

## Kim on Causation and Mental Causation

---

Panu Raatikainen<sup>1</sup>

---

**Abstract:** Jaegwon Kim's views on mental causation and the exclusion argument are evaluated systematically. Particular attention is paid to different theories of causation. It is argued that the exclusion argument and its premises do not cohere well with any systematic view of causation.

**Keywords:** Mental causation, exclusion argument, non-reductive physicalism, causation, Jaegwon Kim

---

<sup>1</sup> Department Philosophy, Pinni B4147, FIN-33014 University of Tampere, Finland, e-mail: [panu.raatikainen@staff.uta.fi](mailto:panu.raatikainen@staff.uta.fi)

# 1 Introduction

Jaegwon Kim is a towering figure in the contemporary philosophy of mind. A considerable part of the literature in this area nowadays revolves around Kim's views on mental causation. The problem of mental causation has been a central problem in the philosophy of mind at least since Princess Elizabeth's devastating critique on Descartes' interactive dualism in the 1640s. The topic has again become a subject of intense debate over the last half-century. In recent decades, the dispute over mental causation has often focused on the famous causal exclusion argument (for helpful overviews, see Bennett 2007; Yoo 2007; Robb & Heil 2014; Gibb 2014). Kim has been a leading advocate of this argument.<sup>2</sup>

This argument is based on the apparently plausible assumption of "the causal closure of the physical realm" (see below) and the worry that the mental, if it is not physical, would end up being epiphenomenal, that is, causally idle. The closure principle amounts to the assumption that all physical effects are due to physical causes. The argument is then, roughly, that anything that has a physical effect must itself also be physical. Thus, if the mental is capable of causing physical effects, it must itself be physical. Today this argument is standardly called "the causal exclusion argument", or simply "the exclusion argument" (though sometimes also "the overdetermination argument" or "the causal argument").<sup>3</sup>

The aim of this paper is to evaluate systematically Kim's views on mental causation and exclusion by paying particular attention to different theories of causation. My aim in doing this is to clarify the debate by being more explicit about the different views on causation and to determine how the key premises of the exclusion argument fare in relation to them.

The literature on Kim and the exclusion argument is vast, so it is impossible to discuss it comprehensively; it is necessary to be very selective.<sup>4</sup> In particular, the focus in this paper is more on the philosophy of science than on metaphysics (though I don't believe the difference between these two is clear-cut; accordingly, it is not really possible to completely avoid metaphysical issues). In particular, as interesting as they are, I prefer to avoid metaphysical questions concerning constitution, grounding, universals and properties, the details of the Lewisian metaphysics, etc. Although these are worthy considerations, I feel that if a response to the exclusion argument depends essentially on a specific view on such issues, it is much more controversial than the widely accepted conclusions about causation discussed below. I shall argue that the exclusion argument can be shaken without going into such difficult topics. The focus is on Kim's views in particular—on their coherence both internally and with respect to the best available understanding of causation in the philosophy of science.<sup>5</sup>

---

<sup>2</sup> David Papineau has been another important figure (see e.g. Papineau 1993, 2001, 2009). However, Papineau's views differ from Kim's in certain delicate ways, and deserve analysis of their own. In this paper, the focus is on Kim's views.

<sup>3</sup> Kim calls a certain extended variant (see below) of it "the supervenience argument"—in its common form, though, the exclusion argument does not lean on the assumption of supervenience.

<sup>4</sup> Bennett (2007), Yoo (2007), Robb & Heil (2014), and Gibb (2014) include useful surveys of various attempts to resist the exclusion argument.

<sup>5</sup> Certain recent non-standard variants of the exclusion argument philosophers (e.g. Baumgartner 2009, 2010, 2013; Zhong 2014) must also be omitted and left for another occasion.

## 2 The Causal Exclusion Argument

The exclusion argument of course did not originate from Kim. In contemporary philosophy, the argument seems to have several somewhat different roots. Already the early advocates of the mind-body identity theory such as Feigl (1958) and Smart (1959) proposed that we should identify mental states with brain states, for otherwise those mental states would be “nomological danglers,” lying outside causal laws and the causal structure of reality, and in particular playing no role in the causal explanation of behavior.<sup>6</sup> As Smart memorably concluded, mental states are therefore “nothing over and above brain processes”.

In the same tradition, Armstrong (1968) and Lewis (1966, 1972) argued, first, that mental states are picked out by their causal roles, and, further, that we know that physical states play these roles: they concluded that mental states must be identical with those physical states. Lewis (1966) made the assumption of *causal closure* explicit (under the label “the explanatory adequacy of physics”). It is not implausible to see a rough version of the exclusion argument already here. At least the *closure* principle also played a role in Davidson’s once influential Anomalous Monism (Davidson 1970, 1974; cf. Yalowitz 2014).

Usually though, the exclusion argument proper is credited to Malcolm (1968), who himself used it to argue that reasons could *not* possibly be causes of action (it is the question of which premise to give up). Later, Peacocke (1979) and Schiffer (1987), for example, have used the argument, but more recently, it has been undeniably Kim above all who has clarified and pressed the exclusion argument in defense of strong reductionist form of physicalism, or the identity theory (see Kim 1989a, b, 1992, 1993a, b, 1998, 2005, 2006a, b, 2009).

Since the 1960s, the identity theory has become much less popular. The most important reason for this is Putnam’s (1967) argument from multiple realizability. This argument has convinced a number of philosophers to reject the identity theory and reductionism. Non-reductive views have thus been prevailing. This is the context in which Kim and others have now presented the exclusion argument.<sup>7</sup> Kim in particular argues that non-reductive physicalism is an unstable and a much more problematic position than it first appears: because of the exclusion argument, it threatens to make mental properties entirely epiphenomenal; therefore, we should after all favor reductive physicalism—that is, the identity theory in some form.

Like everything in philosophy, the argument from multiple realizability is, of course, not completely uncontroversial.<sup>8</sup> Nevertheless, this is not the place to be absorbed in that issue. Our focus here is on the exclusion argument and specifically on whether it forces us to give up non-reductive physicalism and accept a more austere physicalism. For the sake of argument, we simply assume that mental properties are multiply realizable and that non-reductive physicalism is true. As a part of it, we shall also presuppose that everything, whether reducible or not,

---

<sup>6</sup> The talk of “nomological” (“relating to laws”) suggests that they presupposed the regularity view of causation.

<sup>7</sup> At some points, Kim has suggested that the exclusion argument also rebuts dualism; Papineau (1993) and Bennett (2008), for example, seem to agree. Kroedel (2015) argues that even this is not true. I shall not consider this question explicitly, but obviously some of the arguments discussed here have relevance to it as well.

<sup>8</sup> For different critical responses, see, e.g., Funkhouser (2007), Bickle (2008), and the references given therein. Bickle (1998), Bechtel & Mundale (1999), Shapiro (2000), and Polger (2004) in particular deserve to be mentioned.

supervenes on the physical with metaphysical necessity. We will then reflect on whether this view necessarily runs into the kind of troubles Kim argues it does.

Let us run over the exclusion argument. Assume first that strong physicalism, that is, reductive materialism or the mind-body identity theory, is false:

*Assumption (distinctness)*: Mental properties are distinct from (not type-identical with) physical properties.

The following four premises are—so the argument goes—apparently unexceptionable. To begin with, it is maintained that unless something like the following closure principle holds, physical reality would be mysteriously full of gaps:

*Premise 1 (the causal closure of the physical)*: Every physical occurrence has a physical cause.

The general idea behind this principle is that as we trace back the causal history of any physical effect, there will never be a need to appeal to anything essentially non-physical (vital spirits, souls, gods, demons, or whatever). Note that this principle does not as such amount to strong physicalism, or the identity theory; it is even compatible with some versions of dualism. It only states that, whatever else, non-physical or “spiritual” entities or properties there may exist, the physical realm is “causally closed”. Even an opponent of physicalism may find this principle plausible and even incontestable. It is common to think that a non-reductive physicalist must necessarily commit herself to it.<sup>9</sup>

Then again, it seems to be both a common-sense truism and something that much of psychology and the social sciences presuppose: mental states such as beliefs and desires bring about bodily behavior. That is:

*Premise 2 (causal efficacy)*: Mental events sometimes cause physical events, and sometimes do so by virtue of their mental properties.

But could not both a mental state and the underlying physical state (e.g., the brain state) be the cause of certain behavior? Philosophers have indeed reflected on peculiar cases in which an event has more than one cause. In such a case, the event is said to be “overdetermined” by its causes. A standard example is a death caused by several members of a firing squad shooting simultaneously. However, there is wide agreement that such cases of overdetermination are relatively rare coincidences, or intentionally orchestrated, and that behavioral events cannot frequently be overdetermined in this way. This leads to the next premise:

*Premise 3 (no universal overdetermination)*: The physical effects of mental causes are not all overdetermined.

Finally, the obvious-looking principle of causal exclusion (not to be conflated with the exclusion *argument*) is added:

*Premise 4 (exclusion)*: No effect has more than one cause unless it is overdetermined.

Kim, for example, writes: “the causal exclusion principle is widely accepted and I believe it is virtually an analytic truth” (Kim 2005, p. 51). In another context (Kim 2014), he calls it “virtually analytic and essentially vacuous”.

The causal exclusion argument now states that the assumption and the premises are inconsistent: premise 3 (*no overdetermination*) denies what the rest apparently entail: the

---

<sup>9</sup> For science-based arguments in favor of the principle, see Papineau (2001, 2009); for some critical reflections, see Montero (2003) and Ney (2016).

mental property and the underlying physical property do overdetermine the behavioral effect at stake. Kim (2005, p. 15) takes the premises to be “unexceptionable” and concludes that the assumption must be rejected.<sup>10</sup>

*Conclusion:* Mental properties must be identical to physical properties.

*Extended Exclusion Argument.* Sometimes Kim and others also want to extend the argument from mental-to-physical (“downward”) causation to the case of (inter-level) mental-to-mental causation, and argue that the non-reductive view would make such inter-level causation impossible as well.

Roughly, the idea of this extended argument is the following: Let  $M$  be a mental property at time  $t$ , and  $M^*$  another mental property at later time  $t'$ . Could  $M$  be the cause of  $M^*$ ? Assume that they are realized by, and supervene on, two physical properties  $P$  and  $P^*$ , at time  $t$  and  $t'$ , respectively. Because of the supervenience, if  $P^*$  occurs,  $M^*$  must occur too. It seems that only way that  $M$  can cause  $M^*$  is by causing  $P^*$ . But then, the standard exclusion argument shows—so the argument continues—that the latter is impossible. Therefore,  $M$  cannot be the cause of  $M^*$ .

But it is important to recognize that this really is an extension (even if perhaps an easy corollary) of the standard exclusion argument—or at least to be clear which one is under discussion. Indeed, Kim often calls such an extended argument, with the supervenience assumption included in its premises, “the supervenience argument” in order to distinguish it from the basic exclusion argument, which essentially concerns physical effects and their alleged mental causes, that is, downward causation.<sup>11</sup>

### 3 The Relevance of Theories of Causation

Does it matter, for the issue of mental causation, and for the exclusion problem in particular, what theory of causation is presupposed? On the one hand, both the exclusion argument and responses to it have been frequently presented without reference to any specific views about causation, but they have been put forward as if they held quite independently of any particular theory of causation. One has often leaned only on some general intuitions about causation.

Sometimes, on the other hand, it is claimed that all that matters is the choice of the theory of causation: for example, it has been suggested now and then that given the dependence view of causation, mental causation is no problem at all, whereas if the production view of causation is assumed, mental and other higher-level causation is immediately impossible even without any

---

<sup>10</sup> One could alternatively refute, for example, the efficacy premise—that is, conclude that mental properties do not cause any physical effects (as, e.g., Malcolm did)—but most contemporary philosophers find this epiphenomenalist option unattractive.

<sup>11</sup> Thus Kim writes: “The ... argument is the so-called ‘exclusion argument,’ since it aims to show that a mental cause of a physical event is always excluded by a distinct physical cause” (Kim 2006b, p. 197; my emphasis). And again: “Thus the problem of causal exclusion is to answer this question: Given that every physical event that has a cause has a physical cause, how is a mental cause also possible?” (Kim 1998, p. 38; my emphasis). “In essence” the supervenience argument is, in Kim’s own words, “the result of superimposing mind-body supervenience on the causal exclusion problem” (Kim 1998, p. 39). Bennett, in turn, writes: “The exclusion argument is also known as the overdetermination argument. (Kim’s recent ‘supervenience argument’ is an extension)” (Bennett 2007, p. 324; my emphasis).

exclusion argument.<sup>12</sup> Bennett (2008), for one, disagrees: “But while I certainly agree that the production view [of causation] is often in the background of discussions of the problem ... I do not agree that the problem itself actually requires it. I do not agree that rejecting it makes the issue go away.” Kim (2002, p. 675) himself at least admits that “Loewer is right... in saying that my thinking about causation and mental causation involves a conception of causation as ‘production’ or ‘generation’”. Bennett, however, thinks it is “wrong to assume that the pure dependence notion alone would dissolve the [exclusion] problem completely” (*ibid.*).

It is indeed old news (see, e.g., LePore & Loewer 1987; Baker 1993), that from the perspective of the dependence notion of causation (e.g., the counterfactual approach), mental states or properties seem perfectly suitable for causing, for example, bodily behavior, and from this perspective, indeed often seem to do so. I agree with Bennett, though, that this observation does not yet make the exclusion problem go away, and it does not really tell us where exactly the exclusion argument then goes wrong—that is, it does not provide us with any analysis or diagnosis of the problem. Yet, I submit that a careful analysis based on an up-to-date picture of causation *can* illuminate the problem, and perhaps even dissolve the problem. The resolution does not, though, fall out trivially from the most plausible theories; it requires a little elaboration.

## 4 Kim’s Progress

Kim is indeed showing increasing sympathy towards the “production” or “generation” view of causation. But it would be a mistake to conclude that this has been his view of causation all along and he has always understood the exclusion argument in terms of such a theory of causation.

To begin with, Kim has often presented the exclusion argument as if it is sound quite independently of any specific theory of causation—and so have many others. In his 1998 book, *Mind in a Physical World*, for example, Kim reflects on alternative views of causation, namely, nomological sufficiency and counterfactual dependence, and states that from either perspective, the physical state *P* underlying the relevant mental state *M* qualifies as a cause of the subsequent physical occurrence *P\** (p. 43), thus implicitly suggesting (more or less) that the exclusion argument does not essentially depend on any particular view of causation. A couple of pages later (p. 45), though, Kim clearly sympathizes with the view of causation as production (and refers to Salmon 1984)) and leans on it while arguing that mental-to-mental regularities may not be genuinely causal. In a later work, Kim (2006a) suggests that the exclusion argument succeeds in establishing the same conclusion, whether causation is understood as nomological sufficiency or in terms of counterfactuals. In an even more recent work, Kim (2014) again argues that the conclusion follows whether one understands causation as nomological sufficiency, as counterfactual dependence, or as production or generation.<sup>13</sup>

More importantly—and this has not apparently been widely recognized—in his early articles on the exclusion problem (Kim 1989a, 1993), Kim is seemingly rather leaning on the traditional regularist-nomological view. Already in these earlier papers, Kim undeniably shows a certain awareness of the limitations of at least the covering-law model of explanation. But it is fair to

---

<sup>12</sup> Loewer (2002), for example, seems to think along these lines.

<sup>13</sup> As will become apparent, I do not agree.

say that by and large, he presupposes the regularist-nomological view of causation and causal explanation in those papers. Thus, in his 1993 article, Kim writes:

We shall assume here a broadly *nomological* conception of causality, roughly in the following sense: an instance of *M* *causes* an instance of *N* just in case there is an appropriate causal law that invokes the instantiation of *M* as a *sufficient condition* for the instantiation of *N*. (Kim 1993, p. 204; my emphasis)

In his more recent writings, Kim instead advocates explicitly the “production” or “generation” view of causation; he now seems to sometimes even think that mental causation, and the exclusion argument in particular, somehow requires such a view of causation. Accordingly, Kim writes in his 2007 article:

... even if we succeed in showing that mental causation, with causation construed as dependence, is real, that would not suffice to vindicate mental causal efficacy and thereby dissipate our epiphenomenalist worries. In my view, what we need to show is that mentality has causal powers to bring about their effects in a continuous process of generation and production. ... We care about mental causation because we care about *human agency*, and agency requires the productive/generative conception of causation. ... These causal processes all involve *real connectedness* between cause and effect, and the connection is constituted by phenomena like energy flow and momentum transfer, an actual movement of some (conserved) physical quantity. (Kim 2007, p. 236)

In an endnote, Kim specifies: “I am of course referring to the so-called conserved quantity approach to causation,” and refers to Dowe (1992, 2000) and Salmon (1994) (Kim 2007, p. 240, en 13).

Furthermore, commenting on Loewer’s claim that the exclusion and supervenience arguments rely on a productive/generative concept of causation, Kim (2009, p. 44) writes: “I believe that Loewer is correct in saying that a robust notion of generation or production is involved in the concept of causation driving the [exclusion and supervenience] arguments”. He further adds: “An approach to causation that is quite congenial to the productive/generative view of causation is the conserved-quantity account” (*ibid.*).

## 5 Developments in the Theories of Causation

In order to properly evaluate Kim’s views, it is necessary to take a closer look at the developments in theories of causation (cf. Psillos 2002, Hitchcock 2008; Woodward 2016). It has become conventional to divide theories of causation into two broad families, first, the *causation-as-dependence* views,<sup>14</sup> and second, *causation-as-production*, or *causation-as-transmission*.<sup>15</sup> The former includes the regularity view and various counterfactual approaches.

---

<sup>14</sup> Some philosophers use “causation as difference-making” more or less interchangeably with “causation-as-dependence”. I am reluctant to use the former label for this wide class, because some theories in it, such as the regularity theory, fail quite badly in capturing the idea that causes are difference-makers. Consequently, I prefer to use only the more general label here. I shall, though, talk about “causes as difference-makers” as referring to the intuitive picture of causation that emerges from the critique of the regularity theory. As I use it here, if a little more specific than causation-as-dependence, it does not refer to any particular exact theory (some philosophers by contrast use it to refer only to the Lewisian counterfactual theory).

<sup>15</sup> (Hall 2004) is a standard reference. Hall, though, suggests that there are two independent and equally plausible concepts of causation. All the same, the distinction is now often used to classify competing theories of causation, without such pluralistic implications; see, e.g., Hitchcock (2008).

More developed examples of the latter are Salmon's mark transmission account (1984) and Dowe's conserved quantity account (1992, 2000).

*The regularity theory.* For a long time, the received view on causation, especially among empiristically minded philosophers, was *the regularity theory of causation*: *c* causes *e* if and only if all events of type *C* (i.e., events that are like *c*) are regularly followed by events of type *E* (i.e., events like *e*). The standard picture of scientific explanation, grounded on this view of causation, was the covering-law model—or the deductive-nomological model—of explanation.

This approach has long been criticized, and problems have accumulated. The fundamental problem is that it cannot distinguish truly causally relevant properties from mere correlations. For example, a sudden drop in the reading of a barometer is regularly succeeded by the occurrence of a storm. However, it does not follow—*pace* the regularity theory—that the barometric reading caused the storm; rather, a drop in atmospheric pressure caused both the barometric reading and the storm.

To give another example, imagine Mrs. Smith aims to kill her husband. She puts some cyanide in Mr. Smith's drink. However, deeply superstitious as she is, and because she wants to guarantee the result, she also casts a spell over the poison. Now digesting a sufficient dose of cursed cyanide is regularly followed by death. Still, it is not causally relevant to the husband's death that the poison was also cursed.<sup>16</sup>

Salmon (1971) gave an amusing and vivid example: Start with the fact that John Jones, a male, fails to get pregnant. In the example, John Jones, for some strange reason, has regularly taken birth control pills for an entire year. Finally, it is a fact that:

All males who take birth control pills regularly fail to get pregnant.

Nevertheless, it would be absurd to consider the use of birth control pills as the cause of John Jones's not getting pregnant.

All these examples show, in different ways, that the regularity theory is unable to distinguish between difference-making causes and contingent, irrelevant factors. At the same time, they show the fact that an event or a property being sufficient for another to occur is not the same thing as an event's or property's being a difference-making cause—the former is not a sufficient condition for being a cause (the regularity theory blurs this distinction).

Scriven (1959) noticed that the occurrence of paresis can, in some cases, be causally explained in terms of the fact that the patient had latent syphilis untreated by penicillin. However, given someone with latent untreated syphilis, the chance that he or she will develop paresis is only about one in four, and there is no known way to separate those who will develop paresis from those who will not. That is, untreated syphilis is, in some cases, the cause of the occurrence of paresis, but the former is not regularly followed by the latter. The former does not even make the latter particularly probable.

I would like to add that an even more dramatic example is provided, for example, by low-level radiation and cancer: One sievert carries with it a 5.5% chance of eventually developing cancer. Consequently, if 10,000 people are irradiated with a dose of 10 millisieverts of ionizing radiation, this probably causes in the long run five extra cancer deaths. Again, the exposure is *not* regularly followed by cancer, and not even at all probable. And still, in the case of the few cancers, it is the cause.

---

<sup>16</sup> This example is my own, but is adapted from the classic example of hexed salt dissolving into water by Henry Kyburg (1965).

Finally, consider Mackie's (1965) classic example: There has been a fire in a certain house, and the investigation concludes that it was caused by an electrical short-circuit. But surely the short-circuit was not alone sufficient for the fire. There are a number of other causally relevant factors: the presence of oxygen, the absence of an efficient automatic sprinkler, the presence of inflammable material nearby, etc. If the case with any of these had been different, the fire would not have occurred.<sup>17</sup>

More generally, hardly any normal cause (in the standard sense of the word "cause")<sup>18</sup> is in itself sufficient for the effect, but almost all events have many simultaneous relevant causes. Still, such everyday cases are not usually considered cases of overdetermination (see also below). In sum, that a property or an event should be sufficient for the effect is not a necessary condition for it to be truly causally relevant either.

*Counterfactual theories* emerged from dissatisfaction with the regularity view of causation and the related covering-law model of explanation. The basic idea of the counterfactual theories of causation is that the meaning of causal claims can be explained in terms of counterfactual conditionals of the form:

If *C* had not occurred, *E* would not have occurred.

Such counterfactual considerations can easily deal with many counter-examples to the regularity theory and distinguish truly causally relevant factors from mere correlations. For example, even if cursed cyanide is sufficient for death, it is not the difference-making cause. The counterfactual conditional

If Mrs. Smith had not given cursed cyanide to Mr. Smith, he would not have died

is false; even if cyanide was not cursed, he would have died.

The best known, more developed counterfactual analysis of causation is David Lewis's (1973, 1986a, 1986b, 2000, 2004) theory. In terms of counterfactuals, Lewis first defines a notion of causal dependence between events, and then causation in terms of chains of such causal dependence. Nevertheless, vivid philosophical debate over four decades has made it doubtful whether any theory along these lines could work. Difficulties with so-called "preemption" and "trumping" have proven to be insurmountable problems for Lewis's theory (see Menzies 2014).

*Context-sensitivity.* Another serious problem with Lewis's theory—one it shares with many traditional theories—is that it views causation as an absolute relation that is unqualified between two events. On closer scrutiny, though, it is more plausible to understand causal judgments as essentially relative to context and contrasts (cf. Menzies 2014, Section 3).

First, consider again Mackie's short-circuit example. Under normal circumstances, it is natural to consider the short-circuit as the cause, and the presence of oxygen, flammable material and

---

<sup>17</sup> Mackie himself went on to offer his own "INUS" analysis of causation, which is not without problems. However, Mackie's example has turned out to be fruitful, and many philosophers have used it independently of Mackie's specific positive theory.

<sup>18</sup> Mill (1843) anticipated Mackie's argument; Mill, however, suggested that that "the real cause" (as he called it) is only the sum of all causally relevant factors, and that it would be illegitimate to call any one of these "the cause". Though it would indeed be somewhat arbitrary to choose one as "the cause" and call all the others mere "background conditions" (such a distinction would depend more on our interests and expectations rather than on objective facts), it is now common to call any such causally relevant factor equally "a cause". In other words, it is granted that an effect typically has a number of causes. Reserving the name "cause" only for the sum of all them (something that may often be opaque for a scientist) would not harmonize well with the practice of causal talk in the various sciences.

such as background conditions. But what is viewed as the cause and what as background conditions depends solely on our contingent beliefs and expectations, or on our interests. In a different context, say, if the fire took place, for example, in a laboratory that was designed to be oxygen-free and in which short-circuits occur all the time, it would be natural to count the unexpected presence of oxygen as the cause (cf. Hart & Honore 1985, p. 35).

Second, it is increasingly common to think that causal judgments in fact involve (even if often implicitly) a class of contrasts for both cause and effect—that they contrast the putative cause and effect with some alternatives. Different choices of contrasts for the same cause and effect result in different causal judgments, some of which may be true, some false. Unless the contrasts are obvious from the context, it is necessary to fix some contrast class or another in order to make a causal judgment unambiguous (see, e.g., Garfinkel 1981; Woodward 1984; Hitchcock 1996a, b; Maslen 2004; Schaffer 2005; Menzies 2007; Northcott 2008).

Inasmuch as Lewis's counterfactual theory is also unable to accommodate such relativity, it is an inadequate theory of causation.

*Causation as Production.* The basic idea of the causation-as-production or causation-as-transmission view, by contrast, is that causation involves objects coming into contact and exchanging or transmitting something. Although the general intuitive idea is also a classic, the first really well-developed theory of causation of this sort was Salmon's 1984 *Mark Transmission* (MT) account. Salmon suggested that we need to change the conceptual apparatus: instead of taking distinct events or facts to be the causal relata, Salmon thinks we should try to characterize directly when a process is causal. Salmon proposed that a process is causal if it is capable of *transmitting a mark*. Otherwise, we have a non-causal process, or, a pseudo-process. However, Kitcher (1989) and others soon argued that Salmon's definition excluded some genuine causal processes, and allowed some clearly non-causal processes. Moreover, it was important for Salmon to treat causation as an empirical phenomenon, and this involved—in his view—avoiding any appeal to counterfactuals. When Salmon realized that his MT theory in fact makes a tacit appeal to counterfactual relations, he abandoned it (Salmon 1994), and adopted Phil Dowe's *Conserved Quantity* approach (Dowe 1992). In this theory, a causal process is defined as a process that transmits a non-zero amount of a conserved quantity (for example mass, energy, momentum, charge) at each moment in its history.

However, Dowe's theory has likewise serious difficulties in identifying the causally relevant quantities. As Hitchcock (1995) points out, often in causal interactions, several conserved quantities are exchanged. An example: A pool cue strikes a cue ball, imparting both momentum and a blue dot of chalk. In the former case, momentum is exchanged. In the latter, matter is exchanged. Yet only the first is relevant to the trajectory of the cue ball. It is unclear how to determine which exchanges are relevant. A natural response would be to rely, again, on counterfactuals. If we had removed the dot, or had changed it from blue to red, for example, the trajectory of the ball would have been the same. But this solution is not open for Salmon and Dowe, who are trying to avoid any use of counterfactuals. Ironically, as Hitchcock notes, the counterexamples formulated earlier by Salmon himself against the regularity theory—e.g., John Jones taking birth control pills—can be turned against the Salmon-Dowe approach. Salmon acknowledged such difficulties and admitted that they are severe problems for his approach. Dowe attempted to address these worries in his 2000 book. Hausman (2002) and Ehring (2003), for example, have in turn presented a severe critique against it, and it seems that the majority of philosophers remain skeptical about the success of his replies. For these and other reasons, the conserved quantity approach is no longer that popular. More generally, there is no well-developed theory of causation available in the causation-as-transmission or causation-as-

production group which would commonly be regarded as plausible. It seems we must look elsewhere.

*The interventionist theory.* Particularly in the philosophy of science, the theory of causation that is now becoming prevalent is the so-called “interventionist” theory. This theory has been developed especially by James Woodward (1997, 2000, 2003).<sup>19</sup> It can be regarded as a new variant of the counterfactual view of causation, and also belongs to the causation-as-dependence family. The theory is quite sophisticated and flexible, and can deal with many well-known problems of the more traditional counterfactual theory.<sup>20</sup> The interventionist framework can also easily handle cases with many simultaneous causes, such as Mackie’s fire example. Even so, the outline of the interventionist theory can be explained quite simply: It connects causal claims with specific counterfactual claims concerning what would happen to an effect under hypothetical interventions on its putative cause. Roughly, *C* causes *E* if and only if an intervention on *C* would bring about a change in *E*.

To be a little more precise, according to the interventionist approach, causal claims relate variables—say *X* and *Y*—that can take at least two values. These may be magnitudes (such as temperature, electric charge, or pressure), but in simple cases, they may also be just discrete alternative events or states of affairs, or the presence and the absence of a property. The basic idea is that were there a possible intervention on the value of *X*, this would also result in a change in the value of *Y*; in that case, and only in that case, *X* is causally relevant for *Y*. Heuristically, one may think of interventions as manipulations that might be carried out by a human agent in an idealized experiment. However, the approach is not really problematically anthropocentric, and intervention can be defined in purely causal terms (the theory does not even aim to give a reductive analysis of causation, so this is not a problem). Though a hypothetical intervention need not be possible in practice, Woodward is perfectly explicit that it should not be metaphysically impossible, in other words, it must be “well-defined”; otherwise, the related causal judgment does not make clear sense (see, e.g., Woodward 2003, pp. 46, 113–, 127–33).

The interventionist approach, which relates variables, also embodies the important idea that causal claims are essentially contrastive: variables can take different values, and different choices of possible alternative values result in different contrast classes. The natural contrast (what Woodward calls “the default contrast”), though, is often—unless there is some specific reason to choose differently—that the presence rather than absence of the property at issue is caused by the presence of another appropriate property rather than absence of it (see Woodward 2003, pp. 67–68, 145–146; cf. Woodward 2008, pp. 235–236; see also Menzies 2008). This is also the setting that best corresponds to philosophers’ customary discussions of causation, events, and properties outside the interventionist tradition (cf. Woodward 2008, p. 222).

*General lesson: Causes are difference-makers.* The interventionist theory is indeed a promising and attractive recent approach, and it is becoming increasingly popular—especially in the philosophy of science. Furthermore, there has been recently a number of attempts to undermine the exclusion argument based on the interventionist theory (see, e.g., Menzies 2008; List & Menzies 2009; Raatikainen 2010; Shapiro & Sober 2007; Shapiro 2010; Woodward 2008,

---

<sup>19</sup> Related ideas have also been put forward by Pearl (2000), and Spirtes, Glymour, and Scheines (2000); see also Hitchcock (2001).

<sup>20</sup> The interventionist theory builds on the framework of causal models and structured equations which has been employed for some time in econometrics and epidemiology, for example. This toolbox enables much more fine-grained analyses than the heavily metaphysical Lewisian framework with its all-inclusive possible worlds.

2015). There have been critical reactions to these attempts by, for example, Baumgartner (2009, 2010, 2013), Hoffmann-Kolss (2014), and Pernu (2013a, 2014a, 2014b). Woodward (2015, 2017) has in turn produced some responses. Unfortunately, these controversies become quickly very technical and focus on different interpretations of the fine details of the interventionist theory (see, e.g., Baumgartner 2013; Woodward 2015). Though I think these are important issues and deserve careful discussion, they are beyond the scope of this paper. Our main focus here is simply on the coherence and plausibility of Kim's views in relation to certain rather general observations about causation.

I do share the enthusiasm concerning the interventionist approach, and I shall also refer to the basic ideas of these intervention arguments against the exclusion argument. Nevertheless, I prefer a slightly different emphasis here: I want to “step back” and lean as much as possible only on the more general lessons from the development in the theories of causation—especially on certain widely accepted critical conclusions concerning the regularity theory. I try as far as possible to avoid depending on details of this or that specific theory.

Whatever the true theory of causation looks like exactly, we have noted in particular two general philosophical points that emerge from the critical discussion concerning the regularity theory and other traditional approaches: First, just like one should clearly distinguish between causation and correlation, it is important not to conflate causes with mere sufficient conditions: genuine causes are difference-makers for their effects. Second, causation is arguably not an absolute relation, but causal judgments depend on contexts and contrasts. “Causes as difference-makers” is understood here as an intuitive philosophical picture that drives our more rigorous theorizing on causation (cf. Menzies 2004, 2007, 2009). The interventionist theory, in its own way, nicely reflects both points, but these quite general and widely accepted lessons do not depend on the minutiae of any developed theory of causation.

## 6 Exclusion, Causation, and Sufficient Conditions

Let us now revisit the exclusion argument and its premises from the perspective of different theories of causation, and look more closely at the details.<sup>21</sup> The *distinctness* (or non-identity) assumption does not involve causes at all, so it can immediately be put aside. How about the *efficacy* premise? Though Kim has sometimes suggested in his more recent writings that it in fact requires a causation-as-production view (I think it is fair to say that his reasons are not particularly conclusive or clear),<sup>22</sup> the *efficacy* premise does not at any rate explicitly involve any specific view of causation. Much the same can be said about the *no overdetermination* premise (but see below).

The situation with the two remaining key premises, the *closure* principle and the *exclusion* principle, is very different: their more precise formulations often explicitly involve something called “sufficient cause” (see, e.g., Kim 2006b, 2009; Bennett 2007; Yoo 2007). That is, they regularly read:

*Closure*: Every physical occurrence has a *sufficient* physical *cause*.

*Exclusion*: No effect has more than one *sufficient cause* unless it is overdetermined.

In other words, whereas the *efficacy* and the *no overdetermination* premises talk about causes *simpliciter*, the *closure* and *exclusion* principles (so formulated) talk, by contrast, about

<sup>21</sup> In Section 2, the initial formulation of the premises was intentionally left general.

<sup>22</sup> For critical responses, see e.g., Loewer (2007) and Kroedel (2015).

something called “sufficient causes”. Kim and many others seem to think this makes no difference.

However, given the advances in the theorizing about causation (reviewed above)—in particular the moral of the critiques of the regularity view—it begins to look as if both these premises involve conflating causes with sufficient conditions: Genuine causes are relevant difference-makers and are often not as such sufficient for their effects; sufficient conditions are something else, and not necessarily causes—that is, not the things that make a difference. The talk of “sufficient causes” in the exclusion argument mixes up these two fundamentally different things. Therefore, I am inclined to say that those two principles are not so much false (or true), but are mongrels based on a conceptual confusion and they fail to make clear sense.

I contend that even if Kim has more recently distanced himself from the regularist-nomological view of causation, certain remnants of it survive in the exclusion argument: namely, the tendency to equate causes and sufficient conditions (cf. Section 4). In early and later writings, Kim regularly slides from speaking about causes to speaking about “sufficient causes”, and *vice versa*. He often seems to simply assume that being a sufficient condition is either necessary or sufficient for being a cause. And we have seen above that this is quite problematic. This makes the whole exclusion argument quite unstable.<sup>23</sup>

To be sure, the premises in question can be cleared from such confusions and reformulated—either in terms of causes proper, or in terms of sufficient conditions. The former choice is certainly better motivated: after all, the whole point of the exclusion argument and the debate surrounding it is to ask whether a mental property is capable of being a *cause* of something physical. But then, surely the argument and its premises should talk about causes and not be formulated in terms of sufficient conditions. I shall nevertheless also reflect on this latter choice.

Let us now look more closely at the relevant premises. I shall review, if only with broad brush strokes, some plausible ways in which these principles could and have been questioned.<sup>24</sup> All these principles, the critical arguments, and responses to them would certainly merit much more detailed discussion, but here the modest aim is only to give a general idea of the selected arguments (the references given provide further details), and at least show that these premises are more controversial than Kim and some others suggest.

***Overdetermination and Exclusion.*** Many philosophers including Kim have taken the *exclusion* principle as little more than a tautology, and obviously true. Nevertheless, the principle has been challenged. Árnadóttir and Crane (2013), for example, refer to Mackie’s classic fire example (see Section 5 above). From the point of view of the causes as difference-makers picture, the short-circuit, the presence of oxygen, the presence of inflammable material nearby, etc., are all relevant difference-making causes of the fire. Moreover, the fire was *not* overdetermined by them: It is not the case that the fire would still have occurred even if one (or more) of these causes had been absent. Nevertheless, none of these excludes the others as causes; they are all equally causally relevant. The *exclusion* principle, as it is ordinarily

---

<sup>23</sup> Menzies (2008, pp. 208, 211–12), Raatikainen (2010 pp. 358, 360), and Woodward (2008, *passim*) all notice the relevance of this difference for the dispute.

<sup>24</sup> As was noted in the beginning, the critical literature on the exclusion argument is extensive. I shall focus only on responses that are most salient to the general spirit of this paper.

formulated, is thus false: an effect may well have simultaneously several relevant causes without being causally overdetermined.

Furthermore, Pernu (2013b) argues that at least certain formulations of the *exclusion* also involve both causes *simpliciter* and “sufficient causes”<sup>25</sup>—and consequently fail to make clear sense. Moreover, this undermines the overdetermination part: why exactly would having both a “sufficient cause” and a cause *simpliciter* make the effect overdetermined?

As to causal overdetermination, the standard way to explain it is in counterfactual terms: the common idea is that an effect *E* is overdetermined by two causes *A* and *B* only if, roughly, *E* would have occurred even if *A* had not occurred, and *E* would have occurred even if *B* had not occurred (and, finally, *E* would *not* have occurred if neither *A* nor *B* had occurred).<sup>26</sup> For example, in the case of the firing squad shooting simultaneously: the victim would have died even if a particular shooter had not fired his gun. Even Kim—no friend of the counterfactual theory—often explicates overdetermination in terms of counterfactual dependencies.<sup>27</sup>

By contrast, it is not too clear how overdetermination should be even defined in the context of the pure regularity view—that is, in terms of mere sufficient conditions and without any help of counterfactuals. Apparently, the best one can do is simply something like the following: to say that *A* and *B* causally “overdetermine” *E* is to say *A* is sufficient for *E*, and *B* is sufficient for *E*.<sup>28</sup>

At any rate, the exclusion argument proceeds by deriving the opposite of the *no overdetermination* premise—that is, the behavioral effect *is* causally overdetermined from the other premises and the assumption. Many critical analyses focus on this. Funkhouser (2002), Bennett (2003, 2008), and Casey (2012) in particular have developed in different ways the observation (also present in Yablo 1992; Shoemaker 2001; Loewer 2002; Pereboom 2002) that the relation between the (alleged) mental and the physical causes of a behavioral effect is—even if not type-identity—much more intimate than in the paradigmatic cases of overdetermination.

Funkhouser (2002) distinguishes different types of overdetermination: paradigmatic examples of overdetermination, such as the firing squad example, are cases of (what he calls) *independent overdetermination*: in them, overdetermining causes work through separate mechanisms and distinct spatial regions. In so-called *incorporating overdetermination*, by contrast, the alleged overdetermining causes share a mechanism or causal pathway and differ only in their properties within this shared mechanism.<sup>29</sup> The purported causation with supervenient psychological or economical properties, for example, together with their subvening physical properties, is an example of the latter. Now a case of independent overdetermination is either a rare coincidence or premeditated, and it is indeed implausible that *such* overdetermination would occur

---

<sup>25</sup> For example, Kim (2005, p. 17; my emphasis): “If an event *e* has a sufficient cause *c* at *t*, no event at *t* distinct from *c* can be a cause of *e* (unless this is a genuine case of causal overdetermination).”

<sup>26</sup> If the interventionist theory is followed, the relevant counterfactual conditionals must be accordingly formulated in terms of it: “If an intervention had prevented *A* from occurring, *E* would still have occurred,” etc.

<sup>27</sup> There is, of course, nothing wrong in that: one can well analyze overdetermination with the help of counterfactual conditionals, without committing oneself to the counterfactual theory of causation; cf. Bennett (2008, p. 288).

<sup>28</sup> Zhong (2014) apparently interprets overdetermination in this broad way.

<sup>29</sup> Funkhouser further distinguishes what he calls “iterative causal overdetermination” (Loewer 2002 briefly discusses a distinction similar to Funkhouser’s first two types).

regularly. That is, in its case, *no overdetermination* is plausible. Things are, though, quite different with incorporating overdetermination: if non-reductive physicalism is granted, at least for the sake of argument, such overdetermination is destined to be prevalent; if such overdetermination occurs at all, it is necessarily systematic. In other words, in the case of this kind of overdetermination, there is no reason to endorse *no overdetermination*.

Bennett (2003, 2008), on the other hand, focuses on the standard examples of overdetermination,<sup>30</sup> and contends that the holding of the appropriate counterfactuals (see above) is at least a necessary condition for causal overdetermination. She adds an important qualification, though: she argues convincingly that the relevant counterfactual conditionals must be not only true but also non-vacuously true (a counterfactual conditional is vacuously true if the antecedent is metaphysically impossible; clearly that is not the case with standard cases of overdetermination of the firing squad kind: the occurrence of one overdetermining cause without the others is in all cases quite possible.)

Bennett then argues that, given the supervenience assumption, the critical counterfactual

If *P* had occurred without *M*, *B* still would have happened

is true, but vacuously so: the presence of *P* entails the presence of *M* with metaphysical necessity. Consequently, the causal efficacy of mental properties does not in the end lead, in the context of non-reductive physicalism, to overdetermination (in the intended standard sense)—contrary to what the essential step of the exclusion argument claims. Bennett concludes that, assuming non-reductive physicalism (with metaphysical supervenience) and the counterfactual test for overdetermination, the exclusion principle is, after all, false.

Partly building on Bennett's arguments, Carey (2011) considers four possible definitions of overdetermination, and demonstrates that, in all four, either *exclusion* or *no overdetermination* fail.<sup>31</sup> As far as they go, the analyses of Bennett and Carey are quite persuasive, but I think that Funkhouser's analysis in terms of distinct and shared mechanisms (see above) offers a more fundamental explanation of why the relevant counterfactuals behave the way they do (cf. Funkhouser 2014).

Furthermore, in the interventionist framework, the whole issue of overdetermination takes a surprising turn: It has been argued by several philosophers (see Shapiro & Sober 2007; Shapiro 2010; Woodward 2008; Raatikainen 2010) that the interventionist theory of causation and the widely accepted supervenience assumption together make overdetermination problematic: From the interventionist perspective, reflecting on whether a mental state *M* and the physical state *P* realizing it really overdetermine the behavioral effect *B* or not requires that we can construct a causal system that includes a variable for both *M* and *P* (and their alternatives). However, this in turn requires that one can, at least in principle, vary their values independently of each other (like one could, by a hypothetical intervention, prevent one shooter firing his gun without affecting the others in the paradigmatic firing squad example of overdetermination).<sup>32</sup> But if the supervenience thesis is true—that is, if the facts of the physical level determine the mental level (with metaphysical necessity)—constructing the relevant causal system is simply impossible. Consequently, the question of overdetermination does not—so the argument

---

<sup>30</sup> What Funkhouser calls "independent overdetermination".

<sup>31</sup> Carey in fact structures the premises a little differently, but this is, in our terms, the outcome.

<sup>32</sup> Only then it is possible to evaluate the truth or the falsity of the critical counterfactual:

If an intervention were to change *n* from having *M* to not having *M*, while *P* is kept fixed, *n* would have still done *B*.

continues—even make clear sense in this context. According to these philosophers, this gives another reason for doubting the soundness of the whole exclusion argument.<sup>33</sup> Note that the paradigmatic examples of overdetermination (e.g., the firing squad) result in no such problems in the interventionist framework; this suggests again that the inference to the conclusion that *M* and *P* do overdetermine the effect *B* is suspect. Actually, all this harmonizes well with the above-mentioned analyses of Funkhouser, Bennett, and Carey; indeed, these latter analyses may offer a further explanation of why such troubles emerge.

The above arguments, as I have presented them, take seriously the idea that causes are genuine difference-makers rather than mere sufficient conditions.<sup>34</sup> For completeness, let us also contemplate how the *no overdetermination* and *exclusion* premises fare if we pretend for a moment that sufficient conditions can be passed off as causes and consistently follow that line. Philosophers seem to have widely failed to recognize that in this case, it is in fact unexceptional that the supervenient property and the subvenient property both are “sufficient causes” for the effect.<sup>35</sup> Consider, for example, the case in which a victim has ingested 1 g of cyanide and died. In fact, a 200 mg dose of cyanide is sufficient for death, and it would obviously not be possible to digest 1 g but not 200 mg; thus, the former supervenes over the latter. However, either was sufficient for death. To give another example, consider Yablo’s already classic pigeon example (see Yablo 1992): A pigeon has been trained peck at (any shade of) red objects. Now the property of being red supervenes on the property of being scarlet. But clearly both being red and being scarlet are sufficient for pecking. Thus, there are numerous clear cases of such “overdetermination”. In other words, if causes are understood in this way, the *no overdetermination* principle is patently false. If this line is consistently followed, there no longer seems to be any principled reason to insist that the mental and the physical property cannot “overdetermine” the behavioral effect.

Furthermore, if causes are identified with sufficient conditions, the *exclusion* principle is also arguably either problematic or useless—depending on whether cases such as the above examples of cyanide and Yablo’s pigeon are counted as cases of overdetermination or not. If they are, the *exclusion* principle may survive, but it becomes redundant for the purposes of the exclusion argument, because the *no overdetermination* premise fails (as we have just seen). If it is rather assumed that these examples (the cyanide case and the pigeon case) are not real cases of overdetermination, but causes are still identified with sufficient conditions, then these same examples seem to refute the *exclusion* principle.

Of course, all these arguments concerning overdetermination and the *exclusion* principle could be disputed further. But in any case, it is at least fair to conclude that the crucial inference from the assumption and the premises (other than *no overdetermination*) to the conclusion that *B* is

---

<sup>33</sup> It should be mentioned that not everyone agrees with this interpretation: Baumgartner (2009, 2010, 2013), Hoffmann-Kolss (2014), and Pernu (2013a, 2014a, 2014b), for example, do not accept this argument. They interpret the whole thing as rather showing that (given the assumptions) *M* is epiphenomenal. As was noted earlier, dealing with all this properly is beyond the scope of this paper. Suffice it here to say that Woodward (2015, 2016) presents rather detailed responses; and I am inclined to (more or less) agree with them.

<sup>34</sup> Árnadóttir and Crane (2013) suggest that their argument is completely independent of any view of causation. I think this is an exaggeration: While I agree that it does not presuppose any explicitly formulated specific theory of causation, I think that it is based on the broad picture of causation as difference-making and the general counterfactualist picture of causation. It is not clear that the critical argument would work, e.g., in the context of the causation-as-production view.

<sup>35</sup> Hitchcock (2012, p. 50), though, apparently makes a similar observation (under the label “Laplacean Causation”).

causally overdetermined (contradicting *no overdetermination*) is much more problematic and open to controversy than Kim and many others have suggested.

On second thought, it seems that in whatever more exact way causation and causal overdetermination are explicated, either *exclusion* is false, *no overdetermination* is false,<sup>36</sup> or both are false. At the very least, both the intended notion of overdetermination and the relevant *exclusion* principle should be formulated in much more exact and qualified forms than have been yet done in the literature. The burden is obviously on the advocate of the exclusion argument to present a coherent and more rigorous formulation—one that would moreover both be plausible and support the exclusion argument.

**The Closure Principle.** What should we make of the principle of the causal closure of the physical realm? If causation is understood as mere sufficiency, the “closure” hypothesis may perhaps be defensible—but as we have seen, then either the *no overdetermination* principle or the *exclusion* principle, or both, fail; consequently, no exclusion problem threatens.

From the point of view of interventionism, and the more general picture of causes as difference-makers, by contrast, the *closure* principle does not even make clear sense. First, as we have noted above, it<sup>37</sup> conflates causes, which are relevant difference-makers, with sufficient conditions.<sup>38</sup> Second, even if the formulation is revised such that it talks about real, difference-making causes, the principle is not unambiguous unless some contrast class or another is fixed. There are many possible options, but in the case of such a general principle, the default contrasts are a natural choice:

*The causal closure with the default contrast:*

If a physical property  $P_1$  has a cause at time  $t$ ,

then there is a *physical* property  $P_2$  at time  $t$

such that  $P_2$  (rather than not- $P_2$ ) causes  $P_1$  (rather than not- $P_1$ )

Assume then that the focus is on a mental property  $M$  (at time  $t$ ) which is multiply realizable,<sup>39</sup>  $P_1$  is (supposedly) its behavioral effect, and  $P_2$  is the physical property that realizes  $M$  in this particular occasion. Menzies (2008, 2013) and Raatikainen (2010) (see also List & Menzies 2009; Menzies & List 2010; cf. Woodward 2008) have argued that formulated this way, the principle fails. More precisely, they contend that, in the interventionist setting, the following relevant counterfactual conditional:

If an intervention were to change the person from having  $P_2$  to not having it, then  $P_1$  would not occur

---

<sup>36</sup> Or, make no clear sense.

<sup>37</sup> At least many conventional formulations of it.

<sup>38</sup> As I interpret it, the interventionist theory is intended to be a genuine alternative to the regularity-nomological view of causation, one that takes for granted the received wisdom about the shortcomings of the latter (cf. Woodward 2003, *passim*). Though a more specific semi-formal theory, it apparently agrees with the general moral of the philosophical critique, and about the fundamental difference between mere sufficient conditions and genuine causes as difference-makers (discussed above in Section 5), and aims to reflect them. It seems to me that some critics of the interventionist responses to the exclusion argument fail to sufficiently recognize this.

<sup>39</sup> This is, again, a part of our dialectical setting, and something that does not need a defense.

is false. This is because the mere absence of  $P_2$  does not as such rule out the presence of some other physical property  $P_3$  that also realizes  $M$ , and  $P_1$  would still occur. Therefore, the physical property  $P_2$ , though perhaps sufficient for  $P_1$  to occur, is not in such a case its difference-making cause.<sup>40</sup> Against the background of our earlier re-examination of the developments in the theories of causation, this argument simply highlights the general observation that mere sufficient conditions and genuine causes as relevant difference-makers should be clearly distinguished.<sup>41</sup>

One should not misunderstand this challenge to the *causal closure* principle. The idea is emphatically not to suggest that some changes in physical properties are caused by something “from the outside”—that is, the preceding physical state was not itself sufficient for the occurrence of the physical change, but required some additional mental cause.<sup>42</sup> Rather, the point is that the *closure* principle (and the whole exclusion argument) seems to be implicitly based on an overly naïve and simple picture of causation and its relata as forming “the Great Chain of Causes”—a view in which reality, and the micro-physical realm in particular, is naturally carved into events ordered by the causal relation; for every physical event, there exists at any given preceding time *the Cause* of it—an absolute, unambiguous, determinate cause that is sufficient for the event’s occurrence. The point here is simply that real-life causes—causes as difference-makers—are just something more local and partial, and more relative to contexts, interests, and contrasts than this naïve picture suggests.

It is important to understand that all this is consistent with the plausible view that is perhaps often behind the intuitive appeal of *closure*: for every physical occurrence  $B$  which has a cause at  $t$ , there is a physical state of the world  $P$  at  $t$  that is *sufficient* for  $B$ . The argument simply points out that the more specific micro-physical property is not always and automatically—in

---

<sup>40</sup> Woodward (2015, 2017) appears to disagree. However, under closer scrutiny, he only criticizes certain overly simple and strong versions of the argument. First, it is indeed implausible to assume that there is always some sort of back-up mechanism such that whenever  $P_2$  fails to occur, some other realizer of  $M$  occurs instead. (It is not clear to me, though, that, as Woodward suggests, List and Menzies (2009), for example, actually make such an assumption.) As I interpret the argument, by contrast, it only notes that the mere absence of  $P_2$  does not necessarily result in  $\text{not-}P_1$ , as it allows, among other possibilities, cases such as the presence of  $P_3$ , which also realizes  $M$  and therefore rather still brings about  $P_1$ . The latter weaker interpretation is sufficient to make the relevant counterfactual false.

Second, it is certainly unwarranted to conclude that  $P_2$  is, in some more absolute sense, causally irrelevant. Whether it is a cause or not is sensitive to the chosen contrast. The argument, in its defensible form, concerns only “the default contrast”, i.e., the binary case where the variable for the alleged cause can only get two values representing the presence and absence of  $P_2$ . It is consistent with the argument to note, as Woodward does, that if the variable is instead allowed to get a number of values  $P_2, P_3, P_4, \dots$  for alternative physical states, the variable apparently is causally relevant for the behavioral effect. There is thus no assumption that causation in general requires proportionality. All this was perfectly explicit at least in Raatikainen (2010, pp. 356–57); cf. Menzies (2008, pp. 207–10).

Woodward (2008, pp. 232–40) seems to agree about the key points; and also agrees (2017) that, in any event, the relevant counterfactual is false.

<sup>41</sup> Yablo (1992) presents a related argument in terms of more traditional counterfactual theory; cf. also Crane 2001, pp. 64–65.

<sup>42</sup> Accordingly, Raatikainen (2010 p. 358) writes: “Of course, the occurrence of [the subvening physical state] is surely sufficient for the effect, John’s behavior, but that does not make it the cause of the latter. Being sufficient condition for the occurrence of something, and being its difference-making cause, must thus be clearly distinguished.”

every possible way of conceptualizing the situation—a superior candidate for being the difference-maker (as the exclusion argument suggests).

All in all, if causes are consistently taken as difference-makers, as the counterfactual and interventionist approaches suggest, the *exclusion* principle is arguably false and the *no overdetermination* principle becomes problematic, as does the *closure* principle. Then again, if the focus is on causation-as-sufficiency, at least one of *exclusion* and *no overdetermination* is false. (The causation-as-production view is discussed below.) Though it may not be transparent, the choice of theory of causation is relevant for the exclusion argument. There is always room for further debate, but at the moment the prospects of the exclusion argument do not look particularly promising: three key premises falter. At least, Kim and other advocates of the exclusion argument owe us much more rigorous formulations of these premises and the relevant notions, and a much more convincing justification of why exactly we should accept them.

## 7 Causation-as-Production: A Dilemma

As we have noted above, Kim now sympathizes with Dowe's conserved quantity theory, and sometimes even seems to suggest that the exclusion argument in fact requires or presupposes this theory, or something very similar. I contend that this line is less workable than it may first appear to be.

To begin with, Kim seems to greatly underestimate the troubles with Dowe's theory. Though it enjoyed at first certain amount of popularity, as we have seen, the problems have accumulated. I think it is fair to say that few philosophers still find it attractive. Consequently, if the exclusion argument indeed presupposes this theory, or its kin, it is arguably on thin ice.

Furthermore, quite independently of such difficulties—even if we take Dowe's theory for granted—tying the exclusion argument to this theory of causation, presupposing it in this context, is argumentatively problematic. Dowe himself writes:

According to Dowe the relata in true “manifest” (common sense) claims of causation *must be translated* to physical states of the sort discussed above (“object *a* has a value *q* of a conserved quantity”) such that the manifest causal claim supervenes on some physical causation. Even for purely physical cases such as “chalking the ball” this is a complicated matter, and it is not obvious that it can be carried through.

Even if this could be made to work in purely physical cases, there remain questions about mental causation, causation in history, and causation in other branches of science besides physics ... In any case, to suppose that the conserved quantity theory will deal with causation in other branches of science also requires commitment to a *fairly thorough going reductionism*, since clearly there is nothing in economics or psychology that could pass for a conservation law. (Dowe 2008; my emphasis)

Dowe also reflects on some kind of causal pluralism as an alternative: perhaps the conserved quantity theory works at the level of fundamental physics, but a different theory is needed to take care of causation in various higher-level special sciences. In any case, it is undeniable that Dowe's theory directly applies only in the domain of fundamental physics.

This results in a dilemma: Either we must presuppose an all-encompassing and strong reductionism from the outset, or we must allow the possibility that Dowe's conserved quantity theory simply does not at all cover mental causation and other higher-level causation. In the

former case, the exclusion argument becomes redundant<sup>43</sup> and the overall argument circular: The hoped-for conclusion is already presupposed. In the latter case, considerations about mental causation and other (allegedly) higher-level causation must be based on a different theory of causation—and Kim is simply wrong in assuming that a theory like Dowe’s can and must be applied here.

In other words, if the exclusion argument in reality presupposes a specific theory of causation, and that theory in turn presupposes a strong and all-encompassing reductive view, it seems to beg the question to appeal to the exclusion argument in defense of a reductive view. That is to say, the exclusion argument seems to become quite pointless. If we assume that the exclusion argument presupposes such a theory of causation, some other argument, distinct from the exclusion argument and independent from the theory of causation, is needed in favor of such a strong reductionist view. However, there is no such argument on the horizon.

Alternatively, we need a wholly different theory of causation for the higher-level special sciences, and mental causation in particular: Dowe’s theory applies at best at the fundamental physical level. But in that case, it is not legitimate to begin with assuming Dowe’s theory as a generally correct account of causation for mental properties and such. In either case, the causation-as-production view cannot be used to back up the exclusion argument.

## 8 Kim against the Counterfactual Theory of Causation

Kim admits that the counterfactual theory of causation has “much intuitive plausibility” (Kim 2006b, p. 189). In the end, though, he contends that it cannot be the ultimate truth about causation (see Kim 2006b, 2007).<sup>44</sup> He argues that counterfactuals depend on laws and regularities, and that, consequently, the counterfactualist approach in fact reduces the regularity theory and shares its well-known problems: “embracing the counterfactual approach to causation will have no advantages over the regularist-nomological approach ... the counterfactual approach seems to presuppose, or collapse to, the nomological conception and thereby inherit the latter’s shortcomings” (Kim 2007, pp. 233–236).

Kim invites us to consider the following causal claim:

The striking of the match caused it to light.

He then notes that according to the counterfactual theory, whether this is true or not depends on (roughly) whether or not the following counterfactual conditional is true:

(C) If this match had not been struck, it would not have lighted.

But how do we know that the latter is true? Kim replies thus: “the obvious, and the only possible answer seems to be that, in the actual world, dry matches struck in the presence of oxygen usually and reliably ignite” (Kim 2007, p. 234). It is “our knowledge of this regularity, or law, combined with knowledge of the actual circumstances in which the match was struck (e.g., it was dry, oxygen was present, etc.)” that according to Kim accounts for our judgment that (C) is true (*ibid.*).

---

<sup>43</sup> Hitchcock (2012) makes the same observation in passing. And perhaps this is, more or less, what Loewer (2002, 2007) had in mind too. Still, I think that the point is important and insufficiently understood in the literature, and, consequently, deserves to be developed in greater detail here in the form of a dilemma.

<sup>44</sup> For a further useful discussion of Kim and the counterfactual theory, with a somewhat different emphasis, see Loewer 2007; cf. Loewer 2002.

Kim is thus apparently claiming that our knowledge of counterfactual conditionals of the form:

If *c* had not occurred, *e* would not have occurred

reduces, in general, to knowing the regularity that events of type *C* (events that are like *c*) are recurrently followed by events of type *E* (events like *e*). However, cases such as our earlier example of low-level radiation and cancer (see Section 5) clearly show that that this is emphatically not true. Though various regularities (often regularities other than the one between *c* and *e*) may indeed provide us clues for our *knowledge* of counterfactual conditionals, the relation between them is in reality much more complicated than Kim seems to think. Our knowledge of such counterfactual dependencies may be based, for example, on our careful analysis of the underlying mechanism, or/and on active experiments, and so on.

Kim's critique suggests that he implicitly views us as mere passive observers of "constant conjunctions", that is, observable regularities. But as researchers, we can also actively examine reality and conduct experiments; by manipulating the world, we can often distinguish misleading regularities that have a preceding common cause, for example, from genuine cases of causation. Assume, let us say, that we have observed a regular connection between certain values of *X* and *Y*. According to the regularity theory, this amounts to causation, period. According to the interventionist approach, by contrast, we may begin to examine whether this is, after all, a mere correlation, perhaps due to a common cause, (a certain value of) *Z*. We can then, for example, keep the value of *X* fixed and vary the value of *Z*; if the latter results changes in the value of *Y*, we can, at least tentatively, conclude that it is *Z* rather than *X* that is causally relevant to *Y*. Of course, we can later begin to wonder whether some further *W* is a common cause of *Z* and *Y*, and so forth *ad infinitum*—but that is how scientific knowledge makes progress. Consequently, the counterfactual approach, especially the interventionist approach, does not, *pace* Kim, collapse into the regularity theory and inherit its problems.

Furthermore, Kim seems to conflate two distinct issues in his critique: the local epistemological question of how we can know whether a particular pair of events is a case of genuine causation (and not mere correlation), and the general metaphysical question of what it takes for a pair of events to count as a case of real causation. Our knowledge of particular causal relations is empirical and, of course, always fallible and in principle revisable by further observations and experiments. In the worst case, perhaps, particular causal relations may be unknowable, at least in practice. But as philosophers, we may nevertheless be able to analyze what is required for *c* to cause *e*, in contrast to mere correlation. The latter does not require that we can actually decide in every concrete situation whether it really is a case of genuine causation.

In sum, Kim's critique of the counterfactual theory focuses only on an intuitive and quite naïve form of the counterfactual approach, and it makes some problematic moves. The difficulties he presents for it can well be circumvented by more sophisticated theories, and by the interventionist theory in particular.

## 9 Conclusions

We have paid close attention to Kim's views on causation, mental causation, and the exclusion argument, and their relation to different theories of causation. It turns out that Kim's view is less stable than is often thought. It seems that, of the different theories of causation, Kim bets on the wrong horse. In any case, we have shown that under closer scrutiny, the exclusion argument and its premises do not really cohere that well with any well-developed theory of causation. And from the perspective of the arguably most defensible theory—the interventionist

theory—the exclusion argument begins to waver in many respects. Consequently, it does not threaten non-reductive views.

## References

- Armstrong, David M. 1968. *A Materialist Theory of the Mind*. London: Routledge.
- Árnadóttir, Steinvör Thöll & Crane, Tim. 2013. “There is No Exclusion Problem.” In *Mental Causation and Ontology*. Eds. S.C. Gibb et al. Oxford: Oxford University Press, pp. 248–66.
- Baker, Lynne Rudder. 1993. “Metaphysics and Mental Causation.” In *Mental Causation*. Eds. J. Heil and A. Mele. Oxford: Clarendon Press, pp. 75–96.
- Baumgartner, Michael. 2009. “Interventionist Causal Exclusion and Non-reductive Physicalism.” *International Studies in the Philosophy of Science* 23 (2): 161–178.
- \_\_\_\_\_. 2010. “Interventionism and Epiphenomenalism.” *Canadian Journal of Philosophy* 40 (3): 359–383.
- \_\_\_\_\_. 2013. “Rendering Interventionism and Non-Reductive Physicalism Compatible.” *Dialectica* 67: 1–27.
- Bechtel, William and Jennifer Mundale. 1999. “Multiple Realizability Revisited: Linking Cognitive and Neural States.” *Philosophy of Science* 66: 175–207.
- Bennett, Karen. 2003. “Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract It.” *Nous* 37 (3): 471–497.
- \_\_\_\_\_. 2007. “Mental Causation.” *Philosophy Compass* 2 (2): 316–337.
- \_\_\_\_\_. 2008. “Exclusion Again.” In *Being Reduced: New Essays on Reduction, Explanation, and Causation*. Eds. Jakob Hohwy & Jesper Kallestrup. Oxford University Press, pp. 280–305.
- Bickle, John. 1998. *Psychoneural Reduction: The New Wave*. Cambridge, MA: MIT Press.
- \_\_\_\_\_. 2008. “Multiple Realizability.” *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition). Ed. Edward N. Zalta. URL = <<http://plato.stanford.edu/archives/fall2008/entries/multiple-realizability/>>.
- Carey, Brandon. 2011. “Overdetermination and the Exclusion Problem.” *Australasian Journal of Philosophy*, 89: 251–262.
- Collins, J., Hall, E., and Paul, L., 2004. *Causation and Counterfactuals*. Cambridge, MA: MIT Press.
- Crane, Tim. 2001. *Elements of Mind*. Oxford: Oxford University Press.
- Davidson, Donald. 1970. “Mental Events.” Reprinted in *Essays on Actions and Events* (Oxford: Clarendon Press, 1980), pp. 207–224.
- Davidson, Donald. 1974. “Psychology as Philosophy.” Reprinted in *Essays on Actions and Events* (Oxford: Clarendon Press, 1980), pp. 229–239.
- Dowe, Phil. 1992. “Wesley Salmon’s Process Theory of Causality and the Conserved Quantity Theory.” *Philosophy of Science* 59: 195–216.
- \_\_\_\_\_. 2000. *Physical Causation*. New York: Cambridge University Press, 2000.
- \_\_\_\_\_. 2008. “Causal Processes.” In *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition). Ed. Edward N. Zalta. URL = <<http://plato.stanford.edu/archives/fall2008/entries/causation-process/>>.
- Feigl, Herbert. 1958. “The ‘Mental’ and the ‘Physical’.” In *Concepts, Theories and the Mind-Body Problem* (Minnesota Studies in the Philosophy of Science, Volume 2). Ed. H. Feigl, M. Scriven and G. Maxwell. Minneapolis: University of Minnesota Press, pp. 370–497.
- Funkhouser, Eric. 2002. “Three Varieties of Causal Overdetermination.” *Pacific Philosophical Quarterly* 83: 335–351.
- \_\_\_\_\_. 2007. “Multiple Realizability.” *Philosophy Compass* 2/2 (2007): 303–315.
- \_\_\_\_\_. 2014. “A Call for Modesty: A Priori Philosophy and the Mind-Body Problem.” In *New Waves in Philosophy of Mind*. Eds. M. Sprevak & J. Kallestrup. Hampshire: Palgrave, pp. 21–37.

- Garfinkel, Alan. 1981. *Forms of Explanation*. New Haven: Yale University Press.
- Gibb, Sophie. 2014. "Mental Causation." *Analysis* 74(2): 327–338.
- Hall, Ned. 2004. "Two Concepts of Causation." In Collins et al. 2004, pp. 225–276.
- Hart, H. L. and Honore, A., 1965. *Causation in the Law*, Oxford: Clarendon Press. Second edition 1985.
- Hausman, Daniel M. 2002. "Physical Causation." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 33 (4): 717–724.
- Hitchcock, Christopher. 1995. "Salmon on Explanatory Relevance." *Philosophy of Science* 62 (2): 304–320.
- \_\_\_\_\_. 1996a. "The Role of Contrast in Causal and Explanatory Claims." *Synthese* 107 (3): 395–419.
- \_\_\_\_\_. 1996b. "Farewell to Binary Causation." *Canadian Journal of Philosophy* 26: 335–364.
- \_\_\_\_\_. 2008. "Causation." In *The Routledge Companion to Philosophy of Science*. Eds. M. Curd & S. Psillos. London: Routledge, pp. 361–370.
- \_\_\_\_\_. 2012. "Theories of Causation and the Causal Exclusion Argument." *Journal of Consciousness Studies* 19 (5–6): 40–56.
- Hoffmann-Kolss, Vera. 2014. "Interventionism and Higher-Level Causation." *International Studies in the Philosophy of Science* 28 (1): 49–64.
- Kim, Jaegwon. 1989a. "The Myth of Nonreductive Materialism." *Proceedings and Addresses of the American Philosophical Association* 63 (3): 31–47.
- \_\_\_\_\_. 1989b. "Mechanism, Purpose, and Explanatory Exclusion." *Philosophical Perspectives* 3: 77–108.
- \_\_\_\_\_. 1992. "'Downward Causation' in Emergentism and Nonreductive Physicalism." In *Emergence or Reduction?* Ed. A. Beckermann et al. Berlin: Walter de Gruyter, pp. 19–138.
- \_\_\_\_\_. 1993a. "The Non-Reductivist's Troubles with Mental Causation." In *Mental Causation*. Eds. J. Heil and A. Mele Oxford: Clarendon Press, pp. 189–210.
- \_\_\_\_\_. 1993b. "Postscripts on Mental Causation." In J. Kim, *Supervenience and Mind*. Cambridge: Cambridge University Press, pp. 358–367.
- \_\_\_\_\_. 1998. *Mind in a Physical World*. Cambridge, MA: MIT Press.
- \_\_\_\_\_. 2002. "Responses." *Philosophy and Phenomenological Research* 65: 671–680.
- \_\_\_\_\_. 2005. *Physicalism, or Something Near Enough*. Princeton: Princeton University Press.
- \_\_\_\_\_. 2006a. "Emergence: Core Ideas and Issues." *Synthese* 151 (3): 547–559.
- \_\_\_\_\_. 2006b. *Philosophy of Mind*, 2nd edition. Boulder: Westview Press.
- \_\_\_\_\_. 2007. "Causation and Mental Causation". In *Contemporary Debates in Philosophy of Mind*. Eds. Brian P. McLaughlin & Jonathan D. Cohen. Oxford: Blackwell, pp. 227–242.
- \_\_\_\_\_. 2009. "Mental Causation." In *The Oxford Handbook of Philosophy of Mind*. Eds. Brian McLaughlin, Ansgar Beckermann & Sven Walter. Oxford: Oxford University Press, pp. 29–52.
- \_\_\_\_\_. 2014. "Supervenience and Causal Closure: Two Exclusion Arguments and How in the End They Reduce to One." *CHEOLHAK*, Korean Philosophical Association Vol. 118, February 2014, 1–25.
- Kitcher, Philip. 1989. "Explanatory Unification and the Causal Structure of the World." In *Scientific Explanation. Minnesota Studies in the Philosophy of Science Volume XIII*. Ed. Philip Kitcher and Wesley C. Salmon. Minneapolis: University of Minnesota Press, pp. 410–505.
- Kroedel, Thomas. 2015. "Dualist Mental Causation and the Exclusion Problem." *Noûs* 49: 357–375.
- Kyburg, Henry. 1964. "Comment." *Philosophy of Science* 32: 147–51.
- LePore, Ernst & Barry Loewer 1987. "Mind Matters." *Journal of Philosophy* 84: 630–642.
- Lewis, David. 1966. "An Argument for the Identity Theory." *Journal of Philosophy* 63: 17–25.
- \_\_\_\_\_. 1972. "Psychophysical and Theoretical Identifications." *Australasian Journal of Philosophy* 50: 249–258.
- \_\_\_\_\_. 1973. "Causation." *Journal of Philosophy*, 70: 556–567.

- \_\_\_\_\_. 1986a. "Postscripts to 'Causation'." In D. Lewis, *Philosophical Papers: Volume II*. Oxford: Oxford University Press, pp. 172–213.
- \_\_\_\_\_. 1986b. "Causal Explanation." In D. Lewis, *Philosophical Papers: Volume II*. Oxford: Oxford University Press, pp. 214–240.
- \_\_\_\_\_. 2000. "Causation as Influence." *Journal of Philosophy* 97: 182–197.
- \_\_\_\_\_. 2004. "Causation as Influence." In Collins et al. 2004, pp. 75–106.
- Lipton, Peter. 1991. *Inference to the Best Explanation*. London: Routledge.
- List, Christian & Peter Menzies. 2009. "Nonreductive Physicalism and the Limits of the Exclusion Principle." *The Journal of Philosophy* CVI (9): 182–197.
- Loewer, Barry. 2002. "Comments on Jaegwon Kim's *Mind in a Physical World*." *Philosophy and Phenomenological Research*, 65: 655–662.
- \_\_\_\_\_. 2007. "Mental Causation, or Something Near Enough." In *Contemporary Debates in the Philosophy of Mind*. Eds. B. McLaughlin & J. Cohen. Malden: Blackwell, pp. 243–264.
- Mackie, John L. 1965. "Causes and Conditions." *American Philosophical Quarterly* 2: 245–264.
- Malcolm, Norman. 1968. "The Conceivability of Mechanism." *The Philosophical Review* 77: 45–72.
- Maslen, Cei. 2004. "Causes, Contrasts, and the Nontransitivity of Causation". In Collins et al. 2004, pp. 341–357.
- Menzies, Peter. 2004. "Difference-Making in Context", in Collins et al. 2004, pp. 139–80.
- \_\_\_\_\_. 2007. "Causation in Context." In *Causation, Physics, and the Constitution of Reality*. Eds. H. Price & R. Corry. Oxford: Oxford University Press, pp. 191–223.
- \_\_\_\_\_. 2008. "Exclusion Problem, the Determination Relation, and Contrastive Causation." In *Being Reduced—New Essays on Reduction, Explanation and Causation*. Ed. J. Hohwy and J. Kallestrup. Oxford: Oxford University Press, pp. 196–217.
- \_\_\_\_\_. 2009. "Platitudes and Counterexamples." In *The Oxford Handbook of Causation*. Eds. H. Beebe et al. Oxford: Oxford University Press, pp. 341–367.
- \_\_\_\_\_. 2014. "Counterfactual Theories of Causation." In *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition). Ed. Edward N. Zalta. URL = <http://plato.stanford.edu/archives/spr2014/entries/causation-counterfactual/>
- Menzies, Peter & Christian List. 2010. "The Causal Autonomy of the Special Sciences." In *Emergence in Mind*. Ed. G. and C. Macdonald. Oxford: Oxford University Press, pp. 108–128.
- Mill, John Stuart. 1843/1882. *A System of Logic*. 8th edition. New York: Harper & Brothers.
- Montero, Barbara. 2003. "Varieties of Causal Closure." In *Physicalism and Mental Causation*. Eds. Sven Walter & Heinz-Dieter Heckmann. Exeter: Imprint Academic, pp. 173–187.
- Ney, Alyssa. 2016. "Microphysical Causation and the Case for Physicalism." *Analytic Philosophy* 57: 141–164.
- Northcott, Robert. 2008. "Causation and Contrast Classes." *Philosophical Studies* 139: 111–123.
- Papineau, David. 1993. *Philosophical Naturalism*. Oxford: Blackwell.
- \_\_\_\_\_. 2001. "The Rise of Physicalism." In *Physicalism and Its Discontents*. Eds. B. Loewer & C. Gillett. Cambridge: Cambridge University Press, pp. 3–36.
- \_\_\_\_\_. 2009. "The Causal Closure of the Physical and Naturalism." In *The Oxford Handbook of Philosophy of Mind*. Eds. Brian McLaughlin, Ansgar Beckermann & Sven Walter. Oxford: Oxford University Press, pp. 53–65.
- Peacocke, Christopher. 1979. *Holistic Explanation*. Oxford: Clarendon Press.
- Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
- Pereboom, Derk. 2002. "Robust Nonreductive Materialism." *Journal of Philosophy* 99 (10): 499–531.
- Psillos, Stathis. 2002. *Causation and Explanation*. Stocksfield: Acumen.

- Pernu, Tuomas. 2013a. "Does the Interventionist Notion of Causation Deliver Us From the Fear of Epiphenomenalism?" *International Studies in the Philosophy of Science* 27: 157–172.
- \_\_\_\_\_. 2013b. "The Principle of Causal Exclusion Does Not Make Sense." *The Philosophical Forum* 44 (1): 89–95.
- \_\_\_\_\_. 2014a. "Interventions on Causal Exclusion." *Philosophical Explorations* 17: 255–63.
- \_\_\_\_\_. 2014b. "Causal Exclusion and Multiple Realizations." *Topoi* 33: 525–530.
- Polger, Thomas. 2004. *Natural Minds*. Cambridge, MA: MIT Press.
- Putnam, Hilary. 1967. "Psychological Predicates." In *Art, Mind, and Religion*. Ed. W. H. Capitan and D. D. Merrill. Pittsburgh: University of Pittsburgh Press, pp. 37–48.
- Raatikainen, Panu. 2010. "Causation, Exclusion, and the Special Sciences." *Erkenntnis* 73: 349–363.
- Robb, David, and Heil, John. 2014. "Mental Causation." In *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition.) Ed. Edward N. Zalta. URL = <<http://plato.stanford.edu/archives/spr2014/entries/mental-causation/>>.
- Salmon, Wesley. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.
- \_\_\_\_\_. 1994. "Causality Without Counterfactuals." *Philosophy of Science* 61 (2): 297–312.
- Schaffer, Jonathan. 2005. "Contrastive Causation." *Philosophical Review* 114: 297–328.
- Schiffer, Stephen. 1987. *Remnants of Meaning*. Cambridge: Bradford.
- Shapiro, Lawrence. 2000. "Multiple Realizations." *Journal of Philosophy* 97: 635–654.
- \_\_\_\_\_. 2010. "Lessons from Causal Exclusion." *Philosophy and Phenomenal Research* 81 (3): 594–604.
- Shapiro, Larry & Eliot Sober. 2007. "Epiphenomenalism—The Dos and Don'ts." In *Thinking about Causes*. Ed. Peter Machamer and Gereon Wolters. Pittsburgh: University of Pittsburgh Press, pp. 235–264.
- Shoemaker, Sydney. 2001. "Realization and Mental Causation." In *Physicalism and Its Discontents*. Ed. C. Gillett & B. Loewer. Cambridge: Cambridge University Press, pp. 74–98.
- Smart, J.J.C. 1959. "Sensations and Brain Processes." *Philosophical Review* 68: 141–156.
- Spirtes, Peter, Clark Glymour and Richard Scheines. 2000. *Causation, Prediction, and Search*, 2nd ed. New York: MIT Press.
- Woodward, James. 1984. "A Theory of Singular Causal Explanation." *Erkenntnis* 21: 231–62.
- \_\_\_\_\_. 1997. "Explanation, Invariance, and Intervention." *Philosophy of Science* Vol. 64, Supplement. Proceedings of the 1996 Biennial Meetings of the Philosophy of Science Association. Part II: Symposia Papers (Dec., 1997), pp. S26–S41.
- \_\_\_\_\_. 2000. "Explanation and Invariance in the Special Sciences." *The British Journal for the Philosophy of Science* 51 (2): 197–254.
- \_\_\_\_\_. 2003. *Making Things Happen*. Oxford: Oxford University Press.
- \_\_\_\_\_. 2008. "Mental Causation and Neural Mechanisms." In *Being Reduced—New Essays on Reduction, Explanation and Causation*. Ed. J. Hohwy and J. Kallestrup. Oxford: Oxford University Press, pp. 218–262.
- \_\_\_\_\_. 2015. "Interventionism and Causal Exclusion." *Philosophy and Phenomenological Research* 91: 303–347.
- \_\_\_\_\_. 2016. "Causation in Science." In *The Oxford Handbook of Philosophy of Science*. Ed. P. Humphreys. New York: Oxford University Press, pp. 163–184.
- \_\_\_\_\_. 2017. "Intervening in the Exclusion Argument." In *Making a Difference*. In Eds. H. Beebe et al. Oxford: Oxford University Press, pp. 251–268.
- Yablo, Stephen. 1992. "Mental Causation." *The Philosophical Review* 101 (2): 245–280.
- Yalowitz, Steven. 2014. "Anomalous Monism; Supplement: Causal Closure of the Physical in the Argument for Monism." *The Stanford Encyclopedia of Philosophy* (Winter 2014 Edition). Ed. Edward N. Zalta. URL = <<http://plato.stanford.edu/entries/anomalous-monism/causal-closure.html>>.

- Yoo, Julie. 2007. "Mental Causation." *The Internet Encyclopedia of Philosophy*. <http://www.iep.utm.edu/mental-c/>
- Zhong, Lei. 2014. "Sophisticated Exclusion and Sophisticated Causation." *The Journal of Philosophy* CXI (7): 341–360.