

# Chapter 9

## Causal Exclusion and Overdetermination

Markus E. Schlosser

This paper is about the causal exclusion argument against non-reductive physicalism.<sup>1</sup> Many philosophers think that this argument poses a serious problem for non-reductive theories of the mind—some think that it is decisive against them. In the first part I will outline non-reductive physicalism and the exclusion argument. Then I will distinguish between three versions of the argument that address three different versions of non-reductive physicalism. According to the first, the relation between mental and physical events is token-identity. According to the second, mental events are distinct from physical events, but the latter metaphysically include and determine the former. And on the third version, mental and physical events are entirely distinct. I will argue that the causal exclusion argument is not decisive against non-reductive physicalism in any of the three versions. According to non-reductive physicalism, mental events are dependent on physical events. Causal exclusion and overdetermination, however, requires distinct and independent causes. I will argue that the burden of proof lies with the opponents of non-reductive physicalism, who have to explain how metaphysically *dependent* events can possibly overdetermine an effect or exclude each other from being causally efficacious.

### Non-Reductive Physicalism

Assume that there is a true theory of the mind that employs intentional vocabulary, and call that theory *psychology*. Non-reductive physicalism is the conjunction of physicalism and the claim that psychology is not reducible.

Physicalism, one may think, says that everything is physical. It is common to distinguish physicalism from claims such as that there are only material objects or that everything is composed of matter. Physicalism refers to physical things, rather than material ones, and it leaves it up to the physical sciences to specify what counts as a physical entity. Further, the view concerns not only objects and what

---

<sup>1</sup> I would like to thank my respondent Zoe Payne, the audience of the Mind 2005 conference and Sarah Broadie for their helpful comments and remarks.

they are composed of, but also events and properties. Now, if non-reductive physicalism is to be a consistent position, then physicalism cannot be construed as the view that all objects, events and properties are physical. In order to see why, let us have a brief look at the notion of reduction involved.

In the philosophy of mind it is common to talk about the reducibility of mental types, properties or kinds. However, it is also common to point out that reduction is an issue that concerns *theories* and their explanatory power or autonomy. How do these claims hang together? On the standard model of scientific reduction, whether a theory *T* is reducible to the theory *T\** depends on whether the properties that *T* ranges over can be identified with some of the properties *T\** ranges over. That is, *T* is reducible to *T\** if and only if *T*-types are identical with *T\**-types. On that view, psychology is reducible to some non-mental theory *T\** if and only if mental types are identical with (non-mental) *T\**-types. That claim, however, has been contested as too strong. Some philosophers think that type-identity is not necessary for reduction, since weaker bridge-laws—laws that do not claim type-identities—are sufficient. What is uncontested, though, is that if psychology is not reducible, then mental kinds cannot be identified with non-mental kinds.

So, according to non-reductive positions, there are mental properties that are not identical with physical properties. Hence, non-reductive theories are incompatible with a view that says that all properties are physical. If it is to be compatible with non-reductionism, physicalism must be specified accordingly. A common suggestion goes along the following lines. Physicalism says, firstly, that all concrete objects are composed of, or constituted by, physical entities, and, secondly, that all properties and events are *dependent* on physical properties and physical events.<sup>2</sup> Physicalism says, in other words, that physical particulars and properties are the *basic* or *fundamental* particulars and properties.<sup>3</sup>

---

<sup>2</sup> Compare, for instance, Beckermann, 1992, pp. 1-2, and Crane, 1995, pp. 211-212. Note that the notion of asymmetric dependence is itself in need of explanation. Some have therefore suggested construing the relation in terms of composition or constitution only. The problem with that strategy is that it is difficult to see how it can be applied to objects as well as *events* and *properties* (Compare Crane, p. 212). Others have suggested an account of the dependence involved in terms of *realisation* of mental properties by physical properties (Compare Kim, 2000, pp. 19-24).

<sup>3</sup> That is the standard account of non-reductive physicalism, which faces the problem of causal exclusion. There is, however, an alternative construal that might avoid the exclusion problem altogether. Arguably, we can distinguish between two different conceptions of reduction. On the first, reduction is an ontological issue insofar as it concerns the relation between mental and physical *properties*. On the second conception, reduction concerns the relation between *theories* (their vocabularies and explanatory powers). Non-reductive physicalism is clearly committed to non-reductionism in the second sense. It is not clear, though, whether it is also committed to ontological reductionism. If it is not, then it is not committed to the claim that there are irreducible

## The Causal Exclusion Argument

The so-called causal exclusion argument has been formulated in different ways. A first difference concerns the kinds of entities that are said to exclude each other. On some formulations the exclusion concerns mental and physical *properties*, others talk about mental and physical *events*, and some formulate the argument simply in terms of mental and physical *causes* and *effects*.<sup>4</sup> A second relevant difference concerns the mode of causal exclusion. Some philosophers insist that what is at stake is the causal *efficacy* of mental events or properties. Others, however, say that the efficacy of physical events excludes the causal *relevance* of mental events or properties.

What is common and central to all formulations of the argument, though, is the following intuition concerning causal exclusion and causal overdetermination. Suppose that  $c$  is sufficient to cause the occurrence of  $e$ , and that  $e$  has another cause  $c^*$ , which is distinct from and not part of  $c$ . Let us say that an event  $e_1$  is a *sufficient* cause of the event  $e_2$ , if  $e_1$ 's occurrence is, in the circumstances, sufficient for the occurrence of  $e_2$ . And  $e_1$  is a *partial* cause of  $e_2$ , if  $e_1$  is a cause of  $e_2$  in the sense that  $e_1$  is itself not sufficient for the occurrence of  $e_2$ , but it is part of a complex event, which is sufficient for the occurrence of  $e_2$ . Now,  $c^*$ , the additional cause of  $e$ , is either a sufficient or a partial cause of  $e$ . If  $c^*$  is sufficient, then  $c$  and  $c^*$  exclude each other from being *the* cause of  $e$ , because they overdetermine its occurrence. If  $c^*$  is a partial cause of  $e$ , then  $c$  excludes  $c^*$  from being causally efficacious, since  $c$  is already causally sufficient for the occurrence of  $e$ . (I will call causes that exclude each other or overdetermine their effects in that sense, *rival* causes).

These intuitions concerning exclusion and overdetermination are considerably strong and straightforward *only insofar* as they are formulated in terms of causes and effects. Our intuitions are far less straightforward with respect to causally *relevant properties*. It is not obvious whether instantiations of properties can exclude each other, or overdetermine effects, in the *same way* as causes. To assume

---

mental properties. I can agree that reduction is, first and foremost, a relation between theories. The question, however, whether the reduction of a theory presupposes ontological reduction is beyond the scope of this paper. I will assume, throughout, that the standard construal of non-reductive physicalism is correct.

<sup>4</sup> Formulations in terms of causes and effects can be found, for instance, in Lowe, 2003, and Merricks, 2001. For formulations in terms of instantiations of properties compare Kim, 1993 and 2000, Crane, 1995 and Menzies, 2003. Kim and Yablo, 1992, think that it is of no significance whether the argument is put in terms of causes, events or property instantiations.

that they can is a substantial—and controversial—additional assumption. That is why I introduce the causal exclusion argument purely in *extensional* terms: formulated in terms of causes and effects only. The three basic assumptions behind the argument are the following.

- (1) Mental Causation: Mental phenomena cause physical phenomena.
- (2) Causal Closure of the Physical: Every physical effect has a causally sufficient physical cause.<sup>5</sup>
- (3) Exclusion of Causal Overdetermination: Causal effects are, usually, not causally overdetermined.

The causal exclusion argument goes as follows. Every version of physicalism is committed to the three claims just presented. A non-reductive version of physicalism, however, is incompatible with the conjunction of them. Given that only events can be causes, (1) says that some mental events have physical effects. Assume that the mental event *m* is a cause of the physical event *p*; *m* is either a partial or a sufficient cause of *p*. Applying (3), we exclude that *m* overdetermines the occurrence of *p* (we assume that *p* has only *one* sufficient cause, if it has a sufficient cause). So, if *p* has a sufficient cause, *c*, then *m* either is *c*, or *m* is not a sufficient cause of *p*. According to (2), *p* has a sufficient *physical* cause. Hence, *c* is a sufficient physical cause of *p*. Given that, *m* cannot be a sufficient cause, but it must be a partial cause of *p*. Partial causes are parts of sufficient causes. Since *c* is the only sufficient cause of *p*, *m* must be part of *c*. But since *c* is a physical—that is, non-mental—cause, *m* must be a physical cause of *p*, contrary to the assumption that *m* is a *mental* cause of *p*. The argument shows that the assumptions (2) and (3) *exclude* the causal efficacy of mental events, contrary to (1).

The exclusion argument is generally considered to be a very powerful argument that constitutes a serious problem for non-reductive physicalism.<sup>6</sup> It is

---

<sup>5</sup> Generally, by *causally sufficient* I mean sufficient either for the occurrence of the effect or sufficient to determine its chance. In this paper, though, I have to restrict my considerations to the deterministic case. So, in what follows *causally sufficient* will mean sufficient for the occurrence of the effect.

<sup>6</sup> A well-known response on behalf of non-reductive physicalism is a *reductio ad absurdum* to the conclusion that there must be *something* wrong with the argument. There is no obvious reason to deny that the argument applies to the special sciences in general. If the argument can be generalised, it entails that, for instance, the efficacy of chemical, physiological and biological events is excluded by the efficacy of physical events—which is absurd. Proponents of the exclusion argument dismiss this *reductio* as evasive. What is required, they insist, is a metaphysical solution to the problem of mental causation. Or, at the very least, non-reductive physicalists must be able to spell out *what* is wrong with the argument. Compare Kim, 2000, and Crane, 1995.

acknowledged that there are different versions of non-reductive physicalism that require different versions of the argument. But it is usually thought that differences with respect to the details do not diminish the main thrust of the argument. In the following I will have a closer look at three different versions of the argument. I will show that the differences between them are significant, and I will argue that the causal exclusion argument is not decisive against any version of non-reductive physicalism.

## Events and Property Instantiations

I assumed that the relevant mental phenomena are mental *events*. Philosophers of mind, though, often talk about mental *states* and *properties*. It is common to use the term *mental events* in a broad sense that includes *mental states*.<sup>7</sup> We would, then, obtain two versions of the argument; one in terms of mental events, and the other one in terms of the instantiations of mental properties (for no one should expect the properties themselves to have a causal role). Little significance has been given to this distinction. It has been assumed that the argument is equally compelling in both versions. Let us first consider the version in terms of mental events.

The causal exclusion problem would dissolve under the assumption that mental kinds are identical with physical kinds. But this solution is obviously not an option for non-reductive physicalism, which is committed to the rejection of type-identities. Another possibility, though, is identity between mental and physical event-*tokens*.

There are two ways to distinguish events as *mental events*. An event is a mental event just in case it has a mental description, or, alternatively, just in case it has a mental property. I shall assume that the two definitions are equivalent. Given that, it is certainly possible that mental events are identical with physical events, even though mental properties cannot be identified with physical properties. For it is possible that one and the same event has both a mental and a physical description (or property).<sup>8</sup> Further, it is possible that the physical properties of such an event include those properties that realise and determine the event's mental properties. That is, it is possible that it is one and the same event that has both the mental properties and the physical properties on which they depends. Suppose that this is

---

<sup>7</sup> Compare for instance Horgan and Tye, 1988, who say that they are following a "frequent recent practice" by using "event" in a sense that includes "states, process, and the like" (p. 427).

<sup>8</sup> Davidson has famously argued that this not only possible, but actually the case: every event that has a mental description has a physical description. See Davidson 1980, essays 11 and 12.

generally the case. Then the non-reductive physicalist can say that the mental and physical events are not rival causes, because they are one and the same event—being token identical, they share causal powers.

If we consider property instantiations instead of events, that move—the appeal to token-identity—is not available. For what goes hand and hand with that alternative version is a particular view on the nature and individuation of events, according to which events *are* property instantiations. On that view, events are properties instantiated by a substance at a time. That rules token-identity out, since the instantiation of a mental property and the instantiation of the underlying physical property constitute *distinct* event-tokens.

It seems that we have identified a significant difference between the two versions of the exclusion argument. But it merely seems so, the opponents insist. According to Davidson's theory, which is *the* token-identity theory, the events in question instantiate a causal law only under their physical description. We are warranted in regarding them as cause and effect only insofar as they are covered by physical law. But that means that they are causally efficacious only in virtue of their physical properties; they cause what they cause *not* in virtue of being *mental* events. Therefore, appeal to token-identity does not help, and the difference between the two versions is of no significance. Consider, for instance, how Stephen Yablo has put that point.

To reply with the majority that mental events just *are* certain physical events, whose causal powers they therefore share, only relocates the problem from the particulars to their universal features [...]. Mental events are effective, maybe, but not by way of their mental *properties*; any causal role that the latter might have hoped to play is occupied already by their physical rivals.<sup>9</sup>

A few things, however, have been overlooked in this diagnosis. In my alternative analysis I will distinguish between *three* versions of the argument, whereby each version results from a combination of non-reductive physicalism with a particular view on the individuation of events.

### Version One

The first version of the argument is directed at the already mentioned token-identity theory. On that view, the mental events and the underlying physical events, which realise and determine the mental events, are token-identical. The charge against this view is that such events cause their effects only in virtue of their

---

<sup>9</sup> Yablo, 1992, pp. 248-249.

physical properties – the events in question have physical effects, but not *in virtue of* being mental events.

Let us assume, for the sake of the argument, that events cause their effects *in virtue of* some of their properties.<sup>10</sup> Why do the events in question have their effects not in virtue of their mental properties? How can the opponent justify the claim that they are efficacious only in virtue of their physical properties? We can distinguish between two arguments for that claim.

The first argument appeals to a connection between the causal role of properties and causal laws. It is assumed, firstly, that an event *c* causes the occurrence of the event *e in virtue of* having the property *P* if and only if *P* figures in the causal law that covers *e* and *c* (more precisely, if and only if there is a causal law according to which an event's being *P* is nomologically sufficient or relevant for the occurrence of *e*). Further, if *c* causes *e* in virtue of having *P*, *P* is said to be a causally *relevant property* with respect to *e*'s causing *c*. And it is assumed, secondly, that there are no psychological laws or regularities that ground causal claims about mental events; it is assumed, in other words, that psychological *anomalism* is true. From that it follows that no event causes an effect in virtue of having a mental property.

That first argument, however, is not decisive for the following reasons. Firstly, the second assumption—psychological anomalism—is rather controversial. It is in need of independent justification and cannot just be assumed. The problem is that most philosophers who deploy the exclusion argument against non-reductive physicalism also argue *for* some form of reductive physicalism.<sup>11</sup> But it is hard to see what *independent* reasons a *reductive* physicalist might have to hold the second assumption. In fact, reductive physicalism seems to be committed to the claim that there are psychological laws or regularities that ground causal claims about intentional states and events. Secondly, the argument shifts the focus of the debate in a way that is problematic for the opponent of non-reductive physicalism. Typically, opponents of non-reductive physicalism argue that the problem of mental causation requires a *metaphysical* solution, and that all proposals that establish merely the *explanatory* relevance of mental events or properties are inadequate or beside the point. The problem, they insist, concerns causal efficacy,

---

<sup>10</sup> Davidson's own response is that, given his view on causation and the individuation of events, it simply does *not make any sense* to say that a cause is efficacious *in virtue of* some of its properties. Causation is an extensional relation between *events*. Surely, in order to obtain a causal explanation we have to refer to the events using the right descriptions. But it does not follow that they are efficacious in virtue of some property (Compare Davidson, 1993, especially pp. 12-13). Opponents will either insist on the principle that causes are efficacious in virtue of some of their properties, or they will try to show that Davidson's response merely relocates the problem in a way that does not help to save the view. My task, however, is not to decide on the tenability of Davidson's response.

<sup>11</sup> The most prominent proponent of that strategy is Kim, 1997 and 2000.

not causal explanatory relevance. The argument, though, shifts the focus from the causal efficacy of mental events to the question whether mental properties figure in causal laws and the question whether there are psychological laws; questions which concern the explanatory relevance of the mental, rather than their causal efficacy.<sup>12</sup>

According to a second line of argument, it becomes *obvious* that the events in question have their effects only in virtue of their physical properties once the principle of the causal closure of the physical is understood in the right way. According to Kim, the basic idea behind that principle is that, for every physical event, one will never “leave the physical domain”, if one traces out its complete causal history.<sup>13</sup> The causal history of any physical event consists only of other physical events. Events are *physical* events in virtue of having physical properties. So, the causal history of any physical event can be given in terms of other events and their physical properties only. And that means, it seems obvious, that all the causal relations that constitute that history hold only in virtue of physical properties; there is no room left for mental properties to do any additional causal work.

The closure principle, however, is a purely metaphysical and extensional principle; it talks about causes and effects only. Assume, once more, that events and only events can be causes and effects. Consider a physical event *e* that is caused by *c*, and assume that *e* is both a physical event in virtue of having the physical property *P* and a mental event in virtue of having the mental property *M*. What licences the claim that *c* causes *e* only in virtue of having *P*?

I fail to see why and how the closure principle, by itself, rules out the possibility that *c* causes *e* in virtue of having *M*. Note, firstly, that in order to hold that *c* causes *e* in virtue of having *M* one does not have to “leave” the physical domain. Metaphysically speaking, we cannot leave the physical domain, since each mental event is, by assumption, identical with a physical event. Secondly, the thought behind the argument cannot be that the causal relevance of mental properties is *excluded* by the causal relevance of physical properties. Without the introduction and justification of further metaphysical principles, the notion of exclusion—just as the notion of overdetermination—applies only to causes (that is, events).<sup>14</sup> The closure principle does not entail anything with respect to the causal

<sup>12</sup>Further below I will say more about causal efficacy and causally relevant properties. Compare also note 14.

<sup>13</sup>Compare Kim, 2000, p. 40: “One way of stating the principle of physical causal closure is this: If you pick any physical event and trace out its causal ancestry or posterity, that will never take you outside the physical domain.”

<sup>14</sup>Stephen Yablo says that although “causes and effects are events, properties as well as events can be causally relevant or sufficient” (Yablo, 1992, p. 247, note 5). It is correct that we can talk about causally relevant events as well as causally relevant properties. But is not obvious that the sense of *causal relevance* is the same in both cases. An *event* can be



relevance of properties. It does not exclude the relevance of mental properties, since it does not say that physical properties are sufficient—whatever that might mean. It says that every physical event has a sufficient physical cause. But it is an *event* that is sufficient, not an event's having a certain property rather than another.<sup>15</sup>

In conclusion we can say the causal exclusion argument is not decisive against the token-identity version of non-reductive physicalism. Opponents, however, will object that even if the response is correct, it shows only that it is possible that some events have their effects in virtue of their mental properties—it shows only that mental properties can be causally *relevant*. What needs to be shown, though, is that the mental is causally *efficacious*.

What kinds of things can be causally efficacious? Causes, I suggest, and only causes are causally efficacious. And as the standard view has it, events and only events are causes. Instantiations of properties, on the other hand, can have two different causal roles. Firstly, as being the property *of* an event—as being instantiated by an event—a property can be causally *relevant* in the sense the event causes the effect in virtue of having the property. Secondly, as *being* an event the instantiation of a property—by a substance, at a time—can be causally *efficacious*.

On the present version of the argument, mental event tokens *are* physical event tokens. That is, the instantiation of the mental property and the instantiation of the physical property do *not* constitute two *distinct* events, since they are instantiated by one and the same event. Given the distinctions just made, it follows that instantiations of mental properties can at best be causally *relevant*—they are not and cannot be *causes*. And that means that they cannot be causally efficacious. Further, instantiations of physical and mental properties cannot be causal rivals, since causal exclusion or overdetermination presupposes distinct *causes*. Now,

---

causally relevant in the extensional sense of being a *partial* cause. In that sense, events are causally relevant as causes. If one denies, as Yablo does, that properties can themselves be causes, it remains to be explained in what sense properties can be causally relevant. And whatever that sense is, it is different from the one in which events are relevant, and it is, therefore, not clear at all that the exclusion argument can be restated simply by substituting *property* for *event*, as Yablo suggests.

<sup>15</sup> Some may construe the closure principle as saying that the occurrence of any physical event can be *explained* in physical terms only (only in terms of events and their physical properties). But by reading the argument in that way, the opponent is again shifting the focus from causation to causal explanation. The closure principle is a metaphysical principle. It is about causes, not about causal explanations. It does not say that physical causes have their effect only in virtue of their physical causes, nor does it say that everything can be causally explained in terms of physical properties.

what does *not* follow from all that is that mental *events* are not, or cannot be, causally efficacious.

### Version Two

One the first version, mental and physical properties are properties of one and the same event. If we construe events as property instantiations, we obtain two further versions of the argument. On both versions, the properties are partly *constitutive* of events, rather than being instantiated by events. Kim, for instance, says that “the properties an event exemplifies must be sharply distinguished from its constitutive property (which is exemplified, not by the event, but by the constitutive substance of the event).”<sup>16</sup> According to that, properties can be related to events in at least two different ways. Either the property is instantiated by the event, or its instantiation by a substance at a time constitutes the event. In the former case the property modifies the event, in latter case it modifies the involved substance.

In some cases, however, it will be controversial whether a given property is constitutive of an event or not. In an example Kim compares the event *Sebastian’s stroll* with *Sebastian’s leisurely strolling*. Are there two events happening at the same time, or is there only one event has that more than one true description? On what is known as the Anscombe-Davidson view on the individuation of events, there is only one event that can correctly be described as a stroll and as a leisurely stroll.

According to the view that events are property instantiations, however, whether there is only one event depends on whether *strolling* and *strolling leisurely* are constitutive properties. Further, if two properties constitute events, we must distinguish between two ways in which the events are related. For two events can be distinct, without being *entirely* distinct, as Kim says.<sup>17</sup> Sebastian’s stroll and Sebastian’s leisurely stroll, for instance, are distinct events. But not entirely distinct, since the latter event metaphysically *includes* the former.

Given that, we can now formulate the second version of non-reductive physicalism, according to which the instantiation of mental and physical properties constitute events that are distinct, but not entirely. Rather, physical events *include* mental events.

Kim did not try to spell out how the relation of inclusion between distinct events has to be understood; he says that the notion is intuitively plausible

---

<sup>16</sup> Kim, 1993, p. 43

<sup>17</sup> The distinction between distinct and entirely distinct events is supposed to block a counterintuitive proliferation of events. Compare *ibid*, pp. 42-46.

enough.<sup>18</sup> Stephen Yablo, though, has suggested the following. According to Yablo, the relation between mental and physical events is best understood as one between *determinables* and *determinates*, which, in turn, is best understood as an event's essence *subsuming* the other event's essence. What does that mean? Consider the determinate *being red* of the determinable *being coloured*. Being red is a specific—or *determinate*—way of being coloured. Similarly, Sebastian's leisurely strolling is a determinate of Sebastian's stroll, and the former subsumes the latter. The important thing to note is that the relation of subsumption—of one event's subsuming another event—is precisely the relation of inclusion. I will not go into any further detail of Yablo's account.<sup>19</sup> Rather, let us see how it can be applied to the problem of mental causation.

The lesson to be learned, according to Yablo, is that mental and physical events cannot causally exclude each other, if they stand in the suggested metaphysical relation of inclusion.<sup>20</sup> For we *know*, as Yablo says, that determinates and determinables are not causal rivals. Surely, by citing an object's being red we may be able to causally *explain* an event, or an event's having a certain property, which cannot be explained by referring to that object's being coloured. But it cannot plausibly be suggested that these properties are *rival causes*, simply because the particular instance of being red *just is* the particular instance of being coloured. If we apply this lesson to the problem of mental causation, we can then say with Yablo that “any credible reconstruction of the [problem] must respect the truism that determinates do not contend with their determinables for causal influence.”<sup>21</sup>

It does not matter whether or not Yablo was successful in analysing the relation of metaphysical inclusion correctly. For, on any account of that relation, if the physical event determines the mental event by including it, the two events cannot plausibly be causal rivals. To say they cannot be causal rivals is to say that they can neither exclude each other, nor overdetermine an effect, and we can conclude that

<sup>18</sup> Ibid, p. 45

<sup>19</sup> As indicated, Yablo suggests to construe the relation of inclusion as a relation between the event's essences: an event *e* is said to subsume or include another event *e\**, if the essence of *e\** is a subset of the essence of *e*. Compare Yablo, 1992, section 5, especially pp. 261-262.

<sup>20</sup> One may wonder how a physical event can *include* a mental event. The best way to think about it is in terms of multiple realisation. Assume that systems of type *S* and type *T* can be in mental states of type *M*, and that systems of the two types realise *M* in different ways. Then the states that realise *M* in systems of type *S* and *T* must be alike in certain respects, because they both realise *M*. But given that *M* is multiply realisable, we can assume that those states also differ in some respects; they realise *M* in different ways. Given that, we can say that the way in which the two systems realise *M* are specific or *determinate* ways of being in *M*.

<sup>21</sup> Ibid, p. 259

the causal exclusion argument does not apply to the second version of non-reductive physicalism.

### Version three

On the third version of the argument, instantiations of the mental and the physical properties constitute *entirely distinct* events. None of the responses that have been given so far apply. The mental and physical properties in question are constitutive of metaphysically distinct events. Instantiations of them can therefore be efficacious as causes. Furthermore, it seems that they can be causal rivals, since they constitute *entirely distinct* events.<sup>22</sup>

Yablo argued that, if two events are not entirely distinct, in the sense that one includes the other, then they cannot be causal rivals. It does *not* follow that two events can be causal rivals, if they are entirely distinct. Nor is it obvious that all causes that are entirely distinct causes of one and the same effect are, or can be, rival causes. In order to decide whether they are or can be rival causes, we need to have a closer look at some of the issues involved.

Intuitively, for causal rivalry, exclusion or overdetermination to occur, there have to be two or more *independent* causes of one and the same effect. If that is not the case, if the causes in question are not independent in the relevant sense, the exclusion argument does not apply, *because* the notions of causal rivalry, exclusion and overdetermination do not apply. In other words, whatever is necessary for causal rivalry, exclusion and overdetermination, is necessary for an application of the argument.

In the following I will suggest a characterisation of causal overdetermination and of the involved notion of independence. I will argue that mental and physical

---

<sup>22</sup>Non-reductive physicalists may argue that this way of construing the relation between mental and physical events is not compatible with non-reductive physicalism. According to non-reductive physicalism, mental events depend on physical events. This dependence, one may think, is incompatible with the claim that the events are entirely distinct. But is it impossible that two events are entirely distinct *and* dependent? Consider the following. Brutus killed Caesar by stabbing him with a knife. At one point, Kim suggests that both *Brutus' killing Caesar* and *Brutus' stabbing Caesar* are generic—hence, entirely distinct—events, whereas *Brutus' stabbing Caesar with a knife* merely modifies *Brutus' stabbing Caesar* (compare Kim, 1993, p. 44). Intuitively, though, there is a sense in which the particular killing of Caesar depends on Brutus' stabbing him. To decide whether the third version is consistent would require a detailed discussion of the involved notions of metaphysical distinctness and dependence, which is beyond the scope of this paper. Rather, I will assume, *for the sake of the argument*, that it is consistent. If it is not consistent, so much the worse for the proponents of the exclusion argument—for then the third version collapses into either version one or two.

events, construed as entirely distinct events, cannot overdetermine their effects, because they are not independent in the required sense. Given that, it is then the opponent's burden to clarify why and how the exclusion argument applies to non-reductive physicalism and to the problem of mental causation in general. Let me begin with two observations.

Let us consider a clear case of causal exclusion; a case in which two causes overdetermine an effect. An example that is often used to illustrate causal overdetermination is the case in which two sharpshooters kill their victim, which happens to be one and the same person, at exactly the same time. Each shot is sufficient to cause the victim's death, which is overdetermined by two distinct sufficient and independent causes.

The first observation is that the events in the case of mental causation are not distinct in the same way as the relevant events in the sharpshooter case. In the sharpshooter case, the events are not only distinct with respect to the instantiation of properties, but they are distinct in the further sense that the properties are instantiated by distinct *substances*. This, of course, is not the case for mental causation; the mental and the physical property are instantiated by one and the same substance. That does not show, of course, that being instantiated by distinct substances is necessary or sufficient for causal rivalry. But it *suggests* that the event's being entirely distinct is not sufficient. What is required, in addition, is that they are *independent*, in some sense.

This brings us to the second observation. What we are assuming, and what the opponents of non-reductive physicalism usually assume as well, is that the two purported causes - the mental and the physical event - stand in some intimate relationship. All versions of non-reductive physicalism assume that mental events supervene, in some sense, on physical events, and that mental events are realised by them. It is assumed, in other words, that mental events (or properties) are, in some sense, dependent on physical events (or properties). *By hypothesis*, that is, the mental and their underlying physical events (or properties) are *not* independent.

This independence, however, concerns the *existence* of mental events and properties. Whenever there is a mental event, there is some physical event that realises it. It is possible that there are physical events and no mental events, but it is impossible that there are mental events and no physical events. The mental depends on the physical and is realised by it, but not *vice versa*.

What we are interested in, however, is independence of *causes* as a criterion for the application of the causal exclusion argument. Recall that the third assumption of the causal exclusion argument concerns causal overdetermination. The argument applies only if the notion of causal overdetermination applies. That is, for the argument to apply, there must be causes that *can* overdetermine an effect.

I suggest having a closer look at the notion of causal overdetermination in order to get better a grasp of the relevant notion of independence. Let us begin with a

characterisation of overdetermination, which is adopted from Trenton Merrick's *Objects and Persons*.<sup>23</sup>

Some effect *e* is *causally overdetermined* if:

- (a) *e* is caused by *c*,
- (b) *c* is causally irrelevant as to whether some other numerically distinct cause, *c\**, is a cause of *e*,
- (c) *e* is caused by *c\**, and *c\** is numerically distinct from *c*.

Whereby *c* is *causally irrelevant* as to whether *c\** is a cause of *e* if:<sup>24</sup>

- (d) *c* is not a cause of *c\** (that is, *c* is neither a sufficient nor a partial cause of *c\**, and *c* is not an intermediate in a causal chain that runs from a cause of *c\** to *c\**), or
- (e) *c* does not cause *c\** to cause *e* (nor does it cause any members of *c\** to cause *e*, if *c\** consists of more than one cause jointly causing *e*).

One can agree with Merricks that this extensional construal of overdetermination is straightforward. But, unfortunately, it does not help us any further for two reasons. Firstly, Merricks talks about *objects* as being causes.<sup>25</sup> For overdetermination to occur, two distinct causes have to cause the effect, and distinct causes are, on that view, distinct objects. That means that rival causes—causes that can overdetermine an effect—are numerically distinct substances. That reflects my intuition that we get a clear sense of the phenomenon of overdetermination, if the causes are associated with distinct substances. But since it does not cover the case in which two distinct *events* occurring *in* one and the same substance, it is of no help.

Secondly, the definition recognises only *causal* relations between the two potentially rival causes; what matters, according to clause (b), is whether one cause is causally relevant or irrelevant as to whether the other cause brings about the effect. What we are looking for is a characterisation of the way in which the two causes of the effect, *c* and *c\**, can said to be dependent or independent causes. The presented definition captures the dependence between them in terms of causal relevance; that is, the dependence is construed as *causal* dependence.

Some philosophers have tried to understand the relation between mental states and their physical realisations in causal terms. Most philosophers, however, think

<sup>23</sup> Compare Merricks, 2001, p. 58, who thinks that a definition along such lines is ‘the most literal, straightforward and natural’ definition of causal overdetermination.

<sup>24</sup> Compare *ibid*, p. 57. Note that the involved notion of causal relevance is different from the notion of causal relevance as it applies to properties. Compare also note 14.

<sup>25</sup> It is not obvious whether the view is committed to substance-causation. One might hold, for instance, that causation by objects is causation by change in objects or by the object’s having of a certain property.

that the dependence between the mental and the physical is of a different kind. The notion of supervenience, as many think, provides merely a minimal constraint for that relation, which is itself in need of explanation (given that it holds). A popular solution is to say that supervenience holds, because the physical states *realise* the mental states.<sup>26</sup> What is needed, then, are dependence-conditions that can be applied to different relations between the physical and the mental, such as causation, supervenience, determination and realisation. An obvious candidate for that is *counterfactual* dependence.

It is commonly assumed that there is a close connection between causal dependence and counterfactual dependence. Causal relations are said to entail or support relations of counterfactual dependence. On Merricks' definition, *c* and *c\** do *not* overdetermine *e*, if *c* is causally relevant as to whether *c\** causes *e*. If *c* is causally relevant in that way, then whether *c\** causes *e* depends causally on *c*. This causal dependence entails or supports counterfactuals of the following form:

(CF) Given relevantly similar circumstances, had *c* not occurred, *c\** would not have caused *e*.

My suggestion is to replace the causal dependence employed in Merrick's definition by counterfactual dependence. For what is crucial, it seems, is not the causal connection itself, but the entailed counterfactual. Consider again the case of the two sharpshooters. It is a case of overdetermination, because there are two sufficient and independent causes of the same effect. Specified in causal terms, the independence consists in the absence of causal connections between the two causes. Why is that relevant to question whether the effect is overdetermined? If one shooting depends causally on the other shooting, as one might say, then the two shootings are not independent causes of the victim's death. But that is plainly circular, given that we want know in virtue of what causes are independent causes. A better answer is the following. If the two shootings are causally dependent, then one of the two sharpshooters would not have killed the victim, had the other one not done so.

The case of the two sharpshooters is a case in which it is a *coincidence* that two causes are sufficient to cause the same effect. A causal connection would render the two shootings non-coincidental. But so would a counterfactual connection. That is why I suggest replacing the causal condition on overdetermination by a counterfactual one. Accordingly, the two causes, *c* and *c\**, do *not* overdetermine the effect *e*, if *c\**'s causing *e* is counterfactually dependent on *c*; that is, if CF holds.

---

<sup>26</sup> Compare Kim, 2000, pp. 23-24.

We found Merrick's definition of causal overdetermination to be too narrow, since it covers only cases in which two potentially rival causes are causally dependent. To construe the kind of dependence that is necessary for causal rivalry and overdetermination in terms of counterfactual dependence has the advantage that it covers causal as well as other relationships that might hold between events. In order to see whether the counterfactual approach will get us any further with our problem, let us see whether the relation of realisation—as it is usually employed by non-reductive theories—supports counterfactuals of the right sort.

Suppose that  $p$  is the physical realisation of the mental event  $m$ , and that both  $p$  and  $m$  are causes of the event  $e$ . Is it true that had  $p$  not occurred,  $m$  would not have caused  $e$ ? Let us assume that  $m$  is multiply realisable.<sup>27</sup> We have then to distinguish between two cases. In the first case we assume that each system—or agent—falls under a certain type—or species—such that for all mental event-tokens  $m$  of type  $M$ : for any agent of a certain type, tokens of  $M$  are realised by physical event-tokens  $p$  of type  $P$ . In that case, we can limit our considerations to agents of certain types, and restricted counterfactuals of following form will hold.

(CF<sup>1</sup>) For all agents  $s$  of type  $S$ : had the  $s$ -involving event  $p$  not occurred, the  $s$ -involving event  $m$  would not have caused  $e$ , given relevantly similar circumstances.

That counterfactual holds, because the occurrence of  $m$  depends counterfactually on  $p$ . If the antecedent holds, then the mental event  $m$  does not occur—and if  $m$  does not occur, then, trivially,  $m$  does not cause  $e$ .

In the second case, we do not assume that mental events of a certain type are realised by all agents of a certain type in the same way—realised by physical events of the same type. In that case we have to consider the set  $\{P_1, P_2, P_3 \dots\}$  of all possible realisations of the mental state type—across all individuals of all agent-types. In that case, counterfactuals of the following form will hold.

(CF<sup>2</sup>) For any agent  $s$ : had none of the  $s$ -involving events  $p_i$  (of type  $P_i$ ) occurred, the  $s$ -involving event  $m$  would not have caused  $e$ , given relevantly similar circumstances.

Again, that counterfactual holds, because the occurrence of  $m$  depends counterfactually on  $p_i$ . If the antecedent holds, then the mental event  $m$  does not

---

<sup>27</sup>Note that non-reductive physicalism is typically motivated by the possibility of multiple realisation of mental states. It is therefore safe to make that assumption. (Moreover, if relevant counterfactuals hold in case mental states are multiply realisable, they certainly hold in case they are not.)



occur—and if  $m$  does not occur, then, trivially,  $m$  does not cause  $e$ . What we get, in both cases, is a counterfactual dependence between  $p$ ,  $m$ , and  $m$ 's causing  $e$ .

Given all that, we can conclude that mental events and their physical realisations do *not* overdetermine their effects since they are not independent causes. The mental event is not an independent cause, because its occurrence and its causing the effect depends counterfactually on its physical realisation. No dependency of that sort holds for standard examples of causal overdetermination. The two shootings do not depend on each other—neither causally nor counterfactually. What makes the sharpshooter case a case of causal overdetermination is precisely the fact that had one sharpshooter not killed the victim, the other one would have. They kill their victim independently of each other—and that they do so at the same time, with the same success, is a mere coincidence.

## Conclusion

I argued that the causal exclusion is not decisive against non-reductive physicalism, no matter whether the mental and physical events in question are construed as token-identical, distinct or entirely distinct. What is important to note is that the response to the last version applies to all three versions, since it relies on the claim that mental events are dependent on physical events—a claim that all versions of non-reductive physicalism are committed to.

What causal exclusion and rivalry amount to is fairly straightforward in case there are distinct and independent causes. But the case of mental causation is not of that sort. It is not clear at all what causal exclusion and rivalry amount to, given that the mental depends on the physical. There would have to be a kind of causal exclusion that is either not tied to the outlined notion of overdetermination, or one that is not tied to any notion of overdetermination at all. In any case, it remains to be shown what sort of causal exclusion that is and how it has to be understood. We can conclude that the opponents of non-reductive physicalism have yet to show how the causal exclusion argument applies to mental causation, even if it is assumed that mental events and their physical realisations are entirely distinct.

## References

- Beckermann, Ansgar (1992) 'Introduction – Reductive and Nonreductive Physicalism', in Beckermann *et al* (eds.) *Emergence or Reduction?*, New York : Walter de Gruyter, 1-21.
- Crane, Tim (1995) 'The Mental Causation Debate', *Proceedings of the Aristotelian Society*, Supplementary Volume LXIX, 211-236.
- Davidson, Donald (1980) *Essays on Action and Events*, Oxford: Oxford University Press.
- (1993) 'Thinking Causes', in Heil & Mele (eds.) *Mental Causation*, Oxford: Clarendon Press, 3-18.
- Horgan, T. & Tye, M. (1988) 'Against the Token Identity Theory', in Lepore & McLaughlin (eds.) *Actions and Events*, Oxford: Blackwell, 427-443.
- Kim, Jaegwon (2000) *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*, Cambridge: MIT Press.
- (1993) *Supervenience and Mind: Selected Philosophical Essays*, Cambridge: Cambridge Studies in Philosophy.
- Lowe, E. J. (2003) 'Physical Causal Closure and the Invisibility of Mental Causation', in Walter & Heckman (eds.) *Physicalism and Mental Causation*, Exeter: Imprint Academic, 137-154.
- Menzies, Peter (2003) 'The Causal Efficacy Of Mental States', in Walter & Heckman (eds.) *Physicalism and Mental Causation*, Exeter: Imprint Academic, 195-224.
- Merricks, Trenton (2001) *Objects and Persons*, Oxford: Clarendon Press.
- Yablo, Stephen (1992) 'Mental Causation', *Philosophical Review*, Vol. 101, 2, 245-280.