



Explanatory exclusion and mental explanation

Dwayne Moore

To cite this article: Dwayne Moore (2016) Explanatory exclusion and mental explanation, *Philosophical Psychology*, 29:3, 390-404, DOI: [10.1080/09515089.2015.1119263](https://doi.org/10.1080/09515089.2015.1119263)

To link to this article: <http://dx.doi.org/10.1080/09515089.2015.1119263>



Published online: 22 Feb 2016.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

Explanatory exclusion and mental explanation

Dwayne Moore

Philosophy Department, University of Saskatchewan, Saskatoon, Canada

ABSTRACT

Jaegwon Kim once refrained from excluding distinct mental causes of effects that depend upon the sufficient physical cause of the effect (Kim, 1984). At that time, Kim also refrained from excluding distinct mental explanations of effects that depend upon complete physical explanations of the effect (Kim, 1988, 1989). More recently, he has excluded distinct mental causes of effects that depend upon the sufficient cause of the effect, since the physical cause is individually sufficient for the effect (Kim, 2005). But there has been, to this point, no parallel shift in the explanatory realm, such that distinct mental explanations of effects that depend upon complete physical explanations of the effect are excluded since the physical explanation is objectively complete. In this paper I consider, defend, and apply this update to the principle of explanatory exclusion—an update, which, in the final analysis, demonstrates a significant advantage that non-reductive physicalism has over reductive physicalism.

KEYWORDS

Explanatory exclusion;
Jaegwon Kim; Mental
causation

Jaegwon Kim once refrained from excluding distinct mental causes of effects that depend upon the sufficient physical cause of the effect (Kim, 1984). At that time, Kim also refrained from excluding distinct mental explanations of effects that depend upon complete physical explanations of the effect (1988, p. 233; 1989, p. 89). More recently, he has excluded distinct mental causes of effects that depend upon the sufficient cause of the effect, since the physical cause is individually sufficient for the effect (Kim, 2005, p. 17). But there has been, to this point, no parallel shift in the explanatory realm, such that distinct mental explanations of effects that depend upon complete physical explanations of the effect are excluded since the physical explanation is metaphysically complete. In this paper I consider, defend and apply this update to the principle of explanatory exclusion—an update, which, in the final analysis, demonstrates a significant advantage that non-reductive physicalism has over reductive physicalism.

This paper is divided into seven sections. After outlining (section 1) the problem of explanatory exclusion, I motivate several modifications to the principle of explanatory exclusion (section 2). In particular, I introduce the principle of strong explanatory exclusion, according to which complete physical explanations render mental explanations *metaphysically* unnecessary, and hence *in principle* excludable. I then apply the principle of strong explanatory exclusion to a common reductive position (section 3). I consider a number of objections to this admittedly strong principle of explanatory exclusion (sections 4 to 6). I ultimately conclude that mental explanations are necessary, hence not excludable, according to non-reductive physicalism, but not according to reductive physicalism (section 7).

1. Explanatory Exclusion

Jaegwon Kim's famous principle of causal exclusion states that "no event can have more than one sufficient cause occurring at any given time" (2009, p. 39). The principle of causal exclusion is supported by, among other things, the principle of parsimony. Kim explains:

If this physiological event is indeed *sufficient* for the climbing, the climbing should occur whether or not any *other* event occurred. That is, no other event should be necessary for the occurrence of the climbing ... that the climbing would have occurred whether or not the rationalizing belief and desire occurred surely demonstrates the causal and explanatory irrelevance of the belief and desire. (1989, p. 82)

If the physical event is a sufficient cause for the climbing, then the mental cause is unnecessary, so it is possible to exclude the mental cause.

Less widely circulated is the fact that the principle of causal exclusion has a "companion principle," called the *principle of explanatory exclusion*, which states that "there can be no more than a single *complete* and *independent* explanation for any one event" (Kim, 1988, p. 233; 2005, p. 17). In fact, it is of historical worth to note that Kim's (1988, 1989) inaugural articulation of the exclusion problem occurs in the context of excluding superfluous *explanations*, where Kim only later introduces the principle of *causal* exclusion. What reason is there to accept this original principle of explanatory exclusion? Perhaps it is not surprising that explanatory exclusion, as the originative and companion principle to the principle of causal exclusion, is also supported by the principle of parsimony. This is evident in the passage cited above, where Kim argues for the "causal and *explanatory* irrelevance" of beliefs and desires, provided there is a sufficient/complete neural cause/explanation.

This is also evident in Kim's wider argumentation, according to which explanations track causal relations, and hence explanatory exclusion follows from causal exclusion. With respect to the former claim, Kim indicates that explanations track causation through the use of the principle of explanatory realism. In its barest form, the principle of explanatory realism says: "C is an explanans for E in virtue of the fact that c bears to e some determinate objective relation R" (1988, p. 226).¹ For Kim, it is "obvious" that the causal relation is either an "important case" of, or "the only" case of, objective relation R (1988, p. 226), so, for the purposes of this paper, it can be assumed that the objective relation R is a causal relation. This objective causal relation R between events "grounds" the explanans relation between propositions and serves as its "objective correlate":

The kind of view I have just alluded to, namely that a causal explanation of E in terms of C is a 'correct explanation' only if C is in reality a cause of E, can be called an 'objectivist' or 'realist' conception of explanation. And the view that explanations must be 'real' or 'objective' in this sense can be called explanatory realism. (Kim, 1989, p. 94; See also 1981, p. 307; 1988, p. 226)

The explanation of E in terms of C correlates with the objective causal relation between the event e and c, so a correct explanation will depict this objective relation.

Since the principle of causal exclusion states that there can be no more than one sufficient cause for event e, and since explanatory realism says that this exclusive causal relation correlates with and grounds explanations, the result is that there can be no more than one complete and independent explanation for event e either (Gibb, 2009, p. 3; McIntyre, 2002, p. 95). After all, as the principle of causal exclusion is motivated by Ockham's razor, so parsimony renders unnecessary explanations excludable:

It might be said that ... in order to justify something like the rule of explanatory exclusion, ... all we really need is Ockham's razor, the familiar principle of simplicity, that enjoins us to get by with the fewest possible entities, hypotheses, theoretical principles, and, of course, explanations. In reply ... we need to determine exactly at what point the entities in question begin to be multiplied 'beyond necessity'. In fact, determining where the excess baggage starts is the difficult part; the rest is trivial. The exclusion principle does the difficult work: it says that for any event more than one complete explanation is excess baggage. (Kim, 1989, p. 98)

In this passage Kim argues that the principle of explanatory exclusion is a "special case" of the general simplicity requirement which determines where Ockham's razor should cut (1989, p. 98). According to it, if a complete explanation for the occurrence of an event has already been provided, additional independent explanations are unnecessary. And if additional independent explanations are unnecessary, then, according to the principle of parsimony, additional independent explanations can be excluded.

It is possible to object that the principle of parsimony is properly applied to metaphysical entities, such as events and objects (i.e., ontological simplicity), but is improperly applied to theoretic entities, such as explanations and hypotheses (i.e., syntactic simplicity). It is evident from the preceding passage, however, that this is not Kim's approach. After all, while parsimony clearly applies to metaphysical entities,² parsimony is a general principle that applies equally to theoretical posits. After all, parsimony, in Ockham's original words, is a generic principle of simplicity: "Plurality must not be asserted without necessity" (1990, 47).³ Not surprisingly, then, theoretical parsimony is deeply engrained in the historical tradition as well, dating as far back as Aristotle:

We may assume the superiority *ceteris paribus* of the demonstration which derives from fewer postulates or hypotheses—in short from fewer premises; for ... where they are fewer, knowledge will be more speedily acquired. (*Posterior Analytics*, 86a33–36)

Here Aristotle alludes to the contemporary contention that the grammatical elegance of parsimonious theories facilitates both use and comprehension (Friedman, 1974; Kim, 1994). Others argue that parsimonious theories are more probably true, as they have fewer requisitely true propositions. Famous examples of theoretical parsimony include Copernicus' rejection of the Ptolemaic system on account of theoretical parsimony (Guach, 2003, p. 273), and Laplace's exclusion of God as a hypothesis, on account of God's being unnecessary for Laplace's explanation of the universe.

Assuming that the principle of parsimony adequately motivates the principle of explanatory exclusion, how does this create the explanatory exclusion problem for mental explanations? Simply put, if psychological explanations are independent from complete physical explanations, then they can be excluded. Explanations are independent if they are intensionally (conceptually) or logically inequivalent (Davidson, 1980, p. 171). This is the so-called intensional model of explanatory individuation—the most common model of explanatory individuation (Campbell, 2008, p. 86; Fuhrmann 2002, p. 184ff; Marras, 1998, p. 443)—and it is the model that I follow as well. According to the intensional model, only if two explanations have the same meaning, or are synonymous, do they state the same explanation. Psychological explanations, however, are intensionally inequivalent to physical explanations. For example, Socrates can have the following psychological explanation for drinking poison: he desires to honor the laws and believes that drinking poison will honor the laws, where this belief/desire pairing together constitutes Socrates' reason for drinking the poison. Socrates does not, however, have the physical explanation of this same piece of behavior, since he is oblivious to the neural mechanisms implementing this cognitive process. For his own part, Kim acknowledges this as well (1988, p. 233; 1989, p. 81). Since there is, in the words of Kim's primary target, Norman Malcolm, a "logical difference" between psychological explanations and physical explanations, there is a "collision between the two accounts" (1968, pp. 51–52). The result of this collision, as will be detailed below, is that the complete physical explanation excludes an independent mental explanation of the same occurrence.

2. Strong Explanatory Exclusion

In this section I strengthen the explanatory exclusion principle in one crucial respect. The strengthening of the principle of explanatory exclusion may seem counterintuitive. After all, Kim already suspects that "many will consider" the original principle of explanatory exclusion "absurdly strong and unacceptable" (1989, p. 79). True to form, numerous respondents have taken exactly this attitude towards it. For example, Jeeloo Liu notes: "Taken by itself, explanatory exclusion seems too strong. And indeed, it has aroused some suspicion that Kim is creating a pseudo-problem" (2001, p. 10; see also Marras, 1998, p. 439; McIntyre, 2002, pp. 92–93; Sabates, 1996, p. 96; Worley, 1993, p. 334;). Perhaps some of these concerns can be alleviated by the fact that my position allows a weakening of the explanatory exclusion principle in two ways. In either event, I begin by simply introducing the strengthened principle of explanatory exclusion:

Strong explanatory exclusion. There can be no more than a single complete explanation for any one causal relation between events.

First, a terminological modification. Kim's principle of explanatory exclusion states that there can be no more than a single complete and independent explanation for any *one event*. Strong explanatory exclusion states that there can be no more than a single complete explanation for any *one causal relation between events*. This slight modification occurs because, strictly speaking, "an explanation [is] a complex of propositions or statements ... divisible into two parts, explanans and explanandum proposition" (Kim, 1988, p. 225). An explanans statement is the statement that explains, while an explanandum statement is the statement that is explained, and the explanation statement is the complete explanation, which is the complex of both the explanans and the explanandum statements. The explanandum statement *E* refers to the effect, which is event *e*. The explanans statement *C* refers to the cause, which is the event *c*. The event *c* is related to the event *e* by the causal relation *R*, where the causal relation *R* is the explanatory relation (Kim, 1988, p. 226). Thus, the explanation statement refers to the explanatory relation *R*, which is the causal relation *R*, between cause *c* and effect *e*. For example, if the explanation states "The fire caused the smoke," then "the fire" is an explanans statement referring to the fire, "the smoke" is an explanandum statement referring to the smoke, and "the fire caused the smoke" is an explanation referring to the causal relation between the fire and the smoke.

Aside from this terminological amendment, there is one crucial strengthening of the principle of explanatory exclusion here. Kim's principle of explanatory exclusion countenances a complete explanation plus multiple dependent explanations, but then claims that other explanations beyond the dependent explanations can be excluded. Thus, for example, humans have some capability of recognizing whether one object is a rotated image of another (Shepard & Metzler, 1971). The psychological explanation of these capabilities is that the mind can perform rotation-like operations on mental images. The neural explanation is that the sensory-processing mechanism in the brain is active and realizes these cognitive processes. The latter explanation, for reasons that become apparent below, is a complete explanation. But, so long as the former explanation is dependent, via some sort of metaphysical dependency relation, upon the first, the former explanation can be included as well. The principle of strong explanatory exclusion countenances a complete explanation, but then claims that other explanations beyond the complete explanation, including dependent explanations, can be excluded. For example, tiredness is the traditional psychological explanation of sleep. The physical explanation of sleep may involve fortifying and/or remodeling neural connections (Poe, Walsh, & Bjorness, 2010). The latter physical explanation is complete, so the former psychological explanation is unnecessary and can be excluded.

What motivation is there for accepting the principle of strong explanatory exclusion? First of all, the principle of strong explanatory exclusion is motivated by Kim's recent shift in his views on the principle of causal exclusion. When Kim introduced the principle of explanatory exclusion, he argued that distinct but dependent explanations of the same event were not to be excluded. At the time, he also maintained that distinct but dependent causes of the same event were not to be excluded either:

If a pain causes the sensation of fear an instant later, this account tells the following story: the pain is supervenient on a brain state, this brain state causes another appropriate brain state, and given this second brain state, the fear sensation must occur, for it is supervenient upon that brain state. ... It seems to me that this is sufficient to redeem the causal powers we ordinarily attribute to mental events. (Kim, 1993, pp. 106–107)

Kim goes on to argue that even though there is a sufficient physical cause for fear, pain can still be included as a cause for the fear by virtue of the fact that pain strongly supervenes upon, and is dependent upon, the physical cause of the fear.

More recently, Kim argues that these distinct but dependent mental causes of the effect can be excluded. His reasoning is that the physical cause is sufficient on its own, and these mental causes are like shadows which, though dependent, still only come along for the ride:

So why not say that *M* [the mental event], though it doesn't quite have the causal status of *P* [the physical event] in relation to *P** [the effect], is a derivative cause of *P** in virtue of its supervenience on *P*? ... Some years back, I thought that this might be a plausible way of vindicating mental causation. But it soon began to dawn on me that this was an empty verbal ploy. ... This is only a gimmick with no meaning; the facts are [that *P* causes *P**], and inserting [*M* causes *M**] and calling it supervenient causation ... does not alter the situation one bit. (Kim, 2005, p. 62; see also 1998, p. 37)

An honest look caused Kim to see that the dependent mental event can be excluded, since the physical cause is sufficient on its own to bring about the effect. In effect, the principle of causal exclusion states that since one cause is sufficient, additional causes, though dependent upon the sufficient cause, can be excluded. While the principle of causal exclusion is not universally endorsed (Loewer, 2002; Sider, 2003), it is generally considered a plausible and challenging constraint (Harbecke, 2008, p. 28; Schiffer, 1987, p. 148), if not “virtually an analytic truth” (Kim, 2005, p. 51).

Since the principle of explanatory exclusion is a companion principle that, by virtue of explanatory realism, runs parallel to the principle of causal exclusion, a similar amendment seems warranted for the principle of explanatory exclusion.⁴ Namely, though it might appear plausible to include dependent causes, one can see that even dependent mental causes can be excluded by sufficient physical causes. By parity of reasoning, though it might appear plausible to include dependent explanations, one can see that even dependent mental explanations can be excluded by complete physical explanations—which is the principle of strong explanatory exclusion. In fact, due to this parity of reasoning, to deny the prima facie plausibility of the principle of strong explanatory exclusion is akin to denying the prima facie plausibility of the causal exclusion principle itself, which is not advisable since the principle of causal exclusion is currently wreaking havoc in the contemporary mental causation debate.

Not only does the reasoning leading to the principle of strong explanatory exclusion follow the reasoning leading to the principle of causal exclusion, but also the underlying argumentation used in support of the principle of causal exclusion can be used in support of the principle of strong explanatory exclusion. That is, as the principle of parsimony suggests, the terms ‘necessity’ and ‘sufficiency’ stand in the following relation: if *A* is sufficient for *C*, *B* is unnecessary for *C* (Brand 1976, p. 41; Kim, 1989, p. 82; Marras, 2007, p. 319). Thus, for the causal exclusion principle, if *a* is a sufficient cause of the effect, then *a*, by itself, guarantees that the effect will occur. Therefore, no other cause is necessary for the effect, so *b* is not a necessary cause of the effect (Kim, 1998, pp. 44–45; 2005, pp. 48–49). Likewise, with respect to strong explanatory exclusion, if *A* is a complete explanation of the causal relation between events, then *B* is, in principle, not necessary as an explanation of the same causal relation between events.

Here is where the principle of parsimony, as worked out above, enters the picture. If secondary and tertiary explanations are not, in principle, necessary, then the principle of parsimony indicates that they can be excluded. As it turns out, Ockham’s razor is sharper than Kim originally envisaged. Ockham’s razor states that every additional explanation, including dependent explanations, beyond the complete explanation can be excluded. Since these additional explanations can be excluded, it follows that there can be no more than the single complete explanation for a given effect, which is the principle of strong explanatory exclusion.

3. Ontological Reduction and Conceptual Dualism

The principle of strong explanatory exclusion counts as an objection to the increasingly common doctrine of ontological monism yoked with conceptual dualism (Esfeld, 2007, p. 218ff; Heil, 2003, 2013; Hill & McLaughlin, 1999; Kirk, 2013; Levin, 2007; Loar, 1990; Papineau, 2002; Peacocke, 1989; Tye, 2000). In fact, David Papineau suggests that this position is “quickly becoming the orthodoxy” (2002, p. 5). Here are some recent formulations of the doctrine:

The philosophical mistake is to imagine that sameness of word implies sameness of worldly correspondent. ... Reality has but one level. ... We may find it occasionally useful to speak of levels of description or explanation, but these must not be confused with levels of being or encourage the image of a layered world. (Heil, 2003, pp. 49–50)

Even if conscious states are material states at the ontological level, we have two different ways of thinking about these states at the conceptual level. As well as thinking of them as material states, we can also think of them as feelings, by using special ‘phenomenal concepts’. (Papineau, 2002, pp. 4–5)

New wave materialists hold that (a) phenomenal concepts directly refer to phenomenal properties, that (b) phenomenal concepts are not physical ... concepts, and, that (c) phenomenal properties are identical with physical ... properties. (McLaughlin, 2001, pp. 328–329)

Ontological reduction, combined with conceptual dualism, is often motivated as follows (see Davidson, 1993; Papineau, 2002, p. 17). First, it is probable that mental events have physical effects: my fear experience causes my increased heart rate and alertness, or my cognition of present danger causes my increased heart rate and alertness. However, it is also the case that these physical effects have sufficient physical causes: the amygdala releases hormones that cause my increased heart rate and alertness. The result is that some physical effects have both a sufficient physical cause and a mental cause. However, it is implausible to endorse systematic overdetermination, so it cannot be the case that physical effects have mental causes and distinct sufficient physical causes. Rather, it must be the case that physical effects have mental causes which are identical with sufficient physical causes. Hence the ontological reduction: my fear experience is identical to neural activity in the amygdala, and my cognitive processing of present danger is identical to neural processing in the sensory cortex or sensory data received from the thalamus.

At the same time, there appear to be intuitive distinctions between mental and physical concepts. The mental is essentially normative and rational, whereas the physical is not (Davidson, 1993). For example, the logical relation between my belief that this dog is dangerous, my desire to avoid danger, and my ensuing act of running cannot be translated into the neurochemical vocabulary of action potentials, hormone release, and the like. Others argue that the mental is essentially qualitative, while the physical is not (Papineau, 2002, p. 47ff). Subjective, first-person descriptions of the gripping, innervating experience of fear are distinct from the objective scientific vocabulary describing hormone release by the amygdala. So, mental descriptions, concepts, and explanations are distinct and independent from physical descriptions, concepts, and explanations of the same causal relation between events. Hence the conceptual dualism.

Importantly, the view that all effects have sufficient physical causes is commonly called the principle of physical causal completeness. This principle is motivated by various conservation laws in physics and numerous 20th-century scientific advances (Papineau, 2001). Those who endorse this principle are typically called physicalists. The principle of physical causal completeness has a companion principle in the explanatory realm called the principle of physical explanatory completeness, according to which all causal relations between sufficient physical causes and effects have complete physical explanations (Fuhrmann, 2002, p. 186; Gibb, 2009, p. 209).⁵ After all, since all physical effects have sufficient physical causes, it is reasonable that these sufficient physical causes of physical effects can be given complete physical explanations. As Kim says, “when the causal relation provides a sufficient cause, the explanans can also be said to be complete and sufficient” (1989, p. 234). Or, as David Papineau puts it, “prior physical factors will always suffice to give us as full an explanation of that result as is possible” (1995, pp. 228–229). So the causal relation between the physical cause and the physical effect has a complete physical explanation. Since the mental is ontologically identical to, but conceptually distinct from, the physical cause, this same causal relation between events has an independent mental explanation as well. Therefore, the same causal relation between events has a complete physical explanation and a distinct mental explanation.

This is where the principle of strong explanatory exclusion poses problems. Here is the problem: since the physical explanation of the causal relation between physical events is complete, the mental explanation of the same causal relation between events is, in principle, unnecessary.⁶ Since the mental explanation is, in principle, unnecessary, the mental explanation can be, in theory, excluded. I take it that any position that renders it possible, even in theory, to exclude mental explanations is unpalatable—unpalatable enough to seek out alternatives.

Versions of this argument have been introduced before. Andrew Melnyk, for example, offers the following lucid articulation of the exclusion pressures I am attempting to motivate:

Suppose you are also a physicalist, so that you believe (at least) that every event, without exception, just is some fundamental physical event. It seems to follow that every event, without exception, has a fundamental physical explanation. ... What, therefore, is the point of the explanations apparently supplied by the special sciences? ... They seem, indeed, quite needless, since they explain nothing that is not already explained. But if, like explanations citing phlogiston, they are explanatorily dispensable, surely we should dispense with them, just as we have dispensed with the phlogiston citing explanations. ... People might raise an awkward question: why keep it around? (1996, p. 185; see also Horgan, 1991; Jacqueline, 1994, pp. 34–36; Melnyk, 2003, pp. 166–169)

However, the intuition that Melnyk captures here, which I am attempting to motivate as well, is controversial and lacks a sustained defense. Indeed, even Melnyk does not ultimately endorse the position. And, other ontological monists would surely react incredulously to the suggestion that their conceptual dualism is in any way problematic. After all, as Árnadóttir and Crane explain:

There is not even a *prima facie* difficulty of mental explanations being incompatible with, or ‘crowded out by’ physical explanations. Any occurrence can be explained in countless ways, and there is no incompatibility between any physical explanation of an event and a mental explanation of the same event. So there seems to be little plausibility to the idea that one explanation ‘excludes’ another. (2013, p. 256)

In what remains of this paper, I attempt to provide a sustained defense of the difficulty that the principle of strong explanatory exclusion poses for those yoking ontological monism with conceptual dualism.

Before proceeding to this defense, I want to outline two important caveats, which weaken the principle of strong explanatory exclusion in important ways. First, it is common to accept multiple explanations of the same effect because explanations provide epistemic illumination, and multiplicities of explanations may be necessary to provide epistemic illumination to different people in different contexts with different background knowledge. André Fuhrmann, for example, notes that “the context relativity of full explanations leaves room for the massive occurrence of a plurality of explanations for a given explanandum” (2002, p. 184). To borrow his example, those lacking the requisite background may not gain epistemic illumination from the explanation stating ‘The baby in the tub cried because the mean kinetic energy of the H₂O molecules passed a certain threshold’. Thus, the explanation stating ‘The baby in the tub cried because the water was too hot’ may be epistemically necessary, so not excludable.

Fuhrmann is correct that multiplicities of explanations may be *epistemically* necessary, depending on the context. At the same time, however, multiple explanations are not *metaphysically* necessary. This is an important distinction that requires unpacking. I follow Kim, who argues that explanation is like knowledge in the sense that they both have a metaphysical component and an epistemic component (1988, p. 225). Knowledge has a metaphysical “true” component, and an epistemic “justificatory” component. Likewise, explanation has a metaphysical component that, according to explanatory realism, requires that the explanation correctly refer to the objective causal relation between events that grounds the explanation. And, explanation has an epistemic component that requires the explanation to be illuminating.

So, here is my first caveat: when I say that there can be no more than a single complete explanation for a causal relation between events, I have the metaphysical component of explanations in mind, and I do not have the epistemic component of explanations in mind. That is, I grant that there are contexts in which one explanation is metaphysically complete, even though additional explanations of the same causal relation between events would be necessary to satisfy the epistemic component of explanation. Rather, my claim is that if the first explanation is metaphysically complete, secondary and tertiary explanations are not metaphysically, or *in principle*, necessary. That is, if the first explanation is metaphysically complete, then there is no objective state of affairs that remains undescribed or under-described. Thus, secondary or tertiary explanations are not metaphysically necessary, meaning that secondary or tertiary explanations can, in theory or in practice, be silenced, and every causal relation between events that exists would have still been fully explained.

The second weakening of the principle of strong explanatory exclusion centers on the interpretation of ‘can’. The principle of strong explanatory exclusion states that there *can* be no more than a single complete explanation of a given causal relation between events. I do not need to insist on meaning *can* in the sense of ‘ought’ or ‘must’. That is, I do not need to insist that there must be no more than a single complete explanation of the same causal relation between events. Rather, I only need to insist on the weaker notion of *can*, in the sense of ‘it is possible’. That is, it is possible for there to be no more than a single complete explanation of the same causal relation between events.

Here are three examples which will hopefully clarify these caveats. Rock stars used to make records—physical objects made of vinyl which were spun around on record players. Now, rock stars produce digital files, which are downloaded and played through digital media applications. However,

rock stars still promote their latest ‘record’ as available on iTunes. Now imagine that Wallace, who grew up in the age of physical records, does not understand the concepts of ‘downloading’ and ‘digital files.’ Wallace asks his granddaughter, Suzie why she is so intently listening to her ear buds, and Suzie replies, “I downloaded the latest digital file from Coldplay.” Wallace, lacking the requisite background knowledge, receives no epistemic illumination. In this case, Suzie’s explanation is metaphysically complete—her explanation referred to the causal relation between the entire cause and effect, so no objective state of affairs is left that requires explanation. So additional explanations are metaphysically unnecessary—even if the term ‘record’ fell out of the English language entirely, no objective state of affairs would be left undescribed. Indeed, Suzie’s generation is already forgetting the term! But I do not insist that there cannot be additional explanation provided to Wallace, in terms of buying the latest Coldplay record. In the circumstance involving Wallace, additional explanations are epistemically fruitful. But, it is possible to do without it.

Second, imagine that in the year 2500, a group of brilliant scientists colonizes a new planet and decide to only use scientific vocabulary. When they pass by lakes, they speak of the H_2O in the lake. When the rain falls, they speak of the H_2O falling from the sky. A generation goes by, the word water is forgotten, yet everyone still has baths, of course, and the grass still grows. One day an earthling visits the new planet, and, upon noticing the lush grass, says: “I see you have plenty of water to make the grass grow.” The scientists are perplexed at the unusual nomenclature of the earthling visitor. In this case, the earthling’s water-explanation is metaphysically unnecessary—the colony of scientists omitted the term ‘water’ without missing anything in their accounts of biological and geological phenomena. But the earthling’s water-explanation may be epistemically necessary—for the earthling, the water-explanation may illuminate climate patterns on the new planet. So, the water-explanation can be excluded—after not saying ‘water’ for a generation, nothing was left unsaid for a generation.

Third, Richard Rorty tells the story of a race of Antipodeans who resemble humans in every way except that the first disciplines that they master are neurology and chemistry, so they leave out mental explanations entirely, in favor of the neural explanations alone. He claims that “no predictive or explanatory or descriptive power would be lost if we had spoken Antipodean all our lives” (Rorty, 2009, p. 120). That is, given physical explanatory completeness, neurochemical explanations of human behavior are metaphysically complete. The Antipodean vocabulary leaves nothing out, neither descriptive nor predictive. For humans, therefore, psychological vocabulary may be epistemically necessary, due to the complexity or long-windedness of neural vocabulary. But it is possible to do without psychological vocabulary entirely, as the Antipodeans do.

4. Extensional Individuation

In what remains of this paper I defend the principle of strong explanatory exclusion from a number of considerations and objections. Here is one such consideration: it is possible to avoid strong explanatory exclusion pressures on mental explanations by endorsing a fully extensional model of explanatory individuation. Jaegwon Kim makes this move. He argues that explanatory realism implies that the causal relation between events grounds explanations, so “the content of the explanation consisting of C and E ” is that c and e are related by objective relation R (Kim, 1988, p. 226). The objective relation R between c and e “is what the explanation says” (Kim, 1988, p. 226), so if two nonequivalent causal statements refer to this same causal relation between events, then they state the same explanation. And, if they state the same explanation, then there are not two distinct explanations, but only one explanation (Kim, 1989, p. 87; 1988, p. 233; 1989, p. 80). And strong explanatory exclusion pressures only arise when there is more than a single explanation. So there is no pressure to exclude the mental explanation, for there is no more than a single explanation of the causal relation between events.

A number of criticisms have been leveled against Kim’s extensional model of explanatory individuation. These criticisms largely revolve around the strange consequences that arise from two conceptually distinct causal statements stating the same explanation (Furhmann, 2002; Marras, 1998).

Ausonio Marras, for example, argues that an epistemic component is necessary to explanations since explanations are supposed to provide some sort of epistemic relief—a fact with which Kim agrees. He then asks us to consider the following two explanations: “The earthquake caused the collapse of the building” and “the event that caused the collapse of the building caused the collapse of the building” (Marras, 1998, p. 443). While the first explanation helps us understand what happened, the second explanation is epistemically vacuous. However, on Kim’s model, to have the second explanation is to have the first explanation. The second causal statement refers to the same objective causal relation between events as the first one, so it is the same explanation as the first one. Marras recoils: “Anyone in possession of the one explanation is thereby in possession of the other. And this, I believe, is counter-intuitive” (1998, p. 443). Anyone who has the second explanation does not necessarily know that the earthquake caused the collapse of the building, so we have reason to believe that these two explanations remain distinct.

Kim’s solution also faces appropriately modified strong explanatory exclusion pressures. To see how, let us assume the extensional model is true, so “increased amygdala activation causes my increased heart rate” states the same explanation as “my fear caused my increased heart rate.” Even in this case, the syntax is different between “increased amygdala activation causes my increased heart rate” and “my fear caused my increased heart rate”—while they state the same explanation, they do not use the same vocabulary. Now we return to a familiar question: is the statement “increased amygdala activation (along with any other required physical processes) causes my increased heart rate,” which uses only physical vocabulary, complete or not? Assuming physical explanatory completeness, the answer is yes. Since “increased amygdala activation causes my increased heart rate” is complete, it is not necessary to add “my fear caused my increased heart rate.” Everything has already been said, so nothing more needs to be said.

One can object: but, we *can* add “my fear caused my increased heart rate,” since “my fear caused my increased heart rate” states the same thing as “increased amygdala activation caused my increased heart rate,” so there is nothing new being added.⁷ True, but we *need not* add “my fear caused my increased heart rate,” since “increased amygdala activation caused my increased heart rate” is already complete. We could be Antipodeans. We could live without psychological vocabulary. And, if we do, nothing will be left unspoken, and the world will not be under-described. Any position that renders it possible to eliminate psychological vocabulary in this manner is not as appealing as a position that renders psychological vocabulary necessary.

5. Intensional Individuation

Kim’s solution to the problem of explanatory exclusion formulates ontological identity from within an extensional model of explanatory individuation. As we have seen, Kim’s proposal fails due to the difficulties associated with the extensional model of explanatory individuation. In contrast, there is also a set of philosophers that resolve the explanatory exclusion problem by formulating ontological identity from within an intensional model.

There are two central insights of the intensional model, as pioneered by Donald Davidson and popularized by Ausonio Marras. First, explanations are primarily, if not entirely, epistemic entities. Explanations are epistemic since they relieve confusion. Suppose Johnny is confused about why the baby is crying. Johnny’s mother explains that the water in the tub is too hot, and his confusion is alleviated by virtue of the explanation. If explanations are epistemic, then they are context relative. Johnny wants to know why the baby is crying, so Johnny’s mother must explain the baby’s tears using vocabulary that is capable of illuminating Johnny’s young mind, given Johnny’s limited knowledge base. If explanations are context relative, then the explananda are statements rather than causal relations between events. Johnny wants to know why the baby is crying, and Johnny’s mother’s explanation explains Johnny’s query about the baby’s tears. Not only are explanations primarily epistemic, but explanations are individuated intensionally. Johnny’s mother’s explanation of the baby’s tears is conceptually inequivalent to a scientist’s explanation about the molecular kinetic energy of H₂O

molecules in the tub. So, there are two explanations, even though they both refer to the same causal relation between events.

On the intensional model, the principle of strong explanatory exclusion should therefore state that there can be no more than a single complete explanation of a given *explanandum statement* (Campbell, 2008, p. 84ff; Davidson, 1993, p. 16; Fuhrmann, 2002, p. 195; Liu, 2001, p. 16; Marras, 1998, p. 449; McIntyre, 2002, p. 95). But, the mental explanandum statement is distinct from the physical explanandum statement, so there is no longer any exclusion pressure on mental explanations. For example, the explanandum statement using the mental vocabulary of Susan's act, because of its specific psychological context, requires a mental explanans in terms of reasons. The physical explanandum statement of Susan's bodily behavior, due to its physical context, requires a physical explanans in terms of muscle contractions and the like. There can be no more than this one complete physical explanans for this physical explanandum statement. However, the mental explanandum statement is intensionally distinct, so it requires a distinct mental explanans statement. Thus, the mental explanans is necessary to explain the mental explanandum statement, so the mental explanation is once again necessary.

Unfortunately, this dual explananda solution does not ultimately circumvent the problem of strong explanatory exclusion. To see why, consider two different versions of the intensional model. The first, most radical model is the the *fully* intensional model, according to which explanation is fully epistemic, not requiring a metaphysical component at all. It is uncommon to argue that explanation is *fully* intensional. Even Marras, for example, grants that explanation must have a metaphysical component (1998, p. 443). Failure to appreciate the metaphysical component of explanation results in what Kim calls *explanatory irrealism*, which Kim devotes a considerable portion of time discrediting. Indeed, a fully intensional model is as counterintuitive as a fully extensional model. Imagine, for example, that little Timmy explains the appearance of the gifts under the tree by appealing to Santa Claus. Timmy's explanation is epistemically illuminating to him and his friends, but it lacks an objective referent. On the intensional model, Timmy has a full explanation. But, this is counterintuitive. Timmy does not *really* have a full explanation. While Timmy has the epistemic component to explanation, he lacks the metaphysical component, which explains why he does not really have a full explanation.

The *partly* intensional model, according to which explanation is primarily epistemic but also has a metaphysical component, is more common. Unfortunately, the partly intensional model faces a familiar difficulty. According to the principle of physical explanatory completeness, every causal relation between physical events has a complete physical explanation—the causal relation between cause p and effect p^* is completely explained by “ P caused P^* .” If “ P caused P^* ” is a complete explanation, then additional explanations of the same causal relation between events are metaphysically unnecessary, and can be, in principle, excluded. Thus, the mental explanation “ M caused M^* ” is metaphysically unnecessary, and can be excluded. Or, here is another way of making the point: the mental explanandum statement “ M^* ” is unnecessary, since the physical explanandum statement “ P^* ” is metaphysically complete. So, the mental explanation “ M caused M^* ,” which explains the explanandum statement “ M ,” is itself unnecessary—if there is nothing needing explanation, an explanation is not needed—and can be excluded.

Two objections are worth considering. First, let us consider what Sophie Gibb calls the *many-to-one objection* (Gibb, 2009, pp. 213–215; compare Heil, 2003, p. 43ff). According to the many-to-one objection, multiple explanations of the same causal relation between events are possible. After all, the same thing is commonly described in various ways. The statement ‘the rose is red’, the statement ‘the rose is red-or-green’, and the statement ‘the rose is the color of the apple’ all refer to the same thing. Likewise, the same causal relation between events can be described in physical vocabulary or mental vocabulary:

If the causal explanantia are statements and the causal relata are events ... then there can be a many-to-one relationship between an explanation and a cause. Where $C1$, $C2$ and E are distinct statements, and c and e are distinct events, $C1$ may explain E in virtue of the fact that c bears causal relation R to e , but equally $C2$ may explain E in virtue of the very same causal relation. This is because $C1$ and $C2$ may both describe c To a single cause there can correspond a multiplicity of complete and non-equivalent causal explanations. [Explanatory Exclusion] is therefore false. (Gibb, 2009, p. 215)

Since the same thing can be variously described, the same causal relation between events can be variously described. Hence, the same causal relation between events can have a physical explanation and a mental explanation.

For the sake of the argument, I grant that Gibb is correct in suggesting that ‘the rose is red’ refers to the same thing as ‘the rose is red-or-green.’ So, of course, the same causal relation between events can have a physical explanation and a mental explanation. But the problem is not, after all, whether multiple explanations *can* refer to the same causal relation between events, but whether multiple explanations are *needed* for the same causal relation between events. Most specifically, given that the causal relation between events already has a complete physical explanation, is an additional mental explanation needed as well? According to the argumentation presented above, the answer is no. Additional explanations beyond the already complete physical explanation are not metaphysically necessary, and thus could be excluded. Here is another example: imagine that a cape’s redness (*c*) causes a bull to charge (*e*). According to Gibb, there can be two (or more) explanations referring to this same cause *c* of event *e*. This would include explanations like *C1*, ‘the cape’s redness causes the bull to charge’; *C2*, ‘the cape’s redness-or-greenness causes the bull to charge’; *C3*, ‘the instantiation of the causally relevant property in the object causes the bull to charge’; and *C4*, ‘the cape’s being Sandra’s favorite color causes the bull to charge’. Granting that these four explanations *can* refer to the same causal relation between events, the fact remains that, given that *C1* is a complete explanation, *C2*, *C3*, and *C4* are ultimately unnecessary. We do not *need* these additional explanations, even though we *can* have them and may even, *want* them.

Here is a second objection: while the mental explanation may not be metaphysically necessary, the mental explanation is epistemically necessary. And, on the intensional model, explanation is primarily epistemic, so the epistemic necessity of the mental explanation trumps the metaphysical unnecessary of the mental explanation. So mental explanations are, in the way that matters most, necessary. This view has some merit to it, but it faces difficulties as well. First, on the intensional model, mental explanations are only contingently epistemically necessary. That is, mental explanations are only epistemically necessary in those contexts in which humans lack a complete understanding of neuroscience. As neuroscience advances, humans might increasingly deploy neuroscientific vocabulary in their explanations of behavior, thereby shrinking or even eliminating epistemic needs for mental vocabulary. Secondly, while the intensional model secures the contingent epistemic necessity of mental explanations, it fails to secure the metaphysical necessity of mental explanation. A model that secures both the epistemic necessity and metaphysical necessity of mental explanations would be preferable. I propose such a model in section 7.

6. Conservative Reduction

At this point, ontological identity construed intensionally or extensionally fails to overcome strong explanatory exclusion pressures. Perhaps it is possible to circumvent strong explanatory exclusion pressures by positing an ontological identity without yoking it with either the intensional or the extensional model.

This can be done by emphasizing the *conservative* nature of conservative reductionism. That is, although the physical explanation of the causal relation between events is complete, an ontological reduction of the mental cause to the physical cause implies that the mental cause still exists. For this reason, the mental explanation of the effect is necessary as well and cannot be excluded. To see this move in action, consider first of all the views of some of the early eliminativists. W. V. O. Quine argues that mental states can be eliminated for Ockhamistic reasons, since the physiological states are already present: “the bodily states exist anyway, why add the others?” (Quine, 1960, p. 264; compare Rorty, 1970, pp. 420–421). Assuming that the physical explanation is complete and that behavior can be explained with recourse only to physical explanations, and thus without recourse to mental explanations, then mental explanations are not, in principle, necessary and can therefore be excluded.

James Cornman responds to both Quine and Rorty in the manner suggested above. Namely, since mental causes are physical causes, mental explanations need to remain because conservative reduction implies that there are mental causes, which then require mental explanation:

Even if we grant that a pain is identical with a stimulation of C-fibers, it would seem we shall still need sensation-terms to make the true descriptions of certain pains, or stimulation of C-fibers, as, for example, intense, sharp and throbbing. No neurophysiological sentence is synonymous with ‘This pain (stimulation of C-fibers) is intense, sharp, and throbbing’, and thus no neurophysiological sentence can be used to make the same true description. Thus to eliminate the sensation-terms we apply to what we experience would seem to diminish our ability to describe considerably. (Cornman, 1968, p. 30; compare Fuhrmann, 2002, pp. 195–196)

Cornman argues that a conservative ontological identity implies that pain exists, so explanations in terms of those pains are required. Since physical explanations are independent from mental explanations, it is not viable to suggest that the physical explanation offers these mental explanations. Therefore, the mental explanations continue to be metaphysically necessary.

Cornman’s suggestion leads to a simple question: is the physical explanation of the causal relation between events complete or not? If the physical explanation is not complete, then there is leftover work that the mental explanation needs to perform, and the mental explanation cannot be excluded. Notice, for example, that in the passage above Cornman suggests that physical explanations alone will not completely explain certain aspects of the c-fiber/pain event. He then poses a question: “Descriptions of what there is would be no less accurate if we dropped sensation-terms and observation terms. But would they be complete?” (Cornman, 1968, p. 29). If we are prepared to say that the physical explanation is incomplete, then the mental explanation will of course be metaphysically necessary.

Most people in the debate on mental explanation, however, accept the principle of physical explanatory completeness. If the principle of physical explanatory completeness is correct, then physical explanation is metaphysically complete. If physical explanation is complete, mental explanation is metaphysically unnecessary and can be excluded. Since there is a complete physical explanation of a given event, and since the event can have no more than one complete explanation, an additional mental explanation can still be excluded.

7. Non-Reductive Physicalism

The models presented above all attempt to avoid exclusion pressure in a similar manner. Namely, they all attempt to find methods of including mental explanations, mental vocabulary, or mental explananda, despite the fact that there is already a complete physical explanation, complete physical vocabulary or a complete physical explananda referring to the causal relation between events. As we also saw, the models presented above all fail to avoid strong explanatory exclusion pressure for a similar reason. Namely, since there is already a complete physical explanation, complete physical vocabulary, or a complete physical explananda, the mental explanation, vocabulary, or explananda is not metaphysically necessary, such that it can be, in principle, excluded. In this final section I argue that exclusion pressure can only be avoided if the mental explanation refers to a causal relation between events that is distinct from the causal relation between the physical cause and the effect that has a complete physical explanation. This is, of course, a common non-reductive physicalist position. So, I conclude that non-reductive physicalism has the significant advantage of being able to render mental explanations metaphysically necessary.

The principle of strong explanatory exclusion states that it is possible for there to be no more than a single metaphysically complete explanation for a given effect. This is possible because there is nothing left for a secondary explanation to explain. To avoid exclusion of the secondary explanation, something must remain that requires an explanation.

Certain versions of non-reductive physicalism, rather than reductive physicalism, satisfy this requirement. Consider reductive physicalism first. Reductive physicalism is the doctrine that there is only one causal relation between a physical cause and a given effect. Assuming that the principle

of physical explanatory completeness is true, there is a complete physical explanation of this causal relation between physical cause and the effect. So, since the physical explanation is metaphysically complete, and since there is nothing left requiring explanation, the mental explanation is not, in principle, metaphysically necessary. It is therefore possible to exclude it.

On the contrary, by some accounts, non-reductive physicalism is the doctrine that there is a causal relation between the physical cause and an effect, and a distinct causal relation between the mental cause and the same effect. Assuming that the principle of physical explanatory completeness is true, there is a complete physical explanation of the causal relation between the physical cause and the effect. This physical explanation is a metaphysically complete explanation of the causal relation between the physical cause and the effect. But there is something left requiring explanation, namely, the causal relation between the mental cause and the effect. Since there is something else requiring explanation, the mental explanation is necessary for this causal relation between events and cannot be excluded.

There is a significant objection to this proposal. To see this, it is worth pointing out that Fred Dretske has argued, much as I do, that since explanatory exclusion pressures only arise when a mental explanation refers to the same causal relation between events that already has a complete physical explanation, the solution is that mental explanations must refer to a different objective state than the physical explanation does. In this way, he argues, the physical explanation can be the complete explanation of the causal relation between physical events, while the mental explanation can remain necessary to explain a different objective state (Dretske, 1988, 1995; compare Baker, 1998; Pereboom, 2002).

While this move is successful in rendering the mental explanation necessary, Kim points out a problem with this model (1995, p. 134; compare Gibb, 2009, p. 220). In essence, if Dretske's solution is to double the number of objective states, then Kim's reply is to double the problem by imposing fresh exclusion pressures on both states individually (1995, p. 134). The original causal relation between physical cause and effect, everyone agrees, has a complete physical causal explanation. The mental cause, although it is a different state from the first, is still a physical state (otherwise physicalism is false), and so this second physical state will have a complete physical explanation as well. If this is granted, then the mental explanation is again a second explanation that stands in danger of exclusion.

Dretske concedes this point to Kim (1995, p. 142). I prefer to argue that a resolution to this objection depends upon a non-reductive resolution to the causal exclusion problem. That is, if a causal relation between the mental cause and the effect is established, then a mental explanation is necessary to explain this feature of our ontology. If, however, every event has a sufficient physical cause that excludes causal relations between mental causes and these effects, then the causal relation between the mental cause and the effect will not be included within our ontology, and the mental explanation will be unnecessary.

While a resolution to the causal exclusion problem lies beyond the scope of this paper, I think that a causal relation between the mental cause and the relevant effect is viable, which renders mental explanations necessary. This opinion is certainly shared by many non-reductive physicalists who posit distinct mental causes of effects. I only conclude here, however, that a resolution to the explanatory exclusion problem is only possible after a resolution to the causal exclusion problem. And the possibility that non-reductive physicalism secures necessary mental explanations is a significant advantage over reductive physicalism, which does not.

Notes

- 1 Here, and throughout this paper, the upper case *C* stands for the explanans proposition, while *E* stands for the explanandum proposition. By contrast, the lower case *c* stands for the causal event, while *e* stands for the effect.
- 2 Perhaps the most famous historical example is found at the beginning of the Third Book of Isaac Newton's *Principia*, where his first Rule of Reasoning states: "We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances" (Newton, 1846, p. 384).
- 3 Some even suggest that Ockham's principle is only aimed at epistemic entities, not metaphysical entities (Gauch, 2003, p. 272).

- 4 Indeed, a number of authors highlight the similarities that the argument for causal exclusion has with explanatory exclusion (Fuhrmann, 2002, p. 182; Gibb, 2009, p. 5). Andre Fuhrmann, for example, say that Kim thinks “there is essentially only one exclusion problem which may both be cast in terms of explanation as well as in terms of causation” (2002, p. 184). Kim even goes so far as to blend his two principles together under the heading of causal/explanatory exclusion at times (1989, p. 44; 1993, pp. 281, 291). Not surprisingly, as Gibb points out, numerous critics have assumed the two arguments are the same (Block, 2003; Pereboom, 2002). The closeness of the two doctrines is also borne out in the fact that the principle of explanatory realism says that explanations “track” (Kim, 1994, p. 68) or “mirror” (Kim, 1981, p. 307) objective causal relations. So deep is this assumption that explanations track and mirror causation that Kim sometimes fails to mark a distinction between the lower case *c* (the cause) and the upper case *C* (the explanans) (Kim, 1994, p. 58; 1989, pp. 94–95). Kim simply assumes that the explanation will so closely resemble the actual cause that he blurs the distinction at times.
- 5 More fully, all physical effects have complete physical explananda statements. If an effect is physical, it can, in theory, be given a complete physical description. Additionally, all physical explananda statements have complete physical explanantia statements. If there is a physical explanandum statement, it can, in theory, be explained by virtue of a complete physical explanans statement. Thus, all causal relations between physical causes and physical effects have complete physical explanations. The principle of physical explanatory completeness is motivated by considerations motivating the principle of physical causal completeness. Neither reductive nor non-reductive physicalists dispute this argumentation.
- 6 It is possible to argue that, in the event of a competition between a mental explanation and a physical explanation, the mental explanation excludes the physical explanation. This move would require a principle of mental explanatory completeness, according to which the mental explanation is a complete explanation of the causal relation between events. This move would also require motivation for why, in the case of competition, the mental explanation should prevail over the physical explanation of the same causal relation between events. Some attempt to secure the former principle, the principle of mental explanatory completeness, by appeal to the sufficiency of the mental cause for the effect. Physicalists, however, insist that effects have physical causes, which may preclude the sufficiency of the mental cause for the effect. As for the latter, it is common to assume that, in the event of competition, physical causes exclude mental causes, not vice versa. This is once again motivated by the physicalist view that effects must have physical causes. I would like to thank an anonymous referee for raising this concern.
- 7 I would like to thank an anonymous referee for pointing out this possible objection.

Acknowledgements

I would like to thank two anonymous referees for valuable comments and suggestions. I would also like to thank Neil Campbell for numerous valuable suggestions.

References

- Aristotle. (1941). *The basic works of Aristotle* (R. McKeon, Ed.). New York: Random House.
- Árnadóttir, S., & Crane, T. (2013). There is no exclusion problem. In S. C. Gibb & R. Ingthorsson (Eds.), *Mental causation and ontology* (pp. 248–265). Oxford: Oxford University Press.
- Baker, L. (1998). What we do. In J. Bransen & S. E. Cuypers (Eds.), *Human action, deliberation and causation* (pp. 249–270). Boston: Kluwer.
- Block, N. (2003). Do Causal Powers Drain Away? *Philosophy and Phenomenological Research*, 67, 133–150.
- Brand, M. (1976). Introduction: Defining causes. In M. Brand (Ed.), *The nature of causation* (pp. 1–44). Urbana, IL: University of Illinois Press.
- Campbell, N. (2008). *Mental causation*. New York: Peter Lang.
- Cornman, J. (1968). On the elimination of ‘sensations’ and sensations. *Review of Metaphysics*, 22, 15–35.
- Davidson, D. (1980). *Essays on actions and events*. Oxford: Clarendon Press.
- Davidson, D. (1993). Thinking causes. In J. Heil & A. Mele, *Mental causation* (pp. 3–18). Oxford: Clarendon.
- Dretske, F. (1988). *Explaining behavior*. Cambridge, MA: MIT Press.
- Dretske, F. (1995). Reply: Causal relevance and explanatory exclusion. In C. MacDonald & G. MacDonald (Eds.), *Philosophy of psychology: Debates on psychological explanation* (pp. 142–155). Oxford: Blackwood.
- Esfeld, M. (2007). Mental causation and the metaphysics of causation. *Erkenntnis*, 67(2), 207–220.
- Friedman, M. (1974). Explanation and scientific understanding. *Journal of Philosophy*, 71, 5–19.
- Fuhrmann, A. (2002). *Causal exclusion without explanatory exclusion*. *Manuscript*, 25, 177–198.
- Gauch, H. (2003). *Scientific method in practice*. Cambridge: Cambridge University Press.
- Gibb, S. (2009). Explanatory exclusion and causal exclusion. *Erkenntnis*, 71, 205–221.
- Harbecke, J. (2008). *Mental causation: Investigating the mind's powers in a natural world*. Frankfurt: Ontos Verlag.
- Heil, J. (2003). *From an ontological point of view*. Oxford: Clarendon.

- Heil, J. (2013). Mental Causation. In S. C. Gibb & R. Ingthorsson (Eds.), *Mental causation and ontology* (pp. 18–34). Oxford: Oxford University Press.
- Hill, C., & McLaughlin, B. (1999). There are fewer things in reality than are dreamt of in Chalmers's philosophy. *Philosophy and Phenomenological Research*, 59, 445–454.
- Horgan, T. (1991). Actions, reasons, and the explanatory role of content. In B. McLaughlin (Ed.), *Dretske and his critics* (pp. 73–101). Oxford: Blackwell.
- Jacquette, D. (1994). *Philosophy of mind*. Englewoods Cliffs, NJ: Prentice Hall.
- Kim, J. (1981). Causes as explanations: A critique. *Theory and Decision*, 13, 293–309.
- Kim, J. (1984). "Epiphenomenal and Supervenient Causation", *Midwest Studies in Philosophy*, 9, p. 257–270.
- Kim, J. (1988). Explanatory realism, causal realism, and explanatory exclusion. *Midwest Studies in Philosophy*, 12, 225–239.
- Kim, J. (1989). Mechanism, purpose, and explanatory exclusion. *Philosophical Perspectives*, 3, 77–108.
- Kim, J. (1993). *Supervenience and mind*. Cambridge: Cambridge University Press.
- Kim, J. (1994). Explanatory knowledge and metaphysical dependence. *Philosophical Issues*, 5, 51–69.
- Kim, J. (1995). Explanatory exclusion and the problem of mental causation. In C. MacDonald & G. MacDonald (Eds.), *Philosophy of psychology: Debates on psychological explanation* (pp. 121–141). Oxford: Blackwood.
- Kim, J. (1998). *Mind in a physical world*. Cambridge, MA: MIT Press.
- Kim, J. (2005). *Physicalism, or something near enough*. Princeton, NJ: Princeton University Press.
- Kim, J. (2009). Mental Causation. In B. McLaughlin, A. Beckermann, & S. Walter (Eds.), *Oxford handbook of philosophy of mind* (pp. 29–52). Oxford: Oxford University Press.
- Kirk, R. (2013). *The conceptual link from physical to mental*. Oxford: Oxford University Press.
- Levin, J. (2007). What is phenomenal concept? In T. Alter & S. Walter (Eds.), *Phenomenal concepts and phenomenal knowledge* (pp. 87–110). New York, NY: Oxford University Press.
- Liu, J. (2001). A nonreductionist's solution to Kim's explanatory exclusion problem. *Manuscrito*, 24, 7–47.
- Loar, B. (1990). Phenomenal states. *Philosophical Perspectives*, 4, 81–108.
- Loewer, B. (2002). Comments on Jaegwon Kim's Mind and the Physical World. *Philosophy and Phenomenological Research*, 65, 655–662.
- Malcolm, N. (1968). The conceivability of mechanism. *Philosophical Review*, 77, 45–72.
- Marras, A. (1998). Kim's principle of explanatory exclusion. *Australasian Journal of Philosophy*, 76, 439–451.
- Marras, A. (2007). "Kim's supervenience argument and nonreductive physicalism". *Erkenntnis*, 66, 305–327.
- McIntyre, L. (2002). *Supervenience and explanatory exclusion*. *Crítica*, 34, 87–101.
- McLaughlin, B. (2001). In defense of new wave materialism: A response to Horgan and Tienson. In C. Gillett & B. M. Loewer (Eds.), *Physicalism and its discontents* (pp. 319–330). Cambridge University Press.
- Melnyk, A. (1996). Testament of a recovering eliminativist. *Philosophy of Science*, 63, S185–S193.
- Melnyk, A. (2003). *Physicalist manifesto*. Cambridge: Cambridge University Press.
- Newton, I. (1846). *The mathematical principles of natural philosophy* (A. Motte, Trans.). New York: Daniel Adee. (Original work published 1687).
- Ockham, W. (1990). *Ockham: Philosophical writings* (P. Boehner, Trans.). Indianapolis, IN: Hackett.
- Papineau, D. (1995). Arguments for supervenience and physical realization. In E. E. Savellos & U. Yalcin (Eds.), *Supervenience: New essays* (pp. 226–243). Cambridge: Cambridge University Press.
- Papineau, D. (2001). The rise of physicalism. In C. Gillett & B. Loewer (Eds.), *Physicalism and its discontents* (pp. 3–36). Cambridge: Cambridge University Press.
- Papineau, D. (2002). *Thinking about consciousness*. Oxford: Oxford University Press.
- Peacocke, C. (1989). No resting place: A critical notice of *The View from Nowhere*. *Philosophical Review*, 98, 65–82.
- Pereboom, D. (2002). Robust nonreductive materialism. *Journal of Philosophy*, 99, 499–531.
- Poe, G., Walsh, C., & Bjorness, T. (2010). Cognitive neuroscience of sleep. *Progress in Brain Research*, 185, 1–19.
- Quine, W. (1960). *Word and object*. Cambridge, MA: MIT Press.
- Rorty, R. (1970). Incorrigeability as the mark of the mental. *Journal of Philosophy*, 67, 399–424.
- Rorty, R. (2009). *Philosophy and the mirror of nature: Thirtieth anniversary edition*. Oxford: Oxford University Press.
- Sabates, M. (1996). Kim on the metaphysics of explanation. *Manuscrito*, 19, 93–110.
- Schiffer, S. (1987). Remnants of meaning. Cambridge, MA: MIT Press.
- Shepard, R., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701–703.
- Sider, T. (2003). What's so bad about overdetermination. *Philosophy and Phenomenological Research*, 67, 719–726.
- Tye, M. (2000). *Consciousness, color, and content*. Cambridge, MA: MIT Press.
- Worley, S. (1993). Mental causation and explanatory exclusion. *Erkenntnis*, 39, 333–358.