

# The Causal Closure Argument is No Threat to Non-Reductive Physicalism

*Peter Menzies*

## ABSTRACT

Non-reductive physicalism is the view that mental events cause other events in virtue of their mental properties and that mental properties supervene on, without being identical to, physical properties. Jaegwon Kim has presented several much-discussed arguments against this view. But the much simpler causal closure argument, which purports to establish that every mental property is identical to a physical property, has received less attention than Kim's arguments. This paper aims to show how a non-reductive physicalist should rebut the causal closure argument. A crucial premise in the argument is a principle stating that the physical world is causally closed. It states, roughly, that every physical event has a physical sufficient cause. I argue that when the principle is formulated in this way (and the other premises are given a natural reading), the causal closure argument is invalid. Less frequently, the causal closure principle is formulated in terms of a difference-making conception of causation so that it states that every physical effect has a physical difference-making cause. I argue that the principle, so formulated, is false and the causal closure argument unsound. Either way the argument lacks the apodeictic force to compel acceptance of its conclusion.

Keywords: causal closure argument, difference-making theory of causation, non-reductive physicalism, overdetermination, sufficient causes

## 1. Introduction

Non-reductive physicalism about the mind consists in several characteristic theses: in virtue of the mental properties they instantiate, mental events cause other events, both mental and physical; and the relevant mental properties supervene on physical properties without being identical to them. Different forms of non-reductive physicalism are expressed in terms of different conceptions of

events.<sup>1</sup> To simplify matters I shall formulate non-reductive physicalism in terms of Jaegwon Kim's (1973) fine-grained conception of an event as an instantiation of a property by a particular at a time. One virtue of this conception is that it makes it clear how a mental event can cause another event in virtue of the mental property it instantiates: the relevant mental property is the property whose instantiation by a particular at a time constitutes the event. On this conception, a mental event is identical with a spatiotemporally coincident physical event just in case their constitutive mental and physical properties are identical.

Non-reductive physicalism has come under attack in recent years, most notably from Jaegwon Kim, who has criticized the view as unstable and liable to collapse under pressure into causal epiphenomenalism or reductive physicalism. Kim's exclusion and supervenience arguments (1998; 2005; 2009) have been exhaustively discussed in the literature. But the argument that is, arguably, the basis of Kim's arguments has not been so subjected to the same degree of critical scrutiny. This is the causal closure argument. It differs from Kim's exclusion and supervenience arguments in subtle ways,<sup>2</sup> but poses an equally serious threat to non-reductive physicalism. Non-reductive physicalism needs to have a good answer to the causal closure argument if it is to avoid its unpalatable conclusion that mental properties are identical to physical properties.

The causal closure argument derives from the argument that Lewis (1966) originally gave for the identity theory of mind, but eliminates some inessential features of their argument.<sup>3</sup> Here is David Papineau's (2002, pp. 17–28)

<sup>1</sup> Non-reductive physicalism is sometimes said to be compatible with token-physicalism – the doctrine that mental events are identical with physical events. Donald Davidson's anomalist monism (1970), for example, was an early form of non-reductive physicalism that incorporated a conception of events as concrete particulars that allowed one and the same event to be both mental and physical. However, more recent forms of non-reductive physicalism reject Davidson's token-physicalism on the grounds that his coarse-grained conception of events and his conception of causation as a relation between such events do not allow one to capture the intuitively compelling idea that mental events cause other events in virtue of their properties. See discussion of this issue in McLaughlin (1989) and Davidson (1993); recent non-reductive physicalists have tended to agree with McLaughlin's criticisms of Davidson's position. In any case, I shall be concerned in this paper to defend the more recent versions of non-reductive physicalism that reject token-physicalism.

<sup>2</sup> One way in which the causal closure argument differs from Kim's exclusion and supervenience arguments is that it does not employ a causal closure assumption that is crucial to these arguments. Another difference is that the causal closure argument is a direct argument for the identity of mental and physical properties, while the exclusion and supervenience arguments are *reductio* arguments directed at the assumption that mental events cause other events. Nonetheless, all these arguments fail for related reasons. For critical discussion of Kim's exclusion argument, see Shapiro and Sober (2007), Woodward (2008); List and Menzies (2009).

<sup>3</sup> Modern versions of the causal closure argument simplify Lewis' argument by taking its first premise about the causal efficacy of the mental to be an *a posteriori* truth about the world rather than an *a priori*

formulation of the argument. (He calls it «the causal argument based on the completeness of physics»).

1. Mental events cause physical<sup>4</sup> effects. (The causal efficacy of the mental).
2. Every physical effect has a physical sufficient cause.<sup>5</sup> (Causal closure of the physical).
3. The physical effects of mental events are not systematically overdetermined. (No causal overdetermination).
4. Therefore, the mental events must be identical with physical events.

To see how the argument works, suppose that a mental event (say, a conscious decision to throw a ball) causes some physical effect (the bodily movement of throwing of the ball). The causal closure principle tells us that there is a purely physical sufficient cause for this physical effect (say, an event occurring in the agent's brain). It appears, then, that we have two causes of the effect – a mental cause and a physical cause. But the no causal overdetermination principle tells us that a physical effect like a bodily movement is not systematically overdetermined by two causes each sufficient to bring it about. So we must conclude that the only way in which all the premises can be true together is if the mental event is identical with the physical event. But if these events are identical, it must be because their constitutive mental and physical properties are identical.<sup>6</sup>

Papineau claims that the argument is deductively valid and its premises are all true. He believes that the first premise about the causal efficacy of the conscious mind is a central part of commonsense and so very hard to give up: for example, epiphenomenalists who reject the premise must posit 'causal danglers' – ontologically distinct events with causes but no effects – contravening a principle of theory choice that enjoins us to choose the simplest and most ontologically parsimonious theory. Papineau argues that the second premise

truth implied by conceptual definitions of mental properties in terms of their causal roles. On this point see Papineau (2002, pp. 38–40).

<sup>4</sup> Papineau argues that one can characterize 'physical properties' informatively in negative terms as inanimate properties that are not identifiable as mental or biological properties. I shall adopt this characterization of 'physical' in this paper.

<sup>5</sup> The causal closure principle can be reformulated to accommodate physical events without deterministic causes: i.e. every physical effect has a physical event that is sufficient to determine its chance of occurrence. But it facilitates exposition to work with the simpler formulation above.

<sup>6</sup> The identity of properties follows from the identity of events under the Kim conception of events because identical events must be constituted by the same particular, same property, and same time.

about the causal completeness of the physical world is far from being part of commonsense, but is, nonetheless, an empirical claim that has finally been established in the twentieth century on the basis of extensive scientific investigations. Finally, he claims that the third premise about no causal overdetermination is another expression of the standard principle of theory choice requiring the choice of the simplest and most ontologically parsimonious theory.

The argument represents a powerful challenge to non-reductive physicalism in that its conclusion directly contradicts a central tenet of the position.<sup>7</sup> Most non-reductive physicalists have attempted to rebut the argument by denying the no causal overdetermination premise. (See, for example, Baker, 1993, 2009; Bennett, 2003; Crisp and Warfield, 2001; Loewer, 2007; Pereboom, 2002; Shoemaker, 2007). They concede that every bodily movement is systematically overdetermined by both a mental sufficient cause and a physical sufficient cause, but they argue that this kind of overdetermination is not the problematic kind that violates principles of ontological parsimony because the systematic character of the overdetermination is explained by the fact that the mental cause always supervenes on the physical cause. While the elaborations and defences of this position are ingenious, they do not, to my mind, completely dispel the impression that their accounts of the etiology of bodily movement involve a problematic redundancy: the physical sufficient cause of a bodily movement seems to be a superfluous feature of the causal history of the bodily movement because even if it had been absent, the mental sufficient cause would have brought about the bodily movement – though, of course, it would have had a different physical supervenience base. These currently popular defences of non-reductive physicalism do not question the causal closure principle or the validity of the argument.

In contrast, I shall attempt in the first place to rebut the causal closure argument by claiming that the argument is invalid. My rebuttal will turn on the causal closure principle and, in particular, its formulation using the concept of a physical sufficient cause. This principle is rarely scrutinized because the concept of a 'sufficient cause' is taken to be self-evidently clear. What could be

<sup>7</sup> It is noteworthy that the argument also represents a powerful challenge to property dualism as well. Many property dualists try to circumvent the argument by giving up the premise that mental events cause other events. For example, David Chalmers (1996, pp. 150–160) is property dualist who explores this response, though he is reluctant to call his position epiphenomenalist. But he concedes that it looks like epiphenomenalism; and he attempts to show that the arguments usually advanced against this position are not as powerful as they appear to be. I note that he does not, for a moment, question the validity of the argument, nor the truth of the premise of the causal closure of the physical world.

clearer than the concept of a cause that is sufficient to bring about its effect? But the concept is very far from clear, I shall claim. It is most frequently interpreted in terms of the concept of nomological sufficiency, or a concept of primitive production. In criticism of these interpretations, I shall outline arguments, very familiar in the causation literature, that show that these concepts fall well short of being genuinely causal concepts. Consequently, we will see the argument involves a miscellany of concepts: the first and third premises are couched in terms of the genuine causal concept while the second premise is couched in terms of the non-causal concept of a 'sufficient cause' masquerading as a causal concept. I shall argue, further, that if we interpret the first and third premises in terms of a genuine causal concept – a difference-making conception of causation – we can see that the argument is fallacious. For, even granting that the premises are true under these interpretations, one can consistently embrace a non-reductive physicalist conception of mental properties as supervenient on physical properties without being identical to them.<sup>8</sup>

Of course, those who advocate the causal closure argument think that the concept of a sufficient cause is a causal concept. Doesn't its very name suggest that it is? But labeling a concept 'causal' does not make it so. We have to consider the content of the concept, as it is standardly used, to see whether it merits the name; and I will argue that it does not. Even so, it might be thought that the causal closure argument can be easily modified to frame the causal closure in explicitly causal terms. Let us suppose that we replace the causal closure principle above with the following:

5. Every physical effect has a physical cause. (Modified causal closure of the physical).

The difference is that this formulation says that there are physical *causes* for physical effects, whereas the original formulation says there are physical *sufficient causes* for physical effects. With the causal closure reformulated in this way, there is no question that the causal closure argument is valid.<sup>9</sup> But now I shall ar-

<sup>8</sup> In arguing that the causal closure argument is invalid, I shall adopt the standard non-reductive physicalist view that mental properties are multiply realizable by physical properties, or have multiple possible physical supervenience bases. This is a controversial view, which has been recently contested by Lewis (1980), Kim (1992), Bechtel and Mundale (1999), Shapiro (2004), Bickle (2003). However, there is not space in this paper to explain how non-reductive physicalism should answer these recent criticisms of the multiple realizability arguments. But see Gillett (2003), Aizawa (2009); and Aizawa and Gillett (2009).

<sup>9</sup> Strictly speaking, the argument is valid only if we add the additional assumption that every mental event has a physical effect. One has to add this extra assumption because the argument, as it stands, fails to apply to mental events that do not have physical effects. I shall make this additional assumption

gue that the truth of the reformulated causal closure premise is highly dubious. We have good reason to think that many physical bodily movements have mental events as their causes. But, as I shall show by drawing on some logical results of List and Menzies (2009), when a mental event is a difference-making cause of a physical effect in a way that does not depend on the particular way it is realized, then the most plausible candidate for a physical cause – the neural event that subvenes the mental event – cannot also be a difference-making cause of this effect. I shall go on to argue that we have good evidence that mental events are realization-insensitive causes of bodily movements, so that we can justly conclude that there are physical effects for which there are no physical difference-making causes, in contradiction to the modified causal closure principle.

One loose way of expressing the point of this paper is to say that it attempts to establish that the causal closure argument suffers from a fallacy of equivocation. On the one hand, in order for the pivotal premise concerning the causal closure principle to be true, it must be understood in terms of the erroneously-named concept of a sufficient cause. On the other hand, in order for the argument to be valid, the causal closure principle must be understood in its modified form that appeals to a genuine concept of causation. Hence, the argument appears to be sound – valid with true premises – only because it equivocates in the interpretation of the causal closure principle.<sup>10</sup>

My plan of action is simple. In section 2 I explain why the concept of a sufficient cause is not a genuine causal concept and introduce a counterfactual difference-making condition that stands a much better chance of capturing the concept of causation. I use it to show how the causal closure argument that employs the original, unmodified causal closure principle is fallacious. In section 3 I reconsider the argument in a reformulation that employs the modified version of the causal closure principle. I introduce List and Menzies' (2009) concept of a realization-insensitive difference-making cause and use it to show that there are clear counterexamples to the modified causal closure principle. I conclude with a brief homily about the need to employ substantive accounts of the concept of causation in debates about the metaphysics of mind.

throughout the discussion.

<sup>10</sup> Other philosophers have criticized the causal closure argument on somewhat different grounds. For example, Bishop (2006) objects that the causal closure argument relies on a hidden premise to the effect that the only efficacious causes are physical causes, which in effect begs the question in favour of physicalism. Montero (2003) objects that the inductive advanced in favour of the causal closure principle does not warrant its acceptance.

## 2. The Argument in its Original Formulation is Fallacious

Let us consider the causal closure argument in its customary formulation with the causal closure principle expressed in terms of physical sufficient causes. To repeat, the principle states that every physical effect has a physical sufficient cause. But what exactly is a sufficient cause? Papineau, for instance, provides no elucidation of the concept, presuming that the meaning of the concept is self-evident. So we have to look elsewhere for clarification of the concept.

The most obvious meaning that philosophers have in mind when they talk about sufficient causes is that sufficient causes are events instantiating properties that are nomologically sufficient for the properties instantiated by the effect events. More precisely, an F-event is a sufficient cause of a G-event just in case the property F, in conjunction with laws of nature and other properties instantiated in the background conditions, logically implies property G. Such a conception fits in with a deductive-nomological conception of causation, which often lies in the background of many discussions of mental causation. But this conception's influence is completely unwarranted and philosophers of causation have long lamented its distorting effects. (See Lewis, 1973a; Woodward, 2003, 2008).

This conception of causation is known to have many defects, but this is not the place to catalogue all of them. So let me focus on one particular type of problem faced by the conception – the problem of causal irrelevance – that has been well known since Wesley Salmon's important critique (1971). Salmon argued that in many examples a property F, in conjunction with laws L and background properties K, logically implies a property G, but F is completely causally irrelevant to G. He gave some nice examples to illustrate the problem. For example, on the deductive-nomological model, a man's taking contraceptive pills is a cause of his failure to get pregnant since his taking the contraceptive pills, together with the appropriate law, implies that he does not get pregnant. Similarly, on the model, casting a dissolving spell on an aspirin pill is a cause of its dissolving when immersed in water because casting a dissolving spell on an aspirin pill, together with the appropriate law, implies that it dissolves on immersion in water. In each of these cases, the property that logically implies the effect property is causally irrelevant to it, but nonetheless satisfies the conditions for being a deductive-nomological cause.<sup>11</sup>

<sup>11</sup> Salmon assumes that these causal explanations fit the deductive-nomological model because there are laws to the effect that every male who takes a contraceptive pill fails to become pregnant and that every aspirin pill with a dissolving spell cast on it dissolves when immersed in water. It may be objected that the

The problem of causal irrelevance is not a superficial problem that is easily corrected. Its ramifications are widespread. For example, another kind of problem for the deductive-nomological model of causation – the problem of correlation due to a common cause – is a special instance of the problem of causal irrelevance. Consider the kind of case in which the presence of a virus in a person's body causally determines that the person will develop a rash and later a high temperature. So, in the situation in which a person has the virus, his having the rash is a nomologically sufficient for his developing a high temperature, and so, on the deductive-nomological model, the first counts as a cause of the second. However, the rash is causally irrelevant to the high temperature, as they are joint effects of a common cause – the virus. Or take a slightly different example: suppose that a car travelling along a road casts a shadow on the side of the road. Given the presence of the car at one point in the road at time  $t_1$ , the presence of the shadow at  $t_1$  is nomologically sufficient for the presence of the shadow at  $t_2$  further along the road. Hence, on the deductive-nomological model of causation, the shadow of the car at time  $t_1$  is a sufficient cause of the shadow at time  $t_2$ . But clearly the presence of the shadow at one time is causally irrelevant to its presence at a later time.<sup>12</sup>

Perhaps these counterexamples simply show that the concept of a sufficient cause should not be understood according to the deductive-nomological model of causation. Perhaps this concept should be understood in accordance with a model of causation espoused by Anscombe (1971) that takes causation to consist in a primitive relation of production or generation. On this understanding, a sufficient cause is simply an event that produces or generates another. Jaegwon Kim (2010), who often appeals to the concept of a sufficient cause in his discussions of mental causation, suggests that we should understand the concept in this way. He says (2010, pp. 254–258) that that the important feature of pro-

generalizations involved do not represent genuine laws. In response to this objection, Salmon notes that the generalizations satisfy many of the criteria for law-like generalizations: they are universal, they have modal force, and they are as explanatory as the generalizations they are based on. The most obvious way to exclude them as not law-like is to require that law-like generalizations do not contain any causally irrelevant properties. However, this is not a satisfactory move for a deductive-nomological model of causation that purports to be analyzing the causal concept in terms of independently understood non-causal concepts.

<sup>12</sup> It might be argued that these examples are counterexamples to the deductive-nomological model of causation only if the presence of the common cause is held fixed as part of the background conditions in the deduction. If its presence could be excluded from the background conditions, then the examples would not fit the deductive-nomological template. An exclusion clause might read that the background conditions should not include any factors that are causally independent of the putative cause property but causally relevant to the effect property. But any such exclusion clause would run up against the fact that the deductive-nomological account of causation is usually presented as a reductive analysis.



ductive sufficient causes is that they conform to a locality condition: a productive cause is linked to its effect by a spatiotemporally continuous sequence of causal intermediaries. Further, he suggests that productive causal relations are «constituted by phenomena like energy flow and momentum transfer, an actual movement of some (conserved quantity)», referring in a footnote to conserved physical quantity theories of causation<sup>13</sup> as suitable accounts of production.

There is, of course, an immediate problem if Kim is suggesting that the concept of a sufficient or productive cause is conceptually linked to the concept of the transfer of some conserved physical quantity. For this effectively begs the question against the possibility there are conscious mental causes, ontologically distinct from physical events, since such mental events by their nature could not transfer any physical quantity. However, if Kim is suggesting that the concept of a productive causation is conceptually primitive and that productive causal relations involve, as a matter of empirical rather than analytic fact, the transfer of a physical quantity, then his account avoids the question-begging charge. Nonetheless, his suggestion does not help one iota to solve the problem of causal irrelevance.

A theory of sufficient causes would solve the problem of causal irrelevance if it could sort out the causally relevant properties of sufficient causes from the causally irrelevant ones. But the account of sufficient causes in terms of production, at least as proposed by Kim, cannot do this. For example, notice that the event of Mr Jones' taking the contraceptive pill is a sufficient or productive cause of his failing to get pregnant under Kim's proposal, as there are continuous processes – involving the ingestion of the pill, the metabolizing of its chemicals, and other ongoing physiological processes – leading to his failure to become pregnant. We might even conjecture that these processes, on further empirical investigation, will be shown to involve the transfer of some conserved physical quantities. Nonetheless, the fact that Mr Jones' taking the contraceptive pill is a productive cause of his not getting pregnant does not alter the fact that the first event is causally irrelevant to the second. Similarly, observe that casting a dissolving spell on an aspirin pill as it is immersed in water initiates spatiotemporally continuous processes, perhaps even conserved quantity-preserving processes, leading to the aspirin pill's dissolving in the water. But again the fact that one event produces the other does not change the causal irrelevance of the first event to the second.<sup>14</sup> Once more, the point

<sup>13</sup> See Fair (1979), Salmon (1994), and Dowe (2000).

<sup>14</sup> The literature in the philosophy of causation has long recognized that production theories of causation

is made that ‘sufficient causes’, even under Kim’s account of them in terms of a primitive relation of production, need not be genuine causes – genuine in sense of being causally relevant to their effects.

As many philosophers of causation see the issue, the notion of causal relevance goes beyond what actually happens in the presence of the cause, which is the focus of accounts of sufficient causation. Rather it involves a contrast between what happens in the presence of the cause and what happens in the absence of the cause. Probabilistic theories try to capture this idea of causal relevance by requiring that the probability of the effect conditional on the presence of the cause should be greater than the probability of the effect conditional on the absence of the cause. Counterfactual theories keyed to deterministic causation require that the truth of two counterfactuals, one specifying that the effect would occur if the cause were to occur and the other specifying that the effect would not occur if the cause were not to occur. The essential idea expressed by these kinds of theory is that a cause *must make a difference* to its effect. It must do so in the sense that whether or not the cause occurs must affect the occurrence or non-occurrence of the effect, or at least the probability of its occurrence. In terms of this difference-making idea, we can see that Mr Jones’ taking the contraceptive pill is causally irrelevant to his not getting pregnant because his taking the pill makes no difference to the probability of his getting pregnant; and casting a dissolving spell on the aspirin pill is causally irrelevant to whether it dissolves in the water because the aspirin would have dissolved in the water whether or not the spell had been cast.

It will help make our discussion more concrete if we consider how a difference-making theory of causation might be formulated within a counterfactual framework. David Lewis’ (1973a) counterfactual theory of causation is one such theory.<sup>15</sup> In fact, Lewis offered three different versions of his counterfactual theory, the second two (1986, 2000) being refinements of his original theory. The very simplified difference-making theory of causation I will be discussing will be based on his original theory. At the heart of this theory was Lewis’s definition of the notion of causal dependence, corresponding to what I

in terms of conserved physical quantities face the problem of causal irrelevance. See, for example, Hitchcock (1995) and Woodward (2003, pp. 330–353).

<sup>15</sup> Lewis outlines the reasons why a difference-making account of causation is better formulated in terms of counterfactuals rather than conditional probabilities in his (1986, pp. 175–184). However, I conjecture that many of the arguments and logical results on which my defence of non-reductive physicalism relies could be reproduced within the framework of a probabilistic theory of difference-making causation.

have been calling causal relevance:

6. Event  $e$  causally depends on event  $a$  distinct<sup>16</sup> event  $c$  if and only if (a)  $c$  occurs  $\square \rightarrow e$  occurs; and (b)  $c$  doesn't occur  $\square \rightarrow e$  doesn't occur.<sup>17</sup>

Here the symbol ' $\square \rightarrow$ ' represents the counterfactual operator. The two counterfactuals in this definition are to be understood according to the possible world semantics developed by Lewis (1973b). This semantics expresses the truth conditions of counterfactuals in terms of a similarity relation between possible worlds. The similarity relation is represented by an assignment to each possible world  $w$  of a system of spheres of worlds centred on  $w$ , subject to certain constraints.<sup>18</sup> The idea is that the smaller a sphere is around  $w$ , the more similar are worlds to  $w$  are worlds in it. In terms of this framework, a counterfactual  $P \square \rightarrow Q$  is true in world  $w$  if and only if  $Q$  is true in all the closest  $P$ -worlds to  $w$ .

The semantic framework that I shall use differs from Lewis's in one respect: I adopt a weaker centring requirement than Lewis by allowing that the smallest sphere around  $w$  may contain more than one world. Lewis, by contrast, requires it to contain only  $w$ . A relaxation of Lewis's strong centring requirement is essential for my purposes. Strong centring implies that when  $P$  and  $Q$  are true in some world so is  $P \square \rightarrow Q$ . So if event  $c$  occurs and event  $e$  occurs in the actual world, clause (a) of the definition of causal dependence would hold automatically. But this makes the truth condition too weak. For the purposes of formulating a counterfactual account of causal difference-making, it is essential to require that the counterfactual  $P \square \rightarrow Q$  express the condition that  $P$  would be sufficient in the circumstances for  $Q$ . Accordingly, it is reasonable to weaken the centring requirement to permit the smallest sphere in the system of spheres

<sup>16</sup> The restriction of the definition of causal dependence to distinct events means that a counterfactual dependence between two non-distinct events (such as today's being Monday and tomorrow's being Tuesday) does not count as causal.

<sup>17</sup> For Lewis the concept of causation is more complex than the concept of causal dependence. In order to ensure the transitivity of causation and to handle some examples involving pre-emption, Lewis defines causation in terms of the ancestral of the relation of causal dependence. Since transitivity is not at issue in this paper and since I wish to set aside considerations about pre-emption and overdetermination as being too complex for this paper, I shall concentrate on Lewis's definition of causal dependence. This corresponds most closely to the concept of causal relevance, which we saw was missing from accounts of 'sufficient causation'.

<sup>18</sup> Nestedness: For any two spheres  $S$  and  $T$ , either  $S$  is included in  $T$  or  $T$  is included in  $S$ . Weak centring:  $w$  is contained in every sphere. Exhaustiveness: There is a largest sphere containing all possible worlds. Limit assumption: For any world  $w$  and any proposition  $P$ , there is a smallest sphere around  $w$  containing some  $P$ -world, called the *smallest  $P$ -permitting sphere around  $w$* . The *closest possible  $P$ -worlds* are defined as the  $P$ -worlds within the smallest  $P$ -permitting sphere.

centred, say, on the actual world to contain not just the actual world but many other worlds that count as relevantly similar. In this case, the truth of P and Q in the actual world does not automatically make  $P \Box \rightarrow Q$  true. What is required for its truth is that *every* closest-to-actual P-world is a Q-world.

Given the understanding we have reached about ‘sufficient causes’ versus difference-making causes, let us return to the discussion of the validity of the causal closure argument, in its original formulation. We are now in a position to see that the argument is invalid. Let us suppose that all its premises are true under the interpretations we have now assigned to ‘sufficient cause’ and to ‘cause’. In particular, let us suppose that we understand the first premise concerning the causal efficacy of conscious mental states as saying that conscious mental states can be difference-making causes of physical effects; we understand the second premise concerning the causal closure of the physical world as saying that for every physical event there is another physical event that is nomologically sufficient for it;<sup>19</sup> and we understand that third premise concerning no causal overdetermination as saying that the physical effects of conscious mental events typically do not have multiple difference-making causes. Does it follow from these suppositions that the conscious mental events must be identical to physical events? The answer is a decisive ‘No’.

It is easiest to see this in terms of a concrete example. Let us consider a particular scenario to see whether a non-reductive physicalist could consistently entertain the truth of all the premises of the causal closure argument, as they apply to the scenario, while denying its conclusion. Suppose that my conscious mental state of intending to throw a ball – symbolized as M – is a difference-making cause of a physical effect, say my bodily movement of throwing the ball – symbolized as P. In terms of the counterfactual rendering of the difference-making conception of causation, this means that the following pair of counterfactuals hold true:

7. (a)  $M \Box \rightarrow P$ ; and (b)  $\sim M \Box \rightarrow \sim P$ .

The non-reductive physicalist thinks that the conscious mental event M supervenes on some physical event in the brain without being identical to it. Suppose that in the particular situation the mental event M supervenes on the actual neural event N, where N is a highly specific property of my brain involv-

<sup>19</sup> Without loss of generality, I shall interpret ‘sufficient cause’ as an event that is nomologically sufficient for another. But my argument could also be run if ‘sufficient cause’ was interpreted to mean an event that produces another.

ing the precise neural networks activated in my intention to throw the ball. The non-reductive physicalist insists that *M* is not identical to *N*, at the very least because *M* could have different physical supervenience bases.

So let us suppose that premise 1 about the causal efficacy of conscious mental states holds true in this situation. What about premise 2 concerning the causal closure of the physical world? Let us suppose that, in keeping with this principle, there is a physical event that is nomologically sufficient for the physical event *P*. There is no reason why the physical event we pick as nomologically sufficient must automatically be the event *N*, which is the actual supervenience base for the mental event *M*. For instance, the physical event we pick out as nomologically sufficient for *P* may occur at a different time from *M* and *N*. However, the only coherent physicalist picture of the situation must represent the process leading up to *P* as including the physical event *N* in a sequence of physical events where each event is nomologically sufficient for the succeeding event and ultimately for *P*. For if *N* is not nomologically sufficient for *P*, then the first counterfactual of (7), namely  $M \square \rightarrow P$ , could not be true.<sup>20</sup> So without loss of generality, let us pick *N* to be the physical event that is nomologically sufficient for *P*, in conformity with the causal closure principle. As we shall see, this choice makes it harder rather than easier for the non-reductive physicalist to show the consistency of the imagined situation, as it imposes tighter constraints on its logical structure.

Now suppose that premise 3 about no causal overdetermination is true in this situation. We are interpreting this to mean that the physical effects of conscious mental state are not systematically overdetermined by difference-making causes. However, there is nothing in the situation, as we are now imagining, that would conflict with this principle. For we are imagining a situation in which the mental event *M* is a difference-making cause of the physical effect *P* and *N* – the actual supervenience base of *M* – is nomologically sufficient for *P*. Given the sharp distinction between nomological sufficiency and difference-making causation, we can see that this situation does not involve two difference-making causes of the physical effect *P*: just one difference-making cause and a different nomologically sufficient condition. Since the situation does not represent a violation of the no causal overdetermination principle, we can consistently suppose the third premise holds true of the imagined situation, along with the other two premises.

<sup>20</sup> If *N* is not nomologically sufficient for *P*, then we know that in the actual world *M* does not lead to *P*, so falsifying the 7(a) counterfactual  $M \square \rightarrow P$ .

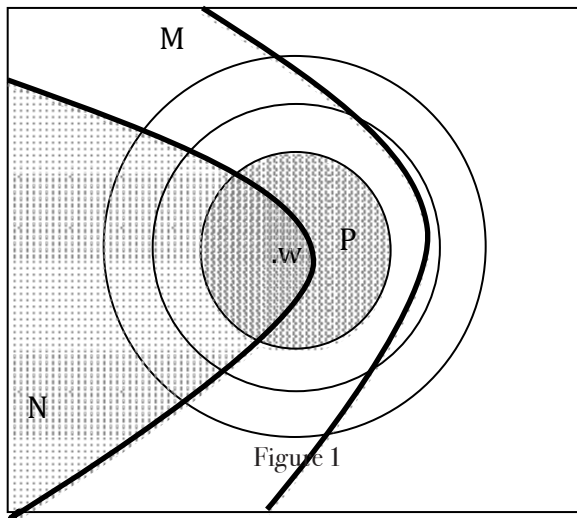
Does it follow from the description of this situation that the conscious mental event *M* must be identical to a physical event that is nomologically sufficient for *P*? If *M* were identical to any such physical event, it would have to be *N*, which occurs at the same time. But we have stipulated from the very start a non-reductive physicalist view of the situation, according to which *M* supervenes on *N* but is not identical to it. One very good reason for this, we are supposing, is that, while *M* has *N* as its actual suopervenience base, it could have had other different supervenience bases.

So we are now in a position to see that the situation we are considering is one in which the three premises of the causal closure argument are true but its conclusion false. So, the argument is invalid under the interpretation we have given of its relevant concepts.

A diagram may assist in seeing that the situation described above is indeed a consistent situation. Figure 1 below represents the universe of possible worlds with a system of spheres centred on the actual world *w*. The figure represents the set of *N*-worlds by the smaller convex region with lighter shading; the set of *M*-worlds by the larger convex region without shading; and the set of *P*-worlds by the innermost sphere with the darker shading. The figure reveals that *M* is a counterfactual difference-making cause of *P* since the two counterfactuals  $M \square \rightarrow P$  and  $\sim M \square \rightarrow \sim P$  are true: every closest *M*-world is a *P*-world and every closest  $\sim M$ -world is a  $\sim P$ -world. The figure also represents the fact that *N* is nomologically sufficient for *P* in terms of the fact that the every closest *N*-world is a *P*-world. But the figure demonstrates clearly that *M* is not identical to *N*, as they occupy different regions of logical space. Nonetheless, *M* supervenes on *N* in virtue of the fact that the set of *N*-worlds is included in the set of *M*-worlds. In this way the figure gives a clear pictorial representation of a consistent situation in which the premises of the causal closure argument hold but its conclusion fails to hold.

Once we recognize that causation is a difference-making relation quite different from ‘sufficient causation’, we can see that there is no inconsistency in saying that the mental event *M* is a difference-making cause of physical event *P*, that its subvening neural event *N* is nomologically sufficient for *P*, and that *M* and *N* do not causally overdetermine *P*. There is no inconsistency because causal overdetermination requires two causes and *N*’s nomological sufficiency for *P* does not qualify it as a cause of *P*. Consider a simple example that is analogous to our example of mental causation. Suppose that my flipping a light switch is the difference-making cause of the light’s going on, regardless of the way in

which I flip the switch, i.e. regardless of whether I flip it quickly or slowly. Notice that my flipping the switch *simpliciter* supervenes on my flipping the switch *quickly* in virtue of the fact that the latter entails the former; but that the events are not identical since my flipping the switch need not have happened quickly.<sup>21</sup> Notice also that my flipping the switch quickly is nomologically sufficient for the light going on in virtue of the fact that my flipping the switch *simpliciter* is nomologically sufficient for it. Further, observe that this is not a case of causal overdetermination, with my flipping the switch quickly counting as a cause alongside my flipping the switch *simpliciter*. My flipping the switch quickly does not count as a cause at all because one of the two difference-making counterfactuals relevant to the example is false: namely, the (b) counterfactual «If I had not flipped the switch quickly, the light would not have gone on» is false since some of closest possible worlds in which I do not flip the switch quickly, i.e. I flip it slowly, are worlds in which the light still goes on.



As the relationships involved in these situations are complex, it is useful to

<sup>21</sup> In this example, I am assuming that the property of flipping the switch *simpliciter* is numerically different from the property of flipping the switch *quickly* since they differ in extensions. The force of the example doesn't depend on an abundant versus a sparse conception of properties. The example can be reworked so that it concerns the way in which the causal powers of a sparse property like spin, mass, or charge differ from those of a property that logically implies such a sparse property.

illustrate my claims about consistency via a different scheme for representing possible worlds and counterfactuals from the traditional scheme.<sup>22</sup> Thus Figure 2 below represents the state space of a mind-brain system at two different times. At time  $t_1$ , the states of the system are partitioned into two macrostates corresponding to the occurrence and non-occurrence of event M. Each of these macrostates can be realized by many different microstates, each corresponding to a point in the state space. At time  $t_2$ , the states of the system are partitioned into two cells, one cell representing the occurrence of the physical event P and the other cell its non-occurrence. The arrows represent the lawful trajectories of the microstates over time. The figure represents just six of these trajectories, but every microstate at  $t_1$  can be mapped onto a single microstate at  $t_2$ . In the figure N represents the actual neural state realizing the macrostate M. The shaded area in the state space at  $t_1$  represents those states that are closest to the actual state in which M and N obtain. The figure represents the fact that M is a difference-making cause of P, since the counterfactuals corresponding to 7(a&b) are true: every closest M-state is mapped onto a P-state and every closest  $\sim$ M-state is mapped onto a  $\sim$ P-state. The figure also shows that while N is nomologically sufficient for P, it is not a difference-making cause of P since some of the closest  $\sim$ N-states are not mapped onto  $\sim$ P states. But, importantly for my purposes, the figure illustrates that we can suppose that macroevent M makes a difference to the subsequent physical event P *consistently* with suppos-

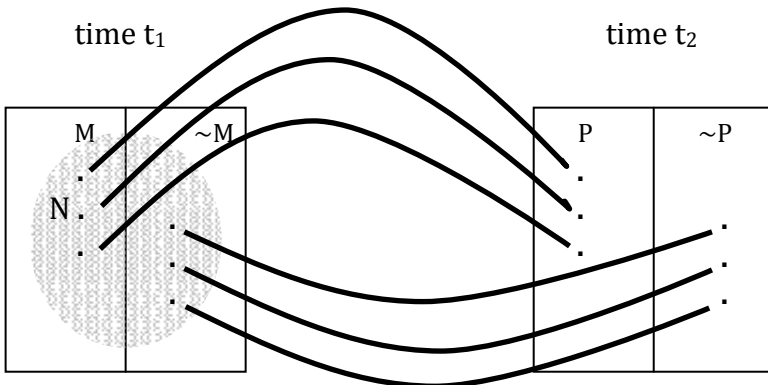


Figure 2

<sup>22</sup> It is easy enough to translate between the new representational scheme and the traditional one. For example, possible worlds are represented as trajectories through state space in the new scheme. The possible world semantics for counterfactuals can be applied to possible worlds so conceived in the new scheme.



ing that its distinct subvening microevent, N, is nomologically sufficient for P<sup>23</sup> and that M and N do not causally overdetermine P.<sup>24</sup>

Let me sum up my argument so far. In this section my point has been to show that the causal closure argument is fallacious when its premises are interpreted naturally so that ‘causal sufficiency’ means nomological sufficiency and ‘causation’ means counterfactual difference-making. I have been at pains to argue that there is a consistent situation in which its three premises are true but its conclusion is false. I have represented this situation schematically as one in which a mental event M is a difference-making cause of physical event P, a neural event N, which is the supervenience base for M, is nomologically sufficient for P but not a cause of it, and yet M is numerically distinct from N. Accordingly, a non-reductive physicalist may accept the premises of the argument without being compelled to accept the conclusion that every mental event is identical to a physical event.

### 3. The Reformulated Argument is Unsound

So we have seen that the causal closure argument is fallacious when its premises are interpreted in terms of a mixture of different concepts. Presumably, the argument’s exponents intend to use the same concept of causation throughout the argument, but mistakenly assume that ‘sufficient causation’ means the same thing as ‘causation’. I have been concerned, however, to argue that this is not so and that so-called ‘sufficient causes’ need not be genuine causes at all. This raises the question whether it is possible to reformulate the argument using a uniform conception of difference-making causation. In order for the argument to go through on this uniform conception of causation, the causal closure principle would need to be reformulated as follows:

5. For every physical effect there is a *physical difference-making cause*.  
(Modified causal closure principle).

<sup>23</sup> In conformity with the causal closure principle, every physical microstate at time  $t_2$  has a nomologically sufficient condition in a physical microstate holding at time  $t_1$ . Similarly, every physical macrostate at time  $t_2$  (consisting of a disjunction of microstates at time  $t_2$ ) has a nomologically sufficient condition in a physical macrostate at time  $t_1$  (consisting of a disjunction of the corresponding sufficient microstates at  $t_1$ ).

<sup>24</sup> It should not be surprising that macrolevel causal dependences are compatible with an underlying determinism of microstates. This is exactly the abstract logical structure of some versions of compatibilism about free will and determinism, for example Lewis’s (1981) version. Whatever objections might be advanced against such versions of compatibilism, it is implausible to think that they contain formal contradictions.

With the causal closure premise reformulated in this way, the argument is clearly valid, as the argument no longer suffers the defect of appealing in one premise to ‘sufficient causation’ and in other premises to genuine causation. Thus, in the schematic example from the last section, if the reformulated causal closure principle tells us that the physical event P has a difference-making cause in the neural event N, then we can suppose that the mental event M is also a difference-making cause of P only by thinking this is a case of overdetermination or, if this is ruled out, only by thinking that M is identical to N.

Before I set out my response to this new argument, I wish to set aside two ways of questioning the validity of the argument. One way of casting doubt on the argument is to point out that an application of the reformulated causal closure principle in the argument does not require that the difference-making physical cause of the given physical effect should occur at the same time as some other putative mental cause, or should act as a supervenience base for such a mental cause. Strictly speaking, the physical cause that makes a difference to the physical effect may occur earlier or later than the putative mental cause. Accordingly, positing two causes, occurring at different times, should not tempt us to conclude that these events are identical. How could they be identical if they occur at different times?

The problem with questioning the validity of the causal closure argument in this way is that it merely postpones the conclusion rather than finessing it. Suppose that we have posited for a physical effect P a mental cause M occurring at one time  $t_1$  and a neural cause  $N^*$  occurring at an earlier time  $t_0$ . Even though M and  $N^*$  cannot be identical, they still constitute a case of problematic overdetermination. And we need to ask how the actual neural supervenience base for M – call it N – fits into the causal picture. The only coherent picture of the causal relations in this situation, the physicalist will argue, is one which represents the chain of difference-making causes from  $N^*$  to P as going through N. The causal closure argument now gets its grip at precisely this point. The only way to avoid the conclusion that M and N constitute a pair of problematic overdetermining causes is to hypothesize that they are one and the same event. Indeed, this conclusion seems to be unavoidable for any mental event that precedes or succeeds M in the causal chain that leads to the physical effect P: any such mental event must be identical with its subvening neural event if there is to be no problematic overdetermination.

Another way of casting doubt on the new causal closure argument is to maintain that, in accepting the mental cause as well as the physical cause of

the physical effect, we are not forced to conclude that either they constitute a problematic pair of overdetermining causes or they are identical. For they could, after all, be distinct events each of which is a contributing rather than an overdetermining cause of the effect. The difference between contributing and overdetermining causes is illustrated by familiar examples. A person's careless discarding of a lighted match into dry grass and the presence of light breeze at the time are contributing causes of a bush fire, whereas two members of a firing squad shooting their victim at the very same time are overdetermining causes. Someone might question the validity of the argument by claiming that the occurrence of the mental and physical causes should be assimilated to the paradigm of contributing causes rather than overdetermining causes.<sup>25</sup>

In response to this line of thought, however, it has to be noted that the mental cause *M* and physical cause *N* of the common physical effect *P* do not have the typical counterfactual profile of contributing causes. In a typical case of contributing causes, it is true that without each cause's contribution, the effect would not have happened. For example, if either the careless discarding of the lighted match or the light breeze had not happened while the other had, the bush fire would not have happened. But the counterfactual profile of mental and physical causes does not fit this pattern. For it is false that if the physical event *N* had not occurred but the mental *M* had, the physical effect would not have occurred.<sup>26</sup> For the mental event *M*, differently realized, would have produced the same effect. So in the light of this, it is difficult to maintain that *M* and *N* should be assimilated to the paradigm of contributing causes.

So in view of the failure of these two skeptical objections, let us accept the validity of the causal closure argument in its new formulation. But this does not mean that we must accept its soundness. For, as I shall now argue, we have very good reason to think that the reformulated causal closure principle is false. The key to seeing this is to recognize that in typical circumstances in which a men-

<sup>25</sup> Loewer (2007) argues for a solution to Kim's exclusion problem that asserts that a mental event and a neural event can be both be causes (though not overdetermining causes) of a bodily movement since the bodily movement is counterfactually dependent on each. Loewer's solution might be viewed as expressing the idea that the mental and neural causes are simply contributing causes.

<sup>26</sup> As Bennett (2003) has pointed out, the situation of the mental cause *M* and the neural cause *N* does not exactly fit the counterfactual pattern of standard cases of overdetermination like the firing squad example, in which it is non-vacuously true that if each cause had not occurred while the other had, the effect would still have occurred. For in the case of the mental and physical causes, it is only vacuously true, if true at all, that if the mental event *M* had not occurred while the neural event *N* had occurred, then the effect would have occurred. But this point does not support the objection in hand, which is that the case of the mental and physical causes should be assimilated to the paradigm of contributing causes.

tal event is a difference-making cause of some physical effect, there is good reason to think that a physical event cannot be a difference-making cause of the same effect.

In order to establish this point I shall rely on a logical result proved by List & Menzies (2009) that bears directly on this issue of whether a mental event and the physical event that realizes it can both be difference-making causes of the same effect. They introduce the useful notion of a realization-insensitive cause, which is crucial to the statement of their result. They say that an event *M* that is actually realized by a physical event *N* is a *realization-insensitive cause* of another event *P* (with respect to *N*) if and only if (a) *M* is a difference-making cause of *P*; and (b) some of the closest  $\sim N$ -worlds in which *M* holds are worlds in which *B* holds. The intuitive idea behind this concept is that an event is a realization-insensitive cause when it makes a difference to its effect and it produces the effect in a manner that does not depend on the actual realizer of the cause – i.e. small perturbations in the way the cause is realized will still result in the effect. To take an example given by List and Menzies, consider the causation of a bodily movement *P* by a mental event *M*, where *M* is actually realized by a highly specific neural event *N* (say, a very specific temporally-ordered sequence of firings of particular neurons; 2009: pp. 478–479). Supposing that what typically matters for the initiation of a bodily movement *P* is an aggregate of firing rates (spikes/second) over a temporal period from a number of neurons,<sup>27</sup> then *M* is likely to be a realization-insensitive cause of *P* in virtue of the fact that small perturbations in the way *M* is realized would have generated the same aggregate firing rate, which would then have produced the movement *P*.

The concept of a realization-insensitive cause plays a crucial role in an important result List and Menzies (2009, pp. 495–499) call the Downwards Exclusion result: if *M* is a realization-insensitive difference-making cause of *P*, the neural state *N* that actually realizes *M* cannot also be a difference-making cause of *P*. Rather than describing their proof, it is easier to explain the result informally. Suppose that *M* is a realization-insensitive cause of *P*, which implies that *M* is a difference-making cause of *P* and that *P* occurs in some of the closest  $\sim N$ -worlds in which *M* occurs. It follows from this straightforwardly that some of the closest  $\sim N$ -worlds are worlds in which  $\sim P$  fails to occur. But this means that the counterfactual  $\sim N \square \rightarrow \sim P$  is false so that *N* cannot be a difference-

<sup>27</sup> The example originally is given in Woodward (2008), which is an elaboration of a study by Mussallam et al. (2004) of the way intentions are encoded as aggregates of firing rates in the brains of macaque monkeys.

making cause of P. Figure 1 above represents this situation exactly. The figure represents the fact that M is a difference-making cause of P, as we saw before. But it also represents that M is realization-insensitive cause of P because P occurs in some the closest  $\sim N$ -worlds in which M occurs. The figure shows that the second of the pair of counterfactuals required to make N a difference-making cause of P – the counterfactual  $\sim N \square \rightarrow \sim P$  – is false. It is easily seen from this figure, then, that when M is a realization-insensitive cause of P, the actual physical event that realizes M cannot also be a difference-making cause of P.<sup>28</sup>

I take this result to provide a basis for saying that the reformulated causal closure principle (5) is not true; or alternatively, that physical effects need not always have physical difference-making causes. We often have good reason, I claim, to believe that the conditions of List & Menzies' Downwards Exclusion result are satisfied. First, we have reason to think that a mental state, say my decision to throw a ball, is a difference-making cause of my bodily movement of throwing the ball. At the crudest level, we have the evidence of a before-and-after observation: before I make the decision, I do not throw the ball and after I make the decision I do. More refined evidence could be gathered through controlled experiments of the kind that social psychology traditionally uses to demonstrate that subjects' intentional states in the form of their beliefs, desires and values affect their behaviour. Secondly, we have reason to think that the precise details of the way in which my decision is realized at the neural level are not material to the generation of the bodily movement. On a plausible rendering of the similarity relation for difference-making counterfactuals, the

<sup>28</sup> This result is stated and proved more formally in the appendix to List & Menzies (2009). See also Menzies and List (2010); Butterfield (2012) for discussion of the implications of the result. List and Menzies say that whether an event is a realization-insensitive cause of another event is an empirical matter that depends on how the similarity relation for the relevant counterfactuals is determined for the system under consideration. In this case, it is an empirical matter whether a mental event excludes its actual neural realizer from being a difference-making cause of a bodily movement. Indeed, they state necessary and sufficient conditions for when a mental event and its neural realizer can both be difference-making causes of the same effect (2009, pp. 490–493). This is an appropriately cautious approach given the weak conception of a realization that they are working with: they say that a mental event is realized by a physical event just when the mental property of the former supervenes on the physical property of the latter. But a stronger conception of realization has it that a mental event is realized by a physical event just when the latter occupies the causal role of the former. On this stronger conception of realization, the Downwards Exclusion result can be strengthened so that it follows *a priori* that if a mental is a difference-making cause of a bodily movement, it is realization-insensitive cause with respect to its actual realizer. For, it follows from the causal role definition of realization that any perturbation in the manner of the mental event's realization will result in a realizer that produces the same bodily movement. Accordingly, the Downwards Exclusion result can be strengthened to show that a mental event of some effect *will automatically without exception* exclude its actual physical realizer from being a difference-making cause of the effect.

closest worlds in which my decision to throw the ball is realized differently are ones where a few more or a few less firings of neurons in the relevant cluster in the supplementary motor area of my brain, or the same firings in a different temporal sequence, would have generated the same aggregate firing rate, so producing the same bodily movement.<sup>29</sup> Accordingly, given the satisfaction of these two conditions, we can infer that the neural state that actually realizes my decision is not a difference-making cause of P. Moreover, since this neural state is the most plausible candidate for the role of the physical difference-making cause of my bodily movement, we can conclude that there is no such physical cause of my bodily movement.<sup>30</sup>

If this reasoning is correct, then it follows that wherever a mental event is a realization-insensitive cause of a physical effect, we will have a violation of the modified causal closure principle. The question of the truth or falsity of the modified causal closure principle has not been discussed in the literature, which, to the extent that it has been clear on the issue, has focused on the formulation of the principle in terms of ‘sufficient causes’.<sup>31</sup> But once we move from the mistaken conception of causes as ‘sufficient causes’ to a more satisfactory conception of them as difference-making causes, we can recognize that the modified formulation is the appropriate one.<sup>32</sup> With this formulation clearly in view, it is easier to see its falsity once we recognize that the existence of realization-insensitive difference-making mental causes excludes the existence of physical difference-making causes of the same effect. Thus, even if we must accept that the new causal closure argument is valid, the falsity of its key premise means that we are not forced to accept its conclusion.

<sup>29</sup> On the stronger notion of realization discussed in the previous footnote, there is no need to justify the claim that a difference-making mental event is realization-insensitive with respect to its actual realizer. For it follows automatically from the causal role definition of realization that a mental cause will be automatically realization-insensitive in this way.

<sup>30</sup> It might be objected that the Downwards Exclusion result does not show that mental difference-making cause excludes every subvening physical event as a difference-making cause. For the disjunction of all the possible physical realizers of the mental event will count as a difference-making cause of the effect. Space limitations prevent me from discussing at proper length the important issues raised by the fact that mental properties are necessarily coextensive with disjunctions of physical properties. However, it seems to me that a non-reductive physicalist should respond to this objection by saying that a disjunction of physical properties need not itself be a physical property, especially in cases in which the disjunct properties are extremely varied and heterogeneous in character.

<sup>31</sup> Some philosophers such as Baker (1993), Sturgeon (2000) and Gillett (2003) have questioned the causal closure principle, but they have not been careful to distinguish the different formulations of the principle.

<sup>32</sup> The falsity of the modified causal closure principle does not imply the falsity of the original principle. One can consistently maintain that the original principle is true while the modified one is false.

#### 4. Conclusion

The discussion of this paper has focused on the role of the causal closure principle in the causal closure argument for reductive physicalism. We have seen that when this principle is formulated as saying that every physical event has *physical sufficient cause*, there is good reason to think the argument is invalid. On the other hand, when the principle is formulated as saying that every physical event has a *physical difference-making cause*, we have seen reason to doubt the truth of the principle and, consequently, the soundness of the argument. Either way the argument does not possess the apodeictic force to compel us to accept its conclusion that every conscious mental event is identical with a physical event.

Apart from this overall conclusion, the discussion of this paper highlights, I believe, the importance of framing arguments about causation in the metaphysics of mind in terms of substantive conceptions of causation. It is surprising how seldom arguments for or against various positions that touch on mental causation draw on substantive theories of causation. One problematic consequence is that many philosophical discussions appeal to the concept of ‘sufficient causation’ as though this captures the significant part of the concept of causation, whereas in fact it is questionable whether it is a genuine causal concept at all. Another problematic tendency is that many discussions neglect to draw on the technical resources of substantive theories of causation such as difference-making theories to clarify and argue for various theses about mental causation. I hope the discussion in this paper can serve as a useful model of how the correction of these two tendencies can shed new light on some familiar arguments in the metaphysics of mind.

#### REFERENCES

- Aizawa, K. (2009). Neuroscience and Multiple Realization: A Reply to Bechtel and Mundale. *Synthese*, 167, 493–510.
- Aizawa, K., & Gillett, G. (2009). The (Multiple) Realization of Psychological and Other Properties in the Sciences. *Mind and Language*, 24, 181–208.
- Anscombe, E. (1971). *Causality and Determination*. Cambridge: Cambridge University Press.

- Baker, L.R. (1993). Metaphysics and Mental Causation. In J. Heil, & A. Mele (Eds.), *Mental Causation*. Oxford: Clarendon Press, 75–95.
- Baker, L. R. (2009). Non-reductive Materialism. In B. McLaughlin, A. Beckermann, & S. Walter (Eds.), *The Oxford Handbook of Philosophy of Mind*. Oxford: Clarendon Press, 109–127.
- Bechtel, W., & Mundale, J. (1999). Multiple Realizability Revisited: Linking Cognitive and Neural States. *Philosophy of Science*, 66, 175–207.
- Bennett, K. (2003). Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract It. *Noûs*, 37, 471–497.
- Bickle, J. (2003). *Philosophy and Neuroscience: A Ruthlessly Reductive Account*. Dordrecht: Kluwer.
- Bishop, R.C. (2006). The Hidden Premiss in the Causal Argument for Physicalism. *Analysis*, 66, 44–52.
- Butterfield, J. (2012). Laws, Causation and Dynamics at Different Levels. *Interface*, 2(1), 101–114.
- Chalmers, D. (1996). *The Conscious Mind*. Oxford: Oxford University Press.
- Crisp, T., & Warfield, T. (2001). Kim's Master Argument. *Noûs*, 35, 304–316.
- Davidson, D. (1970). Mental Events. In L. Foster, & J. Swanson (Eds.), *Experience and Theory*. Amherst, MA: University of Massachusetts Press, 79–101.
- Davidson, D. (1993). Rethinking Causes. In J. Heil, & A. Mele (Eds.), *Mental Causation*. Oxford: Oxford University Press, 3–17.
- Dowe, P. (2000). *Physical Causation*. Cambridge: Cambridge University Press.
- Fair, D. (1979). Causation and the Flow of Energy. *Erkenntnis*, 14, 219–250.
- Gillett, G. (2003). Non-Reductive Realization and Non-Reductive Identity: What Physicalism Does Not Entail. In S. Walter, & H. Heckman (Eds.) *Physicalism and Mental Causation*. Exeter: Imprint Academic, 31–58.
- Gillett, G. (2003). The Metaphysics of Realization, Multiple Realizability, and the Special Sciences. *Journal of Philosophy*, 100, 591–603.



- Hitchcock, C. (1995). Discussion: Salmon on Explanatory Relevance. *Philosophy of Science*, 97, 335–364.
- Kim, J. (1973). Causation, Nomic Subsumption, and the Concept of Event. *Journal of Philosophy*, 70, 217–236.
- Kim, J. (1992). Multiple Realization and the Metaphysics of Reduction. *Philosophy and Phenomenological Research*, 52, 1–26.
- Kim, J. (1998). *Mind in a Physical World*. Cambridge, MA: MIT Press.
- Kim, J. (2005). *Physicalism, or Something Near Enough*. Princeton: Princeton University Press.
- Kim, J. (2009). Mental Causation. In B. McLaughlin, A. Beckermann, & S. Walter (Eds.), *The Oxford Handbook of Philosophy of Mind*. Oxford: Clarendon Press, 29–52.
- Kim, J. (2010). Causation and Mental Causation. In his *Essays in the Metaphysics of Mind*. Oxford: Oxford University Press, 243–262.
- Lewis, D. (1966). An Argument for the Identity Theory. *The Journal of Philosophy*, 63, 17–25.
- Lewis, D. (1973a). Causation. Reprinted with Postscripts. In *Philosophical Papers*, Vol. II. New York: Oxford University Press, 159–172.
- Lewis, D. (1973b). *Counterfactuals*. Oxford: Blackwell.
- Lewis, D. (1980). Mad Pain and Martian Pain. In N. Block (Ed.), *Readings in the Philosophy of Psychology*, 2 Vols. Cambridge, MA.: Harvard University Press, 216–222.
- Lewis, D. (1981). Are we Free to Break the Laws? Reprinted with Postscripts. In *Philosophical Papers*, Vol. II. New York: Oxford University Press, 291–298.
- Lewis, D. (1986). *Philosophical Papers*, Vol. II. New York: Oxford University Press.
- Lewis, D. (2000). Causation as Influence. *Journal of Philosophy*, 97, 182–197.
- List, C., & Menzies, P. (2009). Non-reductive Physicalism and the Limits of the Exclusion Principle. *Journal of Philosophy*, 106, 475–502.

- Loewer, B. (2007). Mental Causation, or Something Near Enough. In B. McLaughlin, & J. Cohen (Eds.), *Contemporary Debates in Philosophy of Mind*. Oxford: Blackwell Publishing, 243–264.
- MacLaughlin, B. (1989). Type Epiphenomenalism, Type Dualism, and the Causal Priority of the Physical. *Philosophical Perspectives*, 3, 109–235.
- Menzies, P., & List, C. (2010). The Causal Autonomy of the Special Sciences. In G. Macdonald, & C. Macdonald (Eds.), *Emergence in Mind*. Oxford: Oxford University Press, 108–128.
- Montero, B. (2003). Varieties of Causal Closure. In S. Walter, & H. Heckman (Eds.), *Physicalism and Mental Causation*. Exeter: Imprint Academic, 173–187.
- Mussallam, S., Corneil, B., Greger, B., & Andersen, R. (2004). Cognitive Control Signals for Neural Prosthetics. *Science*, 305, 258–262.
- Papineau, D. (2002). *Thinking about Consciousness*. Oxford: Oxford University Press.
- Pereboom, D. (2002). Robust Nonreductive Materialism. *Journal of Philosophy*, 99, 499–531.
- Salmon, W. (1971). Statistical Explanation. In W. Salmon (Ed.), *Statistical Explanation and Statistical Relevance*. Pittsburgh: University of Pittsburgh Press, 29–87.
- Salmon, W. (1994). Causality without Counterfactuals. *Philosophy of Science*, 61, 297–312.
- Shapiro, L. (2004). *The Mind Incarnate*. Cambridge, MA: MIT Press.
- Shapiro, L., & Sober, E. (2007). Epiphenomenalism: The Do's and the Don'ts. In P. Machamer, & G. Walters (Eds.), *Thinking about Causes*. Pittsburgh: University of Pittsburgh Press, 235–264.
- Shoemaker, S. (2007). *Physical Realization*. New York: Oxford University Press.
- Sturgeon, S. (2000). *Matters of Mind*. London: Routledge.
- Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford and New York: Oxford University Press.
- Woodward, J. (2008). Mental Causation and Neural Mechanisms. In J. Hohwy, & J. Kallestrup (Eds.), *Being Reduced: New Essays on Reduction, Explanation, and Causation*. New York: Oxford University Press, 218–262.