possessed by it. However it seems to me that it is only a view of this kind that can sustain the notion that mental properties are real causal properties consistent with distinction. Unfortunately, it is not a physicalist account.

View (D) and (E) and in particular Shoemaker's proposal, with the idea that the causal powers of the mental are a proper subset of the causal power of the physical property that realize it in a system is unable to motivate against the charge that the realizer absorbs the mental property. As such this view is unable to give a notion of realization that makes sense of the idea that mental properties are real causal properties. Though this view is thoroughly physicalist, its ontological claims do not suggest a reading of multiple realization that support the claim that the mental is

distinct from the physical. On the contrary, it is plausible to think that such proposal seem to make the mental ripe for reduction.

So the nonreductive physicalist cannot object to premise (4). Our ontological scrutiny of overdetermination seems to point out that it does not make sense to suppose that there is overdetermination, supporting the original claim. As such we have to revise our supposition that mental properties are causal, or embrace emergentism, or adopt some version of type identity theory.

But the first claim, epiphenomenalism, is extremely unwelcoming to a nonreductive physicalist. After all, the reason that one might want to keep the mental with an ontological significance of its own might be to sustain and give sense to the causal explanatory potential of our mental life. But if the epiphenomenal road is taken, such motivation disappears.

While emergentism cannot be an alternative for the nonreductivist without giving up physicalism, identity, the last option, seems to be more appealing. Taking this road would end most of the problems that were expanded in this work.

Bibliography

Baker, Lynne Rudder (1993) "Metaphysics and Mental Causation." In Heil and Mele (1993): 75-96.

Bennett, Karen (2003) "Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract It." *Nous* 37: 471-97. Online at http://www.princeton.edu/~kbennett/exclprobrev.pdf

Beckerman, A., Flohr, H., and Kim, J.. eds.(1992). *Emergence or Reduction*. New York and Birlin: DE Gruyter.

Block, N. (1980a). *Readings in Philosophy of Psychology*, Vol 1 Harvard University Press: Cambridge.

Block, N. (1997). "Anti-reductionism slaps back". *Philosophical Perspectives* 11:107-32.

Burge, T. (1993). "Mind-Body Causation and Explanatory Practice". In Heil, J. and Mele, A., eds. (1993), pp. 97-120.

Chalmers, David (1996). *The Conscious Mind: In Search of a Theory of Conscious Experience*. New York: Oxford University Press.

Clarke, R.1999. "Nonreductive physicalism and the causal powers of the mental". *Erkenntnis* 51:295-322.

Crane, Tim (1995) "Mental Causation Debate." *Proceedings of the Aristotelian Society*, Supplementary Vol. 69: 211-36.

Crane, Tim (2001). "The Significance of Emergence," in Loewer and Gillett, eds., *Physicalism and Its Discontents*. Cambridge: Cambridge University Press.

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

Davidson, Donald (1963) "Actions, Reasons, and Causes." *Journal of Philosophy* 60: 685-99. Reprinted in Davidson (1980): 3-19.

Davidson, Donald (1970) "Mental Events." in Davidson (1980): 207-25.

Davidson, Donald (1974) "Psychology as Philosophy." in Davidson (1980): 231-9.

Davidson, Donald (1980) *Essays on Actions and Events*. Oxford: Clarendon Press.

Davidson, Donald (1993) "Thinking Causes." In Heil and Mele (1993): 3-17.

Fodor, J. A., (1974) "Special Sciences" *Synthese* 28, 97-115. Reprinted in Block 1980a

Gillet, Carl and Barry Lower (2001) *Physicalism and Its Discontents*. Cambridge: Cambridge University Press.

Heil, J. (1998) *Philosophy of Mind: A Contemporary Introduction*. London: Routledge.

Heil, J. (1999) "Multiple Realizability." *American Philosophical Quarterly* 36: 189-208.

Heil, J. (2003) *From an Ontological Point of View*. Oxford: Clarendon Press.

Heil, John, and Alfred Mele, eds. (1993) *Mental Causation*. Oxford: Clarendon Press.

Heil, John and David Robb (2003) "Mental Properties." *American Philosophical Quarterly* 40: 175-96.

Honderich, Ted (1982) "The Argument for Anomalous Monism." *Analysis* 42:59-64.

Horgan, T. (1993). "From Supervenience to Superdupervenience: Meeting the Demands of a Material World," *Mind* 102, pp.555-586.

Horgan, T. (1998). "Kim on Mental Causation and Causal Exclusion". *Philosophical Perspectives* 11: 165-84.

Horgan, T. (2001). "Causal Compatibilism and the Exclusion Problem" Theoria, 16 (2001), 95-116. Issue on mental causation, edited by J. Corbi.

Jackson, Frank (1982). "Epiphenomenal Qualia," *Philosophical Studies*, 32, pp.127-136.

Kim, J. 1984. Concepts of supervenience. Philosophy and Phenomenological Research 45:153-76. Reprinted in *Supervenience and Mind* (Cambridge University Press, 1993).

Kim, J (1989) "Mechanism, Purpose, and Explanatory Exclusion." *Philosophical Perspectives* 3: 77-108. Reprinted in Kim (1993a): 237-264.

Kim, J.(1993). *Supervenience and the mind: Selected Essays*. Cambridge: Cambridge University Press.

Kim, J.(1993a). "Multiple Realization and the Metaphysics of Reduction" In Kim, J (1993a), pp 309-35.

Kim, J.(1993b). "The Nonreductivist's troubles with mental causation" In Kim, J (1993a), pp 336-57.

Kim, J. (1997). "The Mind-Body Problem: Taking Stock after Forty Years", *Philosophical_Perspectives* 11, 185-207.

Kim, J. (1998). *Mind in a_Physical_World*. MIT.

Kim, J (2003) "Blocking Causal Drainage and Other Maintenance Chores with Mental Causation." *Philosophy and Phenomenological Research* 67: 151-76.

Loewer, Barry (2001a). "From Physics to Physicalism," in Loewer and Gillett (eds.), *Physicalism and Its Discontents*. Cambridge: Cambridge University Press.

Lyons, W. (1995) *Modern Philosophy of Mind.* Everyman.

Malcolm, Norman (1968) "The Conceivability of Mechanism." *Philosophical Review* 77: 45-72. Reprinted in Watson (1982): 127-49.

Marras, Ausonio (1998) 'Kim's Principle of Explanatory Exclusion', *Australasian Journal of Philosophy,* Vol. 76 No. 3 (Sept. 1998): 439–451.

McLaughlin, Brian (1989) "Type Epiphenomenalism, Type Dualism, and the Causal Priority of the Physical." *Philosophical Perspectives* 3: 109-35.

McLaughlin, Brian (1992). "The Rise and Fall of British Emergentism," in Beckermann et al. (eds.) *Emergence or Reduction?*

Melchert, Norman, 1986, "What's Wrong With Anomalous Monism," *Journal of Philosophy*, Volume 83, Issue 5, 265-274.

Mills, Eugene (1996) "Interactionism and Overdetermination." *American Philosophical Quarterly* 33: 105-17.

Papineau, David (1993).*Philosophical Naturalism*. Basil Blackwell 1993. Online at http://www.kcl.ac.uk/ip/davidpapineau/Staff/Papineau/PhilNat2nded/PhNatIndexrevised.htm

Papineau, David (2001) "The Rise of Physicalism," in Loewer and Gillett (eds.), *Physicalism and Its Discontents*. Cambridge: Cambridge University Press

Pereboom, D. And Kornblith, H, "The Metaphysics of Irreducibility", *Philosophical Studies* 63, 125-145

Place, U. T. (1956) "Is Consciousness a Brain Process?", reprinted in Lyons,W. (1995), pp 106-116

Putnam, H. (1967) "The Nature of Mental States", reprinted in Putnam (1975).

Putnam, H (1973). "Philosophy and our Mental Life", reprinted in Putnam (1975).

Putnam, H. (1975). *Mind, Language, and Reality.* Cambridge: CUP.

Savellos, E.E. & Yalcin, Ü.D., eds., (1995): *Supervenience. New Essays*. Cambridge University Press, Cambridge.

Shoemaker, Sydney (1980) "Causality and Properties." In van Inwagen (1980): 109-36. Reprinted in Shoemaker (1984): 206-33.

Shoemaker, S. 2001." Realization and mental causation". In *Physicalism and Its discontents*. Gillet & Lower. Cambridge University press 2001.

Smart, J.J.C. (1959) "Sensations and Brain Processes", reprinted in Lyons,W. (1995), pp 117-132.

Sosa, Ernest (1984) "Mind-Body Interaction and Supervenient Causation." *Midwest Studies in Philosophy* 9: 271-83.

Wilson, Jessica (1999) "How Superduper does a Physicalist Supervenience Need to Be?" *Philosophical Quarterly* 49: 33-52.