Efficient Explanations: A Powers Hypothesis for How the Mental Causes

William Nolan

BPhil Dissertation

University of Oxford

14 June, 2023

Word count: 29,987

Abstract

This thesis attempts to offer a powers causation solution to the difficulties posed by mental causation. I do so by arguing against what I call “events-pair” analyses of causation. As is known, mental causation seems to raise the problems of overdetermination and the possible exclusion of the mental cause. Jaegwon Kim presents one possible answer by using his Supervenience Argument to urge reduction, whereas Thomas Kroedel presents another by defending a Lewisian counterfactual condition for causation. I argue that both these views use a common view of causation that depends on the assumption that causation is the relationship between a pair of temporally separated events. Against this assumption, I present and defend a powers view of causation, which can account for both the simultaneity and interruptibility of a causal process. I then apply the powers perspective to mental causation, with the beneficial results that overdetermination no longer poses a threat, double-prevention can be re-analyzed into constituent causal stages, and mental causes can be, without contradiction, one part of the sufficient cause of intentional bodily movement. Overall, I thus present a new powers solution to the difficulties of mental causation.

Contents

Acknowledgements

Thanks to an Ockham Society audience for helpful feedback on material in this thesis. Thanks to Mariana Canales for a well-timed lunch and to Allie Trice for well-timed study sessions. A group of philosophers at the Canterbury Institute merit gratitude in discussing many themes related to this project in a reading group (Henry Philips, Nathan Coundon, Ella Lubell, Grace Regnier, Francesco Banfi, Emerson Salovaara, Fr. Joe Murphy, Fr. Dancho Azagra). Thanks to both Ursula Coope and Daniel de Haan for pertinent and timely references to relevant literature over the past months. I am grateful for incisive feedback from fellow philosopher friends, including Nathan Elvidge, Kevin Zhang, Francesco Banfi, and especially Joe Cobon, as well as from my father, James Nolan. Finally, the greatest thanks are due to Timothy Williamson for his relentless, thorough, and attentive guidance and supervision.

For James Andrew Quinn

0. Introduction

The stakes of the mental causation debate are well known. The failure of mind to matter in the physical world is both philosophically catastrophic and compromises our most dearly held everyday conceptions about the nature of reality. As Jerry Fodor apocalyptically proclaims: "if it isn't literally true that my wanting is causally responsible for my reaching, and my itching is causally responsible for my scratching, and my believing is causally responsible for my saying… if none of that is literally true, then practically everything I believe about anything is false and it's the end of the world."[1]

These days, the difficulties of mental causation are not framed as the purported interaction between a Cartesian soul and a material body. Rather, now the difficulty is the seeming impossibility of causal interaction between mental and physical properties.[2] The problem is typically presented as arising from two separate arguments. The "interaction problem" worries about the causal nexus between physical and mental properties. The "exclusion problem," on the other hand, worries that a bodily physical effect might have a sufficient physical cause, in which case the mental cause would have nothing to contribute.

There are several common strategies to dissolve the problems of interaction and exclusion, and thus preserve mental causation. One is to embrace physical monism. Physicalists come in reductive and non-reductive varieties. The latter—like Jerry Fodor—claim that non-physical descriptions will not ever reduce to physical ones, whereas the former—like Jaegwon Kim—deny that claim. I present and challenge Kim's solution to mental causation in Chapter 1.

---

[1] Fodor (1989: 77).

[2] See Robb, Heil, and Gibb (2023), Kim (2005: Chapter 3), Kim (1998: 29-30), Steward (1997: 1-16), Sosa (1984: 278-280) for differences between current mental causation debates and early modern ones.

A second common strategy used to avoid the problems of mental causation is a Humean counterfactual account of causation. While it's possible to spell out counterfactual accounts of causation in various ways, all agree there are two causal relata, and that if element *A* had not been present, then element *B* would not be present. I present and challenge Thomas Kroedel's recent counterfactual account of mental causation in Chapter 2.

My strategy is to appeal to a neo-Aristotelian account of *efficient*—or *powers*—causation to reframe the issue of mental causation. The first two chapters are deflationary, whereas the second two are positive. In Chapter 1, I present Kim's "Master Argument" for excluding mental causes, arguing that it fails due to an underspecified notion of overdetermination and a fraught exposition of physical causal closure. In Chapter 2, I present Kroedel's counterfactual condition for causation, arguing that it founders on non-causal counterexamples. That is, his condition is satisfied by true but non-causal counterfactuals. With Kim, I endorse a "production" theory, arguing for a *powers* or *dispositional-property* solution to mental causation. With Kroedel, I argue for this conclusion while endorsing metaphysical agnosticism about mind. In Chapter 3, I motivate and defend the powers view of causation. I use the arguments of Nancy Cartwright and C.B. Martin to demonstrate the need for dispositional properties, before presenting a detailed account with help from Mumford and Anjum. Finally, in Chapter 4, I apply the novel powers view to the problem of mental causation.

1. Contemporary Accounts: Kim on Exclusion

"[T]he physicalist conception of the mind has become, in effect, the default position of contemporary analytic philosophers. According to that conception, states of mind are just a special subset of bodily states." – E.J. Lowe (2008: 12).

In this first chapter, I present the first of two different ways of responding to the problems of mental causation in the contemporary literature that I ultimately find unsatisfactory. For this first account and the one in next chapter, I choose a powerful expositor of the view to serve as my interlocutor. The two accounts share many common assumptions, perhaps chiefly that causation is understood as the succession of two nomologically linked events. It will be the contention of Chapters 3 and 4 that both these accounts fail, though in different ways, to give a satisfactory account of mental causation. The ambition of this chapter and the next is to show the difficulties facing these two views, laying the groundwork for a different solution to the problem of mental causation that presents a *tertia via*.

Jaegwon Kim is interested in exposing the untenability of any version of non-reductive physicalism (NRP). His arguments urge a skeptical stance about whether or not, given certain physicalist assumptions, the mental can be causally efficacious from a non-reductive physicalist perspective. Thus, he is one of the foremost representatives of *identity* responses to the problem of mental causation. If he successfully argues that a mental cause is just a physical cause, then the problems of both exclusion and interaction are eliminated. Kim's project is therefore deflationary, trying to use the untenability of NRP as a means to establish reduction, identity, and thus the dissolution of the problems of mental causation. The account attempts to show the incoherence of any variety of NRP about the mind.

1.1 Kim's Supervenience Argument

Kim's strategy is to argue against the tenability of NRP, which he believes to be a species of property dualism, by arguing that it disallows mental causation. In this section I lay out Kim's main argument for his view. In what he calls his "Supervenience Argument," Kim puts forward five principles that he thinks can be combined to produce the desired conclusion that NRP is untenable. (It is important to note that he believes the non-reductive physicalist endorses these principles, though Kim himself only accepts some of them). However, I argue that at least two of the principles do not hold up under scrutiny. I first present the principles, then his argument, and then I call into question its conclusion by attacking two of the principles, namely, *Exclusion* and *Closure*. In this way, even if one of my counterarguments fails, I can still fall back on the second one.

1.1.1 Kim's Principles

Besides his articles on the topic of mental causation, Kim's two main book-related treatments of the issue occur in *Mind in a Physical World* and his later *Physicalism, Or Something Near Enough.*[3] In what follows, I rely mostly on his supervenience argument in the second book, which has several crucial premises. A first and general basis for his formulation is an "exclusion principle." Such a principle is a necessary part of any version of the exclusion argument. His version runs as follows:

> *Exclusion.* No single event can have more than one sufficient cause occurring at any given time—unless it is a genuine case of causal overdetermination.[4]

[3] Kim (1998), (2005).
[4] Kim (2005: 42).

The idea behind an exclusion principle such as this one is that once we have landed on the genuine cause of an effect, we only create superfluity in our causal explanation if we add on a further cause. After laying out the succeeding principles, I will question, in particular, the tenability of Kim's *Exclusion* principle.

The next four principles are all ones he takes his opponent, the non-reductive physicalist, to endorse. Thus, the second principle of Kim's supervenience argument is his notion of supervenience itself. A concept that is employed in philosophical contexts as varied as philosophy of mind and metaethics, supervenience received increased attention in the second half of the twentieth century.[5] Kim articulates this concept as follows:

> *Supervenience.* Mental properties strongly supervene on physical/biological properties. That is, if any system *s* instantiates a mental property M at *t*, there necessarily exists a physical property P such that *s* instantiates P at *t*, and necessarily anything instantiating P at any time instantiates M at that time.[6]

The first sentence simply states that the mental supervenes on the physical, while the second spells out what Kim takes such supervenience to mean. He thinks that there is a nomological link between the mental and physical property such that whenever a certain mental property is instantiated, so too is a certain physical property, and vice versa. It's important for Kim that supervenience applies between mental and physical properties; he is more reserved, however, on the issue of whether or not the *causal powers* exposited by a special science supervene on microphysical causal powers.[7]

Kim advances a third principle, "irreducibility," which he take the non-reductive physicalist to endorse. He writes,

---

[5] See Kim (1984) for discussion of different versions of supervenience, with a brief history, 154-156. There he distinguishes between weak, strong, and global supervenience. Also, McLaughlin (1995), Sosa (1984: 275-277) for helpful discussions. See Davidson (1993: 187-188) for a brief history of the term.
[6] Kim (2005: 33).
[7] Ibid., 57-69.

> *Irreducibility*. Mental properties are not reducible to, and are not identical with, physical properties.[8]

This principle is obviously a basic commitment for anyone who wishes to preserve some autonomy for the mental. Given his reductive sympathies, it is a principle that Kim doesn't personally endorse. *Irreducibility* thus serves as a helpful reminder that Kim formulates his argument on principles he believes his opponent endorses, hoping to create a *reductio ad absurdum*, but not on principles to which he himself is necessarily committed.

Another principle which is also held by non-reductive physicalists and which Kim states before launching his supervenience argument has to do with causation. He writes,

> *Causal efficacy*. Mental properties have causal efficacy—that is, their instantiations can, and do, cause other properties, both mental and physical, to be instantiated.[9]

This principle simply states that mental causation occurs in the world. Similar to *Irreducibility*, Kim himself doesn't need to endorse every interpretation of this principle. For, one might think, with the non-reductivist or dualist, that mental properties have a different kind of causal efficacy than physical properties. But to reach his preferred conclusion, Kim is only minimally committed to defending the perspective that mental properties *can* cause other properties, even if they themselves turn out just to be physical properties. If his project succeeds, he will espouse an identity theory.[10]

Finally, Kim's last principle—that he sees himself as sharing with any other physicalist—is the causal closure of the physical. He writes,

> *Closure*. If a physical event has a cause that occurs at *t*, it has a physical cause that occurs at *t*.[11]

---

[8] Ibid., 34.
[9] Ibid., 35.
[10] Identity theories evade the exclusion problem because the effect has only one cause, which can be described mentally or physically. Kim (1989: 89).
[11] Kim (2005: 43).

Much ink has been spilled on the precise meaning of the causal closure of the physical and I will discuss it further later in the chapter.[12] But notice that Kim's statement of the principle is a relatively weak version of causal closure claims: he states only that a physical event has a physical cause. This principle, on its own, says nothing about the impossibility of mental causes or about the impossibility of physical and mental causes working in tandem. Of course, his ultimate view that mental causes are just physical ones permits him to endorse a specific sense in which mental properties cause and to reject the need for separate physical and mental causes. But the benefit of having a weak closure principle is that Kim can formulate his argument while demanding minimal commitments from his interlocutor.

1.1.2 Kim's Argument

His argument works as a kind of *reductio ad absurdum*. The conclusion of his argument is that mentality, as characterized by NRP, can do no causal work because it is merely epiphenomenal. Kim doesn't want to use his argument to deny mental causation *simpliciter*, but rather to demonstrate that mental properties must be physical properties to possess causal efficacy. Please note, at this early stage, that a property instantiation is for Kim considered an event.[13] This qualification is crucial for any view of causation on which the causal relata are taken to be events. The argument runs as follows:[14]

(1) The mental property M causes another mental property M* [an implication of *Causal efficacy*]

(2) A mental property M supervenes on physical property P [from *Supervenience*]

[12] For discussion of closure principles, see Baker (1993), Lowe (2008: 41-58).
[13] Kim (2005: 42).
[14] I follow the argument, with minor modifications, from Kim (2005: 39-52). Note, I use Kim's "Completion 1" version because it more explicitly integrates supervenience.

Kim thinks that there is a tension between these two initial premises. If we are trying to understand why M* comes about, Kim thinks that we have two possible answers. He writes, these possibilities are "(a) 'Because M caused M* to instantiate on this occasion,' and (b) 'Because P*, a supervenience base of M*, is instantiated on this occasion.'"[15] Option (a) is straightforwardly a rearticulation of (1), while (b) must be seen as a rearticulation of (2), though with the following substitutions: M* for M and P* for P. Kim notices that so long as P* is instantiated, M* will be too, as a result of *Supervenience*, even if M doesn't occur. This observation might seem to make M unnecessary, so long as P* is present. So Kim proposes a way to dissipate this worry about having two different explanations of M*: he posits a connection between them. He writes:

> (3) M caused M* by causing its supervenience base P* [from *Supervenience* and the above considerations]

Kim takes himself to have established that mental-to-mental causation implies, because of supervenience, that it is really a physical property that is caused, since without P*, M* remains uninstantiated.[16] This concludes what he calls Stage 1 of his supervenience argument. Stage 2 consolidates this beginning:

> (4) M has a supervenience base, P [from *Supervenience*]
>
> (5) M causes P* and P causes P* [from (3) and (4)]
>
> (6) M ≠ P [*Irreducibility*]
>
> (7) Physical property P* is not causally overdetermined by M and P [from *Exclusion*]
>
> (8) The mental property M is excluded by the physical property P. P, not M, causes P* [from *Closure* and *Exclusion*]

[15] Ibid., 39.
[16] Kim states a general principle concerning the outcome: "*level-bound causal autonomy is inconsistent with supervenience or dependence between the levels*" ibid., 40.

The contention behind (5) is essentially that taking two ideas into consideration we can conclude that P causes P*. The first idea is that there is a nomological link of supervenience between M and P and the second idea is that M is, *ex hypothesi*, a cause of P*. If M is a genuine cause and is always accompanied by P, then P also would seem to be a cause. Though there is much to be said for this view, it is unclear that (5) follows from what has been established. Unless I misunderstand Kim, it would seem easy to simply counter that P could be "epi-physical"—that is, causally irrelevant—to M's causal work.

For example, consider the case in which extreme anger causes me to punch a hole in the wall:

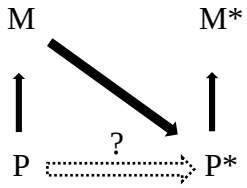

Figure 1.1: Kim's Argument

If my anger (supposing it to be a mental property instantiation M) causes me to punch a hole in the wall (supposing my arm movement to be a physical property instantation P*), one might consider P (the original physical property) to be, for instance, a raised heartrate and any number of other physical realizers. In the diagram, M represents my anger, P my anger-realizing physical properties, P* the muscle contractions of punching, while M* represents a contented happiness at the satisfaction of my anger. Now Kim says that M causes P*, represented by the solid arrow, though we don't yet know the causal relationship between P and P*, as represented by the empty dotted arrow. The smaller vertical arrows represent supervenience relationships between mental and physical property instantiations. What do the physical realizers of P have to do with my punching, P*? It is unclear, on the limited premises given, that P also needs to be a cause of P*.

Perhaps Kim could get (5) from reasserting the principle of *Closure*, but he doesn't in fact appeal to causal closure at this stage.

Granting the obvious (6), it next remains to examine the controversial (7). Recall Kim's statement of *Exclusion*:

> *Exclusion.* No single event can have more than one sufficient cause occurring at any given time—unless it is a genuine case of causal overdetermination.

I will say more about overdetermination shortly, but for the moment assume that the case at hand *is not* one of overdetermination. That leaves us in the territory of the first part of the *Exclusion* principle, preceding the dash. That yields (7), namely, that P* will not be overdetermined by the conjunction of M and P. Notice, though, that Kim can only get the first half of his Exclusion principle to apply by blocking the possibility of the second half—something he does.[17] Whether this move is legitimate remains to be seen.

The conclusion of Kim's argument as written in (8) is that a mental property instantiation, as understood by the non-reductive physicalist or any property dualist, cannot be causally efficacious. He wishes to reach this conclusion not because he wants to rule out mental causation, but because he thinks the version of mental causation defended by NRP is untenable. He thinks it fails to grasp what the conjunction of the above principles entails.

One might think that Kim's argument is unfairly prejudiced in starting with the arguably more difficult case of mental-mental causation rather than mental-physical causation. But one benefit of Kim's approach is that it actually treats both types of causation within a single argument. In order to see his treatment of the mental-physical variety of mental causation, simply take his argument beginning with (3). There he explains that M (a mental property instantiation) causes P* (a physical property instantiation). Of course, his argument about this kind of

[17] "Let us assume that this is not a case of causal overdetermination," ibid., 42.

causation proceeds down through (8) to reach the same conclusion as that starting with (1): even in cases of mental-physical causation, the mental cause is inefficacious (or, at least reducible to the physical cause).

One way of responding to Kim's argument would be to put pressure on one or more of the premises. In this way, an interlocutor might hope to preserve NRP against what some have called "Kim's Master Argument."[18] However, since my aim here is not to preserve NRP against its enemies, I opt for a second strategy. This strategy is to put pressure on some of the five underlying principles that justify the premises. If any of these principles is shown to be untenable, then the argument collapses by relying on untrustworthy foundations.

1.2 Questioning *Exclusion* and *Closure*

In this section, I will put pressure on two of Kim's principles from the above section. In this way, I hope not only to dismantle his Master Argument, but set the stage for a different understanding of mental causation in subsequent chapters. For the sake of my argument, I set aside three of Kim's five principles. I target *Exclusion* and *Closure*, while passing over *Supervenience*, *Irreducibility*, and *Causal Efficacy*, except where they touch on the first two. My reasons for setting aside these three latter principles are distinct, but in general I do so because they are more philosophically acceptable than *Exclusion* and *Closure*. *Supervenience* is commodious to any who wish for a nomological link between properties of the mind and properties of the brain. *Irreducibility* is commodious to all except the full-fledged reductivist (or eliminativist). Finally, *Causal Efficacy* is the least controversial of all: it simply states that mental properties *have causal influence in the world.* I first attack *Exclusion* and then *Closure*.

[18] Crisp and Warfield (2001).

1.2.1 *Exclusion*

Recall Kim's statement of the *Exclusion*: "No single event can have more than one sufficient cause occurring at any given time—unless it is a genuine case of causal overdetermination." Kim's *Exclusion* principle, or other versions of it, feature prominently in formulations of the Exclusion Argument.[19] The basic intuition of this principle and the argument in which it features is that an effect is fully explained once a sufficient cause is identified. Since the effect is fully explained, a further cause is only superfluous and unnecessary. I argue that as stated, the *Exclusion* principle cannot do the work Kim wants it to accomplish.

At first pass, notice that within the stated conditions of the principle one might also question whether it is impossible for two sufficient causes to operate in tandem.[20] Consider cases of so-called late preemption.[21] In this sort of case, two sufficient causes are involved, even though only one of them does all the causal work. If, for example, both Billy and Suzy throw stones at a glass bottle simultaneously, but Billy's stone arrives first, his stone obviously does the causal work of shattering the bottle. Nevertheless, Suzy's throw ensures that had Billy failed to throw his stone, the bottle still would have shattered. To meet Kim's temporal qualification, it might be added that there are two different possible worlds for what happens at time *t*. In one world, Billy gets distracted and Suzy's stone arrives first. In a second world, Suzy gets distracted and Billy's stone arrives first. There are thus two possible sufficient causes of the single effect.

---

[19] See Aimar (2011), Kroedel (2019: Chapter 4).

[20] See Crane and Arnadottir (2013). Though formulated less technically than I attempt here and eliding the distinction between the two parts of Kim's exclusion principle, these authors also reject the principle: "[F]ar from being an analytic truth, the exclusion principle is not even plausible on its face," 257. They conclude thus because while *independent* causes might be an acceptable candidate for exclusion, mere *distinction* between causes is not (and Kim doesn't defend his principle). Relatedly, see my distinction between *Prototype Overdetermination* and *Mental Overdetermination* below and my discussion in Chapters 3 and 4 of *composite* sufficient causes.

[21] My example is from Kroedel (2019: 30-31; 49). On preemption, see Field (2003), Hall (2004), Bernstein (2019).

But one could easily respond by saying that in the actual world, only one sufficient cause is actually operative. Sufficiency needs to be more clearly spelled out to be useful. For if something happens (e.g. the glass bottle is shattered), one can always trivially point out that there is a sufficient cause of the effect. In what follows, I hope to get more precise about the difference between the kind of overdetermination that occurs in cases like late-preemption in contrast to possible overdetermination in mental causation.

Returning to Kim's statement of *Exclusion*, notice, that it really contains two possibilities, *Exclusion'* and *Overdetermination*:

> *Exclusion'*: No single event can have more than one sufficient cause occurring at any given time.
>
> *Overdetermination*: A single event *can* have more than one sufficient cause occurring at any given time when there is a case of genuine causal overdetermination.

Since Kim wants to reach a reductivist conclusion by collapsing the mental cause into the physical cause, his argument only works if he endorses *Exclusion'*, something we saw he indeed does. Does he argue for preferring *Exclusion'* over *Overdetermination*? Since endorsing *Overdetermination* (permitting two sufficient causes) is quite different from endorsing *Exclusion'* (permitting only one), Kim needs to show that *Overdetermination* doesn't feature in mental causation.[22] As first articulated, *Exclusion* on its own cannot get Kim his desired conclusion.

Kim's strategy of responding is expressed mostly as a reply to the critiques of an earlier version of his argument.[23] First, he notes that M (e.g., pain) might be present in a world where P (a specific set of nociceptive neurons) is not present. Nevertheless, P* is caused. Kim's response is to assert that even if P isn't instantiated, a different property P' (a different set of nociceptive

[22] He attempts to respond to this critique (2005: 46-52).
[23] These critiques are from Block (2003), Crisp and Warfield (2001), themselves responding to Kim (1998).

neurons) must accompany M, given *Supervenience*. Given the acceptance of supervenience by all relevant parties—non-reductivists and Kim—this reply works.

Nevertheless, Kim thinks he can rule out the second part of *Exclusion* as articulated in *Overdetermination* and maintain *Exclusion'* because supervenience ensures that mental causation differs sufficiently from standard cases of overdetermination. Consider these two possible ways of unpacking *Overdetermination*:

> *Prototype Overdetermination*: A single event *can* have more than one sufficient—but distinct and independent—causes occurring at any given time when there is a case of genuine causal overdetermination.
>
> *Mental Overdetermination*: A single event *can* have more than one sufficient—but indistinct and dependent—causes occurring at any given time when there is a case of genuine causal overdetermination.

I take the terms "distinct" and "independent" from Kim's text, though he does not cash out the precise alternatives I have presented.[24] His contention seems to be that only *Prototype Overdetermination* qualifies as a full-fledged case of *Overdetermination*. Perhaps his idea is that for an effect to be overdetermined, it needs to involve the coincidence of truly independent causal chains that happen randomly to coincide to produce an effect.[25] Obviously, given supervenience, this type of situation couldn't be true for *Mental Overdetermination*.

To make his response work, Kim would need to re-express his original *Exclusion* principle in the following way: "No single event can have more than one sufficient cause occurring at any given time—unless it is a genuine case of *Prototype Overdetermination*." In the new definition *Overdetermination* has been replaced with *Prototype Overdetermination*. It might be the case that an argument could be made for excluding *Mental Overdetermination* from his articulation of *Exclusion*, but Kim has not provided one yet. For *Mental Overdetermination* is a

[24] Kroedel too has helpful comments on how mental causes are nomologically linked but prototype cases of plural causation aren't (2019: 167-171, 173). Also, Menzies (1995), LePore and Loewer (1987: 635, n. 13).
[25] See Lewis (1986).

genuine third type of case: it is neither a case involving a single cause nor a case involving plural causation by independent and distinct causes. Thus, we do not yet have a basis for interpreting *Exclusion* as Kim does, discounting both possible kinds of *Overdetermination*.

Kim offers another, though, even by his own lights, admittedly less compelling response to the possibility of overdetermination. It will be useful to glance at this strategy briefly, even though ultimately it does not offer a way out for Kim. He says that one might make his argument work by embracing a new version of *Closure*, excising dependence on *Exclusion*. It goes as follows:

> *Strong closure.* Any cause of a physical event is itself a physical event—that is, no nonphysical event can be a cause of a physical event.[26]

Employing this principle, Kim can avoid entirely appeal to the controversial *Exclusion*. Now, he can simply take *Irreducibility* and *Strong closure* together to entail "M does not cause P*."[27]

The trouble with *Strong closure* is that it requires too much from the non-reductive physicalist interlocutor at the outset. It essentially asserts the conclusion of his argument. It would thus be rejected by most of his opponents and fail to get the argument off the ground.

The examination of *Exclusion* so far has yielded this result. First, in order to make his argument work, Kim needs to say that the second clause of *Exclusion*—countenancing the possibility of overdetermination—doesn't apply in the case of mental causation. However, simply stating the way in which mental causation differs from other types of overdetermination (as I expressed in *Prototype Overdetermination* and *Mental Overdetermination*) doesn't seem to get Kim the elimination of overdetermination from *Exclusion* that his argument requires. Thus, *Exclusion* cannot function as he wishes. Second, his suggested appeal to *Strong closure* is also not a viable strategy since it assumes too much from his non-reductive physicalist interlocutor.

---

[26] Kim (2005: 50).
[27] Ibid., 50-52.

1.2.2 *Closure*

If the above counterarguments hold, then Kim's argument fails as it stands because *Exclusion* fails. However, even if he could offer a successful response to my attacks on *Exclusion*, I also wish to argue against his articulation of *Closure*. This way, if a reader finds one of the challenges less plausible, there is still the second attack on which to fall back. I will argue against *Closure* on three different counts. First, I note a *prima facie* reason that *Closure*, as it stands, at least needs further qualification. Second, I outline some considerations proposed by E.J. Lowe that put pressure on the tenability of causal closure. Third and finally, I argue that the principle needs further justification, particularly perhaps from a separate concept like supervenience.

As a preliminary, it should be noted that within the context of this principle alone, one can easily imagine all sorts of mental events or property instantiations causing other mental events or property instantiations, that would not require physical causes. Thus, cases of mental-to-mental causation are wholly unaffected by *Closure.* It therefore should be noted that it is only by the addition of principles like *Supervenience* and *Exclusion*, that Kim's Master Argument can get off the ground in its attempt to reduce mental causes to physical entities.

Returning to the principle itself, it is far from clear that the causal closure of the physical is an empirically grounded principle.[28] Kim states *Closure* in the following way: "If a physical event has a cause that occurs at *t*, it has a physical cause that occurs at *t*." What, precisely, is the "physical" in this context? It could range from that part of reality captured by pure physics alone,

[28] See Papineau (2001) for a history of the claim. See Gibb (2015) for an argument against the exclusion problem by questioning causal closure.

to that part studied by everything from physics, to biology, and neurophysiology.[29] Even once non-physical causes have been ruled out as relevant to the framing of causal closure, further specification is needed to isolate the sense of physical causes required.

Closely related to this issue is the concern that the sciences don't themselves require causality at all, in the way that philosophers often naïvely assume. While history has not been kind to Bertrand Russell's radical suggestion that we should eliminate the concept of causality, there are still ways in which causality within the physical world is difficult to pin down.[30] Hartry Field has pointed out that many features of causation (its directionality, asymmetry, temporal ordering, even its relevance to statistical probabilities in Nancy Cartwright's work) cannot be simply dispensed with even in our best physical theories.[31] These considerations do not point decisively against the veridicality of *Closure*, they simply call for clarification of what is the precise meaning of "physical causes" within this definition.

E. J. Lowe has a somewhat controversial proposal about causal closure principles.[32] He suggests that though causal closure holds, this ought not worry us, for mental causes operate as causes of *facts* as opposed to causes of *events*. First, he notes that there are a surprising variety of causal closure principles on the market, which might lead one to wonder whether these principles coalesce around a focal meaning. In the current context, however, we can focus on Lowe's treatment of causal closure as he extracts the principle from Kim.[33] This formulation runs as follows: "Every physical event contains only other physical events in its transitive causal

---

[29] Menzies (1995: 197): "in the context of the problem of mental causation, the term 'physical' is usually understood more loosely to apply to the particulars and properties of neurophysiology, presumably on the grounds that such entities supervene straightforwardly on the physical." Menzies doesn't necessarily endorse this view, he simply offers it as a likely meaning of 'physical' in this context.

[30] Russell (1913: 1): "The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm."

[31] Field (2003).

[32] Lowe (2008: Chapters 2 and 3).

[33] In fact, Kim (2005: 43, n. 10) formulates his principle in response to transitivity concerns raised by Lowe (2003).

closure."[34] Lowe explains that by 'transitive causal closure' he means "the set of events consisting of the immediate causes of *P*, the immediate causes of those causes, the immediate causes of *those* causes…"[35] First, note that this principle of causal closure is stronger than that advanced in the Master Argument, as stated above. In that case, we saw that mental causes could enter into the causal history, but only that there had to be a physical cause *in addition*. Even under the auspices of this distinct and stronger principle, however, Lowe still believes there can be an intervening mental cause that doesn't violate closure. If it is possible, then closure is not so much violated as made less relevant. For if mental causes can intervene without violating *Closure*, then the force of the principle is much diminished.

He explains how his counterexample works, "Now, if a mental event *M* causes it to be the case that certain physical events, $P_1, P_2, \ldots P_n$, have a certain physical effect, *P*, then, it seems clear, *M* is itself a cause of *P*—but not an immediate cause of *P*, nor an immediate cause of any of the physical events $P_1, P_2, \ldots P_n$."[36] Lowe here relies on what he calls the difference between *event* and *fact* causation. He thinks that operating as causes of *facts* or *states of affairs*, mental causes bring about the broader context in which physical event causes occur. He concludes that "it seems clear that *M* is not included in *P*'s transitive causal closure and hence that *M* can be a non-physical cause without violating [the closure] principle."[37] In this way, Lowe sees the mental cause as bringing about a state of affairs without being an event cause. If we take Lowe's approach, then *Closure* states an unhelpful truth, since it pertains to only one kind of causation.

Such a kind of response to causal closure is controversial, however. For it demands making strong claims about a new kind of causation. Before turning to a final difficulty with *Closure*, one further objection that Lowe raises should be mentioned. He argues that the causal

[34] Lowe (2008: 53). Lowe gets this principle from Kim (1998).
[35] Ibid., 54.
[36] Ibid., 55.
[37] Ibid.

closure principle is not *a priori*. In fact, it is a principle that must be empirically verified. He claims that based on our current scientific understanding of quantum mechanics, there is no good reason for thinking that such a principle holds.[38] Such a claim is similar to the strategy invoked by Ackerill and Keating, but I avoid wading into the murky waters of these empirical debates.[39] Lowe admits that in light of quantum mechanics, causal closure could be reformulated to operate by "fixing probabilities." Nevertheless, it is false in its standard formulations.[40] This contention is speculative and depends on how one interprets quantum mechanics.

A final and more pressing reason to question *Closure* is that it needs external support to be convincing. Recall the principle one more time: "If a physical event has a cause that occurs at *t*, it has a physical cause that occurs at *t*." Notice that in the protasis of the conditional, the appearance of "cause" is unmodified by "physical." This is of course essential, otherwise no new information would be included in the apodosis. Yet, *on its own*, the principle doesn't give reason for thinking that the apodosis follows from the protasis. Say that there is a mental cause of the physical event at time *t*, it does not immediately follow that there is also a physical cause at *t.* The assertion in the apodosis that there is a physical cause could only be necessitated by the conjunction of the apodosis of *Closure* with another principle, like *Supervenience*. In this way, a mental cause would always bring along a physical one because of nomological dependency. This consideration does not show *Closure* to be false; rather, it shows that in its current formulation, it is only a compelling conditional if paired with further principle(s). It is surprising, therefore, that Kim doesn't give motivating reasons to accept this principle in the course of his argument.[41]

---

[38] Ibid., 65-68.
[39] Ackerill and Keating (1981).
[40] Lowe (2008: 67).
[41] Kim (2005: 43). Before stating the principle, he writes only, "for a general theoretical justification [of the last stage of the argument] we may appeal to the causal closure of the physical domain."

In summation, the conclusion of Kim's supervenience argument need not be accepted. I have argued that two of his underlying principles are particularly questionable. First and foremost, *Exclusion*, as formulated, doesn't seem capable of carrying the weight that Kim places upon it. He fails to establish either that mental causation is the bad kind of overdetermination or that if a mental and a physical cause are paired, one needs to be eliminated. As a result, his overall argument doesn't proceed. If someone could offer counterarguments to the concerns raised about *Exclusion*, I have also offered reasons for doubting *Closure*, as it stands. The principle is ambiguous about what exactly qualifies as physical causation. It could admit mental causes as causes of *states of affairs*, it might fail to account for quantum indeterminacy, and its conditional phrasing only makes sense when conjoined to other principles. Thus, like *Exclusion*, if *Closure* fails for the reasons given, then Kim's overall argument also cannot stand. The failure of his argument doesn't mean the failure of mental causation, but it does mean that we need a different account than the reductivist physicalist one Kim seeks.

2. Contemporary Accounts: Kroedel on Counterfactual Causation

"The mind makes a difference to the physical world. Therefore, the mind causes things to happen in the physical world." – Thomas Kroedel (2019: ix).

Humean *counterfactual* accounts of causation have experienced a great deal of popularity over the last five decades, thanks particularly to the work of David Lewis.[42] Hume infamously penned the following passage: "we may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or, in

[42] E.g. Lewis (1973a), (1979), (1986), (2000), Loewer (2002). See Stalnaker (1968) for an influential non-causal account of conditionals. See Bernstein (2019) for an overview Lewis's influence on contemporary philosophy of causation.

other words, where, if the first object had not been, the second never had existed."[43] Counterfactual theories adhere to the second of the two sentences in this passage. Starting with an influential 1973 article, Lewis spent three decades advancing and refining his Humean account of causation. Thomas Kroedel, as one of the latest in the line of those defending a counterfactual account of causation, applies Lewis's analysis to mental causation. He shares the Kimian aspiration to defend the reality of mental causation. Unlike Kim, however, he avoids embracing a specific metaphysical stance, instead trying to offer a flexible picture which can accommodate metaphysical stances ranging from NRP to what he calls "super-nomological dualism."[44] In this conciliatory spirit he offers a Humean counterfactual account of mental causation.

Kroedel's strategy is on display most fully in his recent book *Mental Causation: A Counterfactual Theory*.[45] In presenting his view, I focus on the position put forward in his brief introduction and the first two chapters. His first chapter focuses on theories of mind and counterfactuals, the second on how these each operate in the mental causation debate. Since chapters three and four deal with more fine-grained details and applications of his view, I reference them only when further precision is desirable.[46] My approach proceeds in three stages. First, I examine the theories of mind that Kroedel's account accommodates. Next, I examine his

---

[43] Cited from Lewis (1973a: 556) who references Hume, *An Enquiry concerning Human Understanding*, Section VII. Of course, counterfactual theorists might quibble about details: whether counterfactual dependence is between objects, what an object is, or whether *transitivity* holds in causal chains.

[44] See Kroedel (2019: 4): "Thus, the strategy pursued in this book uses weak assumptions both about the metaphysics of causation and about the metaphysics of mind." Super-nomological dualism is based on Kroedel's understanding of nomological supervenience, a modal articulation of supervenience, applied to the standard dualist view that mental and physical properties are distinct, ibid., 17-18. It's interesting to note that Kim (2005), without direct engagement, makes passing reference to counterfactual theories, saying that he presents a view of "productive causation against counterfactual theories (2005: 38 n. 6, 47 n. 12). Also, Kim (2002), Loewer (2002). My Chapter 3 discusses further.

[45] Kroedel (2019).

[46] These chapters explore Causal Modelling and the Exclusion Problem. Note, that in my analysis, I start with his positive account and turn in later sections to his own objections and responses (2019: 49-59, 68-75, 92-97), when voicing my own counterarguments.

application of counterfactuals to mental causation. Finally, in the last section of this chapter, I raise two key objections to his approach, based on worries about *temporal ordering* and *overlapping realizers*.

2.1 Theories of Mind

Before proceeding to Kroedel's careful presentation of his counterfactual approach to causation, I briefly mention the span of metaphysical views of mind he attempts to accommodate. Kroedel gives a fairly standard definition of non-reductive physicalism (NRP). Where he is perhaps more innovative is his particular definition of dualism. If Kroedel can get his version of dualism off the ground, he hopes to explain how mental causation can work for both non-reductive physicalists and for dualists. I outline how he defines NRP and then how he defines dualism.

On the one hand, Kroedel's definition of NRP depends on his rendering of strong supervenience, which he formulates as follows:

> **Strong supervenience**: A set of properties **A** *strongly supervenes* on a set of properties **B** if and only if, necessarily, if anything instantiates some property *F* in **A** at a given time, then there is a property *G* in **B** such that that thing instantiates *G* at that time, and, necessarily, everything that instantiates *G* at a given time also instantiates *F* at that time.[47]

The key differences between this articulation of supervenience and that of Kim discussed above are twofold. First, Kim's definition deals only with the special case of *mental* and *physical* properties, whereas Kroedel's definition is framed as a general relation between any two sets of properties. This definition can of course be made more specific by adding the assumption (employed by Kroedel) that we take the set of properties **A** to be mental ones and the set of

[47] Kroedel (2019: 15). Also, Kim (1984), McLaughlin (1995) for further discussion of "Modal-Operator Strong Supervenience."

properties **B** to be physical ones. Second, and more subtly, because Kroedel employs the notion of a "set of properties," he allows for a weaker connection between the super- and sub-vening properties. Properties need not be correlated one to one, but rather certain sets of (mental) properties will always supervene on certain sets of (physical) properties. This approach has relevance to the now famous "multiple-realizability" concerns. Multiple-realizability is simply the claim that a given mental property, event, or state—say pain—will have a different physical realizer in its different instantiations, across species, individuals, or even across instances within the same individual.[48] Kroedel's strategy allows him to avoid this challenge, because he leaves it open precisely *which* property *F* in **A** will subvene the property *G* in **B**.

Applying these insights to his definition of NRP, Kroedel takes this view to consist in the conjunction of (1) the claim that the mental and the physical are distinct with (2) strong supervenience. He thus writes:

> **Non-reductive physicalism**: Each mental property is distinct from all physical properties, but mental properties strongly supervene on physical properties.[49]

Note one brief upshot of this definition: momentarily setting aside dualism, throughout the book Kroedel focuses on the NRP variety of physicalism, assuming that if his account works for the non-reductivist it also succeeds *a fortiori* for the reductivist. The reason for this assumption is that the reductivist endorses stronger identity claims. A reductivist would endorse the second clause of the above definition, but not the first, accepting only "mental properties strongly supervene on physical properties" along with an identity claim. Because reductivists endorse the identity of mental and physical properties, the Exclusion Problem does not arise for them. Thus, if mental causation holds for the more difficult case of NRP, it also holds for the easier case of reductive physicalism. Further, since neither the Exclusion Problem nor the Interaction Problem

[48] Putnam (1967); Kim (1992); Shoemaker (2007: 1, 6, 11); Bickle (2020); Robb, Heil, and Gibb (2023: 6.1).
[49] Kroedel (2019: 16).

arise for the reductive physicalist, it is obviously the right dialectical move for Kroedel to focus on the trickier obstacles NRP introduces to mental causation.

Conveniently, Kroedel can also use his notion of Strong Supervenience to help capture his preferred version of dualism. To avoid an overly spooky form of dualism that would be uncongenial to the scientifically minded, Kroedel wishes for a weak form of dualism. He finds what he is looking by defining a new kind of supervenience:

> **Nomological supervenience**: A set of properties **A** *nomologically supervenes* on a set of properties **B** if and only if it is nomologically necessary that if anything instantiates some property *F* in **A** at a given time, then there is a property *G* in **B** such that that thing instantiates *G* at that time, and it is nomologically necessary that everything that instantiates *G* at a given time also instantiates *F* at that time.[50]

Notice that this definition is nearly identical to that of Strong Supervenience. The only difference is that "necessarily" is now replaced with "nomologically necessary." The latter term signifies, for Kroedel, "necessity in view of the laws of nature."[51] A benefit of the new definition is that it endorses a weaker modality than Strong Supervenience.

Having delineated both Strong and Nomological Supervenience, Kroedel believes he can offer a preliminary definition of his preferred brand of dualism. He writes:

> **Naturalistic dualism**: Each mental property is distinct from all physical properties. No subset of mental properties strongly supervenes on physical properties, but mental properties nomologically supervene on physical properties.[52]

Kroedel thus expresses his version of dualism based on his two versions of supervenience, paired with the claim that mental and physical properties are distinct. Henceforward, however, I will call his preferred version of dualism "super-nomological dualism," which includes the additional criterion that nomologicality in the definition means that the relationship of mental and physical is stricter than the nomologicality of the ordinary laws of nature.[53] Since his reasons for adopting

[50] Ibid., 17.
[51] Ibid., 18.
[52] Ibid. David Chalmers coined "naturalistic dualism."
[53] Ibid., 88. He adds this stipulation to follow Lewis's similarity of worlds theory.

the new term are irrelevant to my own concerns, we can say without loss that his version of dualism is the conjunction of three theses:

1. nomological supervenience is true
2. strong supervenience is false
3. mental and physical properties are distinct

Kroedel recognizes that there could be hybrid views between physicalism and dualism, as he has defined them, but doesn't think that specifying them matters for his argument. Further, he recognizes that some might wish to apply his labels to slightly different views. While admitting these possibilities, he proposes that the views he articulates are sufficiently representative of mainstream views in the philosophy of mind.

At first glance, Kroedel's strategy might appear attractively generalized, accomodating various theories within its expansive fold. There is a possible worry, however, that though the spectrum of his views seems wide, he has in fact decided fundamental questions by prematurely narrowing the range of acceptable views. Crucially, the Cartesian or non-Cartesian dualist[54] might not recognize his own view in Super-nomological Dualism. Surely there are dualists who endorse a form of supervenience weaker than Super-nomological Dualism. For example, consider the view of someone who believes that if anything instantiates some property *F* in **A** at a given time, then there is a property *G* in **B** such that that thing instantiates *G* at that time, but only *probabilistically*. Thus, the supervenience no longer holds with nomological necessity. Thus, one might worry that Super-nomological Dualism and NRP are too similar.

2.2 A Counterfactual Condition for Causation

[54] Lowe (2008) is a non-Cartesian dualist.

Kroedel's basic strategy takes a Lewisian account of counterfactuals and applies it to causation (like his predecessor) and, more specifically, to mental causation (unlike his predecessor). At the highest level of generality, Kroedel states that, "Your mind had a physical effect because it made a difference to what was going on in the physical world."[55] He holds onto this basic commitment, developing a sophisticated account for the logical structure of the counterfactuals that capture causation. In this section, I present Kroedel's exposition of counterfactual mental causation, flagging several concerns as I proceed.

2.2.1 Kroedel's General Conception

Note that Kroedel sees himself as providing sufficient, but not necessary, conditions for causation.[56] He states the weak assumptions on which he predicates his project:

> We do not even need a full-blown theory of causation that states necessary and sufficient conditions for causation in terms of difference-making in order to solve the problems for dualism and non-reductive physicalism. It suffices to assume that difference-making is sufficient for causation. This assumption is very plausible. Since it does not amount to a full-blown theory of causation, it is also relatively metaphysically weak.[57]

Kroedel couches his approach on minimal assumptions about 1) sufficiency, 2) difference-making, and 3) relevance to solutions to the issues of mental causation.[58] First, he wishes to explore a theory of causation that captures what is minimally sufficient for causation. Second, Kroedel takes our informal use of the language of "difference-making" to be captured by

[55] Kroedel (2019: ix).
[56] Note that authors like Mumford and Anjum (2011: 52) see sufficient conditions as occupying a "causal necessitarian" viewpoint, which they contest.
[57] Kroedel (2019: 4).
[58] Regarding 3), he states just before this passage that he wants to solve the Interaction and Exclusion Problems.

counterfactual formulations of causation.[59] Finally, Kroedel formulates his account for application to mental causation. We will see whether or not he can reach the goals he sets for himself within the chosen parameters.

Kroedel commences by providing a brief excursus on the kinds of relata between which causation can hold, namely, events.[60] But he wishes to avoid certain difficulties about events. First, he wishes to avoid the insufficiently differentiated Quinean spatio-temporal account of events which fails to distinguish, in a causal analysis, between the *rotation* of a sphere and its *heating up*.[61] Closely related, Kroedel wishes to avoid the famous "*qua* problem," that is, how the mental can cause *qua* mental, in some way, as opposed to *qua* physical.[62] Interestingly, he turns to a Kimian account of events for solutions. Generally, this account says that an event is the threefold combination of an object *x*, with a property *P*, at a time *t*[63]. Kroedel complements this account of events with the Lewisian stipulation that events occur in certain spatiotemporal regions.[64] Since I will address the issue of events in greater detail in the next chapter, I leave aside further exploration for now.

Having stated his preferred view of events, he proceeds to analyze how causation obtains between events. As mentioned above, if one is going to have a counterfactual account of mental causation, *a fortiori* one needs a counterfactual account of causation.[65] Thus, he writes,

> Say that event *e counterfactually depends* on event *c* just in case *e* would not have occurred if *c* had not occurred. Then the principle says that for any two events *c* and *e* that actually occur, if *e* occurs later than *c*, and *e* counterfactually depends on *c*, then *c*

[59] "What makes a difference is a cause. This is the central principle about causation that I shall use in this book" (2019: 30). Also, 30 n. 37.
[60] NB: this phrasing would make many theorists unhappy. For it assumes that causation is the connection (a third entity) between two relata. Since such phrasing is common and endorsed by Kim and Kroedel, however, I leave discussion until Chapter 3.
[61] Kroedel (2019: 20-24). Spatiotemporal identity is not fine-grained enough.
[62] See Sosa (1984), Davidson (1993). Relatedly, Hornsby (1981).
[63] Kim (1976: 160-161).
[64] Kroedel (2019: 28).
[65] Kroedel outlines this account in (2019: Chapter 1: §1.3-§1.7, Chapter 2).

> causes *e*. The principle states a sufficient condition for causation in terms of counterfactual dependence.[66]

This statement is merely Kroedel's provisional attempt to give a sufficient condition for causation in terms of counterfactuals.[67] Kroedel gets the condition from two separate claims. First and centrally, there is the claim of the counterfactual dependence between two events. Second, there is the claim that the first event must occur "later than" the second event (and both events "actually occur"). No doubt, those familiar with the trajectory of Lewis's writings on causation are aware that a counterfactual analysis of causation requires further specifications and stipulations beyond the *mere* counterfactual dependence of events.[68] Kroedel too, in offering his sufficient condition, acknowledges this much. These further specifications—sometimes quite complex—must be introduced to deal with more difficult features of causation like temporal asymmetry, transitivity, preemption, and "chanciness."

A last and crucial aspect of Kroedel's account is his explicit dependence on Lewisian counterfactual analyses.[69] He states: "Lewis's idea is that a counterfactual is true just in case we have to depart further from the actual world to find a world where the antecedent of the conditional is true while its consequent is false than we have to in order to find a world where both the antecedent and the consequent are true."[70] (To endorse this notion of similarity of worlds does not require embracing Lewis's modal realism). A counterfactual, then, comes out true when a set of worlds very similar to the actual world with the antecedent true, in which the consequent is also true, are all closer to the actual world than any world where the consequent

---

[66] Ibid., 30.
[67] I will continue to call it his "sufficient condition," though sometimes he lapses into the more expansive language of an "account" or "theory" of causation.
[68] Menzies and Beebee (2020), Bernstein (2019), Lewis (1973a), (1979), (1986), and (2000), for example. Note two other features of Kroedel's account. First, he wants it to address cases of late preemption, in which one cause preempts another, taking credit for the effect. Second, he wants to avoid the familiar problem of overdetermination.
[69] "I shall assume Lewis's (1973b [*Counterfactuals*]) truth-conditions for counterfactual conditionals" Kroedel (2019: 31).
[70] Ibid.

comes out false.[71] Thus, Kroedel formulates a sufficient, counterfactual condition for causation, predicated on a Kimian view of events and based on Lewis's account, for application to mental causation.

2.2.2 The Logic of Counterfactuals

With this background in place, Kroedel sets himself the task of assessing the validity of a handful of different inferences with counterfactuals. A classic problem arises that inferences that hold between material and strict conditionals don't also always hold between the corresponding counterfactual inferences. Though Kroedel goes through a few different examples of this problem, I focus on one instance for the sake of concision.[72] Consider the following inference using counterfactuals:

(1) If Φ were the case, then X would be the case. ($\Phi \mathrel{\square\!\!\rightarrow} X$)
(2) If X were the case, then Ψ would be the case. ($X \mathrel{\square\!\!\rightarrow} \Psi$)
---
(3) If Φ were the case, then Ψ would be the case. ($\Phi \mathrel{\square\!\!\rightarrow} \Psi$)

The conclusion, according to Kroedel, is false. Consider the following insertion of concrete examples to see why:

(a) If I were king, then I would wear a crown.
(b) If I wore a crown, people would find me ridiculous.

(c) If I were king, people would find me ridiculous.

[71] I don't discuss 'might' conditionals. As Kroedel says, "Such 'might' conditionals are not directly relevant to the counterfactual dependence between events," 32.
[72] Example from Kroedel, 34.

Kroedel uses a similarity of possible worlds analysis to explain why the inference fails.[73] Starting with (a), the antecedent of this counterfactual—that I am king—is *relatively far from the actual world*. Nevertheless, since the consequent—wearing a crown—reliably follows the antecedent, (a) is true. In the case of (b), the world in which I wear a crown—perhaps a tinsel one for a party—is *relatively close to the actual world* and the consequent reliably follows the antecedent. Since the antecedents of (a) and (b) are of different distances from the actual world, however, the combined counterfactual of (c) is not reliable. As a result, the inference fails.

This failure of inference, Kroedel maintains, does not express itself in the more familiar inferences between material conditionals or strict conditionals.[74] Examining just the first of these to see why, consider the following inference:

(1) If Φ is the case, then X is the case. (Φ ⊃ X)
(2) If X is the case, then Ψ is the case. (X ⊃ Ψ)
---
(3) If Φ is the case, then Ψ is the case. (Φ ⊃ Ψ)

Kroedel thinks the outcome is reliable since the logical status of these material conditionals cannot vary in the way that counterfactuals can, so that the inference is always valid.[75] The antecedents of Φ, and X are not contingent as they are in Lewisian possible worlds counterfactuals. As a result, transitivity succeeds and the consequents follow reliably in each counterfactual. The same is true for strict conditionals.

Kroedel considers several ways in which the original counterfactual inference can be "repaired." As mentioned above, Lewis and Kroedel are both well aware that in order to get causation from counterfactuals, further stipulations are required. In some cases, this further

[73] Note that a recent popular response to this issue instead depends upon evaluating counterfactuals by context-sensitivity, e.g. Bernstein (2019: 165-166), Williamson (2020: 10-14, 168-174).
[74] Kroedel, 34.
[75] Lewis (1973b: 4-19), Williamson (2020: Chapter 10).

stipulation will depend upon metaphysical considerations like causal pathways; in other instances, it will simply demand a modification of the logical form of inferences. It is with the second type of modification that we are presently concerned. Kroedel writes that the original counterfactual inference "can be repaired by adding another premise to the effect that the antecedent of the first premise not only counterfactually implies the antecedent of the old second premise, but that the converse is also true."[76] The way this works is as follows:

(1) If Φ were the case, then X would be the case. (Φ □→ X)
(α) If X were the case, then Φ would be the case. (X □→ Φ)
(2) If X were the case, then Ψ would be the case. (X □→ Ψ)
---
(4) If Φ were the case, then Ψ would be the case. (Φ □→ Ψ)

The additional premise (α) added into the original inference would be unneeded for an equivalent inference with material and strict conditionals. What it ensures for counterfactuals, however, is that worlds where Φ and X both hold true are sufficiently close together so that if either Φ or X is true there, the other is as well. This additional premise (α) makes the new inference immune to the critique leveled against the original one.

What is the payoff of this type of analysis? Kroedel suggests that by considering how the logic of counterfactuals differs from those to which it has a family resemblance (material and strict conditionals), we can tailor counterfactuals in a way beneficial to giving a condition for causation. In this vein, he writes,

> The lesson to be learned from the logical peculiarities of counterfactuals is twofold. First, we may not simply use familiar inferences that are valid for other kinds of conditional, for these inferences may fail for counterfactuals. Second, we can find similar inferences that are valid for counterfactuals and that we can substitute for invalid ones… if we are sufficiently careful, we can still build powerful arguments with counterfactuals.[77]

[76] Kroedel, 36.
[77] Ibid., 38.

Thus, he admits that inferences of material and strict conditionals differ from counterfactual ones. Nevertheless, he believes that with the correct modifications, counterfactual inferences can be made valid and still yield arguments helpful to analyzing causation.

2.2.3 Counterfactuals Applied to NRP and Dualism

It is now possible to see how Kroedel's philosophical machinery can explain mental causation within both NRP and super-nomological dualism. I will only go through the formal structure of his argument for NRP because some brief comments will suffice to show how he modifies it for dualism.[78] He writes:

(1) Necessarily, *M* is instantiated if and only if a realizer of *M* is instantiated. ($\Box$ [M$\equiv \cup\mathbf{P}_M$])

(2) If none of *M*'s realizers had been instantiated, then *P** would not have been instantiated. ($\sim\cup\mathbf{P}_M \,\Box\!\!\rightarrow \sim$P*)

(3) If *M* had not been instantiated, then *P** would not have been instantiated. ($\sim$M $\Box\!\!\rightarrow \sim$P*)

(4) If *P** is instantiated later than *M*, and *P** would not have been instantiated if *M* had not been instantiated, then the instance of *M* causes the instance of *P**.

(5) *P** is instantiated later than *M*.

(6) The instance of *M* causes the instance of *P**.

Please note, first of all, that there are some serious worries with this formulation of the argument. First and foremost, note that though causation is supposed to hold between events for Kroedel, he has here structured the argument as an analysis of causation between properties.[79] Each entity represented above by a capital letter is either a property, or in the case of those preceded by the symbol "$\cup$," a set of properties. Second, it is unclear what a material biconditional is supposed to

[78] See Kroedel (2019: 64-65) for the source of this argument.

[79] In what follows, I assume that causation occurs between properties. To make sense, Kroedel would need to defend this claim. Perhaps the presentation can be aided if we grant him Kim's view of the relationship between properties and events (2005: 42).

signify in (1). Perhaps it is to be read as mutual counterfactual entailment. Third, and finally, the use of the word “instance” in (4) and (6) is misleading. Would not the capital letter (representing a property) be sufficient to demarcate that the property occurred? Thus, one would want a serious reformulation of this argument to eliminate these confusions.

With these caveats in place, the argument is supposed to proceed as follows. (1) simply states NRP’s central thesis that if a mental property *M* exists, that means that a physical realizer of *M* must also be instantiated. Kroedel’s employment of the symbols $\cup\mathbf{P}_M$ serves to avoid the threat of multiple realizability, by denoting all possible realizers of *M*.

*P**, we are told, is a physical property occurring later than *M*. In light of the first caveat above, presumably this property is supposed to be the property that is an element of the Kimian event at the initial time. Whatever the precise meaning, this property is the physical effect, subsequent to the cause. (2) then holds that if none of the properties responsible for *M* occurred, then the physical effect *P** also would not have occurred. This premise is a stipulation ensuring that we are dealing with a case of mental causation.

(3) is the logical consequent of (1) and (2). For it simply takes the information from the biconditional in (1) and applies it to (2). Since, necessarily according to (1), *M* and the set of physical realizers of *M* co-instantiate, then we can replace an instance of the set of the physical realizers of *M* with *M* itself. Thus, we can get the simplified counterfactual ($\sim M \mathrel{\square\!\!\rightarrow} \sim P^*$), which is (3).

Premise (4) must bear a lot of weight. Kroedel’s explanation of the premise runs as follows: “We saw that counterfactual dependence is sufficient for causation that is forward in time. Applied to our case, this yields [(4)].”[80] There are two key features of this premise. First, the second property instantiated depends counterfactually on the instantiation of the first.

[80] Kroedel (2019: 64).

Second, this feature is enough, according to Kroedel, to ensure causation. Not as central, but also of relevance, is his assertion that the property classified as an effect occurs later than the property on which it depends. It is this last and less important stipulation that permits the simple statement of (5). As will be explored later, however, a temporal ordering requirement is not demanded by all versions of causation.

The critical conclusion (6) simply states that mental causation occurs; that is, the kind of mental causation in which a mental property causes a physical one. Kroedel asserts that this conclusion follows from the combination of premises (3), (4), and (5). I refrain, for the moment, from targeting vulnerable premises. Instead, I must first mention how he expands the argument for dualism.

Drawing on the nomological necessitation already explored in the previous section, Kroedel believes that psychophysical laws ensure the joint instantiation of mental and physical properties, even in the case of dualism. Thus, he writes,

> According to nomological psychophysical supervenience, it is a matter of nomological necessity that a mental property is accompanied by some physical property whenever it is instantiated, and it is also a matter of nomological necessity that the mental property is instantiated whenever one of the physical properties that can underlie its instantiation is instantiated.[81]

Here he rearticulates his preferred version of dualism. Recall that his strategy is simply to explain that dualism will operate in a manner parallel to NRP. The reason for this parallel structure is that in both, there is a nomological connection between mental and physical properties—even if it is not that of strong supervenience.

Now recall from the above argument concerning NRP that Kroedel got his conclusion from asserting premises (3), (4), and (5). It can be noted that premises (4) and (5) can remain unchanged without detriment to Kroedel's preferred articulation of dualism. Thus, it is still licit

[81] Ibid., 81.

to say both that two events counterfactually linked are a case of causation (4); and that there is a first and second event occurring in temporal sequence (5). The difficulty now is that (1) no longer seems like an acceptable premise for the dualist and (3) depended on (1) for defense. Thus, Kroedel's strategy is to find a new way to generate (3), without reliance on a premise that is no longer viable.

How does this work? Well, notice that (2), from the original argument, *is* acceptable to the dualist. (2) is acceptable, because on Kroedel's version of dualism, *M* depends on the instantiation of a base, even if the mental and physical properties aren't linked by the kind of strict supervenience necessary in the case of NRP. So, we need only to replace (1). Kroedel cleverly accomplishes this task. Essentially, he replaces (1) with two counterfactuals, but avoids the need for a biconditional:

> (1a) If *M* had not been instantiated, then none of *M*'s bases would have been instantiated. ($\sim\cup\mathbf{P}_M \mathrel{\square\!\!\rightarrow} \sim$M)
>
> (1b) If none of *M*'s bases had been instantiated, then *M* would not have been instantiated. ($\sim$M $\mathrel{\square\!\!\rightarrow} \sim\cup\mathbf{P}_M$)

Note that psychophysical nomological linkage is preserved, since if one property fails to be instantiated, the other will too—in both directions. However, the combination of (1a) and (1b) is weaker than (1), which necessitated co-instantiation of the two properties. Nevertheless, (1a) and (1b) combined with (2) conveniently still produce (3). Thus, Kroedel again has the resources—(3), (4), and (5)—to get his desired conclusion (6). (6) states that a mental property (or perhaps more charitably: the mental property *of a certain event*) can cause a subsequent physical property (*of a subsequent event*). Thus, even Kroedel's version of dualism preserves the reality of mental causation based on his counterfactual analysis.

2.3 Epicycles: Problems with a Counterfactual Condition of Causation

The general form of my objection to Kroedel I call the "Epicycle Objection."[82] This objection gets its force from the fact that a counterfactual sufficient condition for causation always needs further specifications[83]—epicycles—and that the criteria for choosing these specifications are arbitrary. I think that knowing what counts as good criteria for establishing the *additional stipulations* to a counterfactual sufficient condition of causation are elusive. Then, since the condition always requires introduction of further specifications, a counterfactual condition for causation could never suffice.

At this point, a very powerful objection to a counterfactual sufficient condition of causation can be formulated. While some of the nuts and bolts were excised,[84] Kroedel's overall argumentative counterfactual strategy was presented in the previous section. Notice that any successful Lewisian-type view of causation is going to include counterfactuals *and* some *additional stipulations*. The reason for the presence of this second feature is that counterfactuals on their own are too general to explain causation. How are we to identify the correct *additional stipulations* that must be supplemented to true counterfactuals? The common form of my objection is as follows: 1) a counterfactual account of causation always ends up embracing counterfactuals *plus* additional stipulations, 2) there are no good criteria on which to decide the *additional stipulations*, 3) therefore, a counterfactual condition for causation cannot be

[82] The pejorative notion of epicycles comes from attempts to explain planetary movement within the Ptolemaic system. The example expresses that small emendations to a bad theory don't rectify it. To eliminate (metaphorical) epicycles, the foundational theory sometimes should be abandoned. Lewis, oddly unconcerned, sometimes seems aware of this attack on counterfactual accounts of causation. E.g.: "the simplest counterfactual analysis breaks down in cases of redundant causation, wherefore we need extra bells and whistles. I have changed my mind once more about how those bells and whistles ought to work" (2000: 182).

[83] Lewis (1986), for example, realizes the need for such criteria. He tries to ensure that *asymmetry* in causation depends on his possible worlds analysis and not on commonsense. Thus, he introduces "backtracking" limitations. Also, Hall (2004: 275, n. 7).

[84] See the Barometer Example, below, for an important complication: namely, Lewis's *miracle approach* criteria; see (2019: Chapter 3) for Causal Modelling; see (2019: Chapter 4) for an essayed solution to Exclusion.

satisfactory. In the following sections, I present two attacks on Kroedel's approach. If successful, they challenge his weight-bearing premise (4)—which essentially states the counterfactual condition for causation—so that, as with Kim, the argument is undermined. Because of its generality, the sufficient counterfactual condition requires (arbitrarily chosen) *additional stipulations* to capture causation. I begin with a critique based on *temporal ordering* and then I proceed to a critique based on *overlapping realizers*.

2.3.1 First Problem: Temporal Ordering in The Barometer Example

I here show the difficulties created by an example where counterfactuals about the world are true, but non-causal. If my analysis holds, then Kroedel relies on epicyclic *additional stipulations* to make his counterfactuals apply to causation. In this section, I posit two counterfactuals which both appear to be non-causal, though one violates temporal ordering and one does not.

To begin, note that both premises (4) and (5) above contained reference to the temporal ordering of events. On Kroedel's view, counterfactuals must, first of all, behave in such a way that the past does not counterfactually depend on the present or the future. Why? Because if the counterfactual has the wrong temporal direction, then it does not describe causation.[85] Thus, Kroedel takes it as a basic commitment that a good account of causation must work in the right temporal order. In a complicated and important example, the Barometer Example, Kroedel introduces two problematic counterfactuals. I quote the passage in full:

> The air pressure drops. A little later, the barometer reading falls. Suppose that our similarity relation had the result that if the barometer reading had not fallen, then the air pressure would not have dropped. We would not like to say that the falling of the

[85] Assuming with Kroedel, for the present, that causal relata are events. I will question this assumption next chapter.

> barometer reading causes the earlier drop in air pressure, and we do not have to say this, for our principle about causation is restricted to putative effects that occur after their putative causes. So far so good. But suppose further that after the barometer falls, there is a storm. If it is true that the air pressure would not have dropped had the barometer reading not fallen, presumably it is also true that the storm would not have occurred had the barometer reading not fallen.[86]

Notice the example contains the following three events in a temporal sequence:

(A) The air pressure drops.
(B) The barometer reading falls.
(C) There is a storm.

According to our commonsense notion of causation, which includes temporal ordering, (A) causes (B), but interestingly, (A) also causes (C).[87] Nevertheless, based on this example, Kroedel can construct two separate counterfactuals that seem to come out true even though they don't describe what we would normally think of as causation. The first is that "if the barometer reading had not fallen, then the air pressure would not have dropped":

(BC1) If ~B were the case, then ~P would be the case.

Now it can be seen that this counterfactual is true, given certain assumptions. Since the world described is a world in which (A), (B), and (C) above all hold, we can assume the standard causal relations between these events hold (whatever precisely those are). Similarly, we assume that the barometer is in working order. If the barometer breaks, (BC1) comes out false. Since neither assumption seems overly demanding, the counterfactual is reliable.

Kroedel nevertheless dismisses this counterfactual, saying: "our principle about causation is restricted to putative effects that occur after their putative causes." Relying on a Lewisian analysis, for these troublesome cases we would have to 'rewind the clock' and examine the historical events occurring before the first event of the counterfactual. Such an analysis is called

[86] Kroedel (2019: 39).
[87] Lewis (1986: 234) for similar example.

a "back-tracking" evaluation.[88] A backtracking evaluation holds fixed the history of the world until right before the first event of the counterfactual, then reexamines prior events to determine whether the relevant counterfactual is actually causal. According to Kroedel, if counterfactuals are to prove useful for analyzing causation, they must *avoid* backtracking.

Indeed, Kroedel spells out two strategies for avoiding backtracking, both involving Lewis's similarity of worlds analysis.[89] I focus on his second response, since he uses the first only for convenience, but emphatically advocates the second strategy.[90] The second strategy is Lewis's so-called "*miracles approach*." Glossing over some details, this approach simply uses avoidance of miracles—first and foremost of "big, widespread, diverse violations of law"—and maximization of spatiotemporal overlap to capture the notion of similarity between worlds.[91] In one brief paragraph, Kroedel claims this approach rules out backtracking and thus banishes (BC1) from the class of causal counterfactuals.

Envision two worlds. The first is a world in which a "small miracle" causes the barometer to break immediately before the needle moves on account of dropping air pressure. The second is a world in which a "small miracle" prevents the air pressure from dropping in the first place. In both worlds, the needle doesn't move. He writes:

> The world where the barometer malfunctions just before the time at which its reading actually fell has more perfect match of particular fact with the actual world than the world where there is no earlier drop in air pressure… the world with the malfunctioning barometer wins the contest for similarity with the actual world… [so] no backtracking ensues.[92]

---

[88] Kroedel (2019: 39); Lewis (1979), Lewis (1973a: 567). Note that for Lewis temporal asymmetry is not assumed, but necessary given the correct counterfactual analysis (1979: 61ff).
[89] Kroedel (2019: 40-41), Lewis (1979).
[90] The first he calls the "*asymmetry-by-fiat approach*" which requires the closest antecedent worlds to match the actual world until just before the time of the effect. The second is explored above.
[91] Lewis (1979: 472). Note also that the *miracles approach* assumes determinism, Kroedel (2019: 45-46). I don't question the assumption here, since chancy accounts of causation raise further issues, likely explainable by rephrasing outcomes as increased likelihoods rather than certainties (see Menzies and Beebee 2019: 1.4, 5.4).
[92] Kroedel (2019: 42-43).

The claim in this passage is that a world with a *later* deviating miracle is closer to the actual world than one with an *earlier* deviating miracle. Thus, assuming the needle moves in the actual world, a barometer that breaks just before moving is closer to the actual world than one in which the air pressure had never dropped. Applying the *miracles approach* allows Kroedel to rule out backtracking. By sidelining back-tracking cases, (BC1) no longer counts as the good kind of counterfactual. Why not? The dependence proposed by (BC1) can only hold by being further from the actual world than worlds in which the needle doesn't move because the barometer breaks. But if the needle breaks and can't move, then (BC1) is false, because dropping air pressure cannot depend on a non-occurring event.

There are several issues with Lewis's *miracles approach* to preserving temporal ordering. The first is evident in Kroedel's conclusion to his argument:

> Instead of claiming that counterfactual dependence as assessed according to the miracles approach is sufficient for (forward-in-time) causation, we claim that counterfactual dependence as assessed according to the miracles approach is sufficient for causation *if* the relevant counterfactual is not evaluated in a backtracking way.[93]

If a condition of counterfactual dependence requires the *additional stipulation* to avoid backtracking, it's no longer "sufficient." While Lewis's similarity of worlds might do the trick, it's no longer a purely counterfactual condition for causation. This restriction, then, is a first epicycle. Perhaps we can forgive this stipulation, however, since most reasonable accounts of causation will need it.

Straight away, however, a much deeper worry arises. We can construct a second counterfactual from the Barometer Example that both comes out true and is even more dubiously causal than (BC1). Oddly, Kroedel does nothing more than passingly dismiss this counterfactual.[94] Kroedel writes: "the storm would not have occurred had the barometer reading

[93] Ibid., 45.
[94] Ibid., 39.

not fallen." Reformulating this counterfactual, it contains the idea that "if the barometer reading had not fallen, the storm would not have occurred." We can formalize this new counterfactual:

(BC2) If ~B were the case, then ~S would be the case.

Unlike (BC1), this counterfactual seems true *and* preserves correct temporal ordering. Thus, (BC2) cannot be excised from an account of causation for its failure to maintain temporal ordering. It again assumes that the barometer is in good working order. This counterfactual also assumes that we inhabit a world where the normal causal relations obtain between (A), (B), and (C), with (A) causing both (B) and (C). Nevertheless, it is surely wrong to suppose that a barometer needle could cause a full-fledged storm.[95] We now have a true counterfactual that is non-causal and is thus a counterexample to Kroedel's sufficiency condition for causation. Thus, if (BC2) is to be ruled out, it will have to be on other grounds. But if it is, it will again be by means of an *additional stipulation*. Epicycles begin to abound.

2.3.2 Second Problem: Overlapping Realizers in The Opacity Example

There is a second kind of case that also presents grave difficulties for Kroedel's project. In many ways, the new type of case is simpler than the counterfactuals of the Barometer Example. These are the family of counterfactuals present in the following Opacity Example.[96]

For the sake of clarity, I similarly quote Kroedel's introduction of this particular problem:

> I hold an aluminium ladder against a power line and subsequently get electrocuted. Being made of aluminium, the ladder is an electrical conductor. Conductivity supervenes on physical properties and can be realized in different ways. If the ladder had not instantiated any realizer of conductivity, I would not have been electrocuted. It follows from the argument for downward causation that the instance of conductivity causes my

[95] I set aside until Chapter 3 consideration of a Millian "total cause."

[96] While I don't delve into the topic of deviant causal chains, note that the Opacity Example has a family resemblance. Deviant chains appear to be instances of the "wrong" kind of causation. Opacity here also seems to be the "wrong" kind of property on which to place causal dependence in a transitive chain.

> electrocution… But being made of aluminium, the ladder is also opaque. Opacity too supervenes on physical properties and can be realized in different ways. The realizers of opacity are closely related to the realizers of conductivity. Almost all conductors are opaque. Some conductors are transparent… but they are not used to make ladders. Thus, it seems that if the ladder had not instantiated any realizer of opacity, I would not have been electrocuted either. It follows from the argument for downward causation that the instance of opacity causes my electrocution.[97]

There is, first of all, a somewhat obvious counterfactual that if the ladder were not conductive, then I would not have been electrocuted. In the Opacity Example, Kroedel introduces the idea that the realizers of both conductivity and opacity are closely related, such that substituting one for the other can yield a new counterfactual. Ignoring for the moment the supervenience-based rationale of the relationship between conductivity and opacity, the relevant counterfactual runs as follows, "if the ladder had not been opaque, I would not have been electrocuted." Formally:[98]

(OC) If ~O were the case, then ~E would be the case.

Is this counterfactual true? Since "[a]lmost all conductors are opaque," it would be unusual for a transparent ladder to be conductive.[99] Thus, in all likelihood, this counterfactual is reliably true. Further, it does not violate the *additional stipulation* given above to avoid backtracking. On the counterfactual account so far presented, then, (OC) seems to be a genuine case of causation. One might worry that one property from a pair of properties that supervenes on the same base should not be able to lay claim to equal causal efficacy with the one that clearly explains the effect.[100] In the remainder of this section, I shall argue that Kroedel's responses to this charge are unsatisfactory, such that we have a new form of the arbitrary epicycle attack.

---

[97] Kroedel (2019: 68).
[98] Ibid., 69, expresses it as follows: "If opacity had not been instantiated, then I would not have been electrocuted. (~O □→~E)"
[99] Kroedel (2019:70): "It is natural to think that if the ladder had not instantiated any opacity-realizer, then it would have been made of some middle-of-the-road transparent material (glass or transparent plastic, say), which would not have been conductive."
[100] Keep in mind the qualification that property instantiations should be seen as partially constituting an event, the actual cause.

Kroedel gives five ways of responding to the counterfactual that would permit us to say that opacity is *not* a cause of the electrocution. The first two deny the closeness of opacity- and conductivity-realizers, so that we cannot presume that the absence of the first entails the absence of the second.[101] The third strategy holds fixed certain surrounding events to ensure that the second event counterfactually depends on the first.[102] Kroedel's fourth and fifth response are more radical. The fourth simply denies the sufficiency of a counterfactual condition for causation —obviously an approach he cannot endorse. The fifth approach bites the bullet, accepting that opacity does cause electrocution.

Each of these approaches creates formidable issues for Kroedel's counterfactual condition for causation, and by extension, the viability of any such account. Since the fourth response would self-undermine his project, we can set it aside. First, strategies one through three all employ a similar idea. They all attempt to separate opacity from conductivity, so that we don't have to count the first as a cause, while maintaining the causality of the second. There are two things to be said about these strategies. First, note briefly that the given example, by relying on a fairly strict sense of supervenience, is much more conducive to elucidating mental causation from an NRP perspective than from any variety of dualism. This fact could be a cause for concern.

Second and more importantly, why does Kroedel ultimately reject strategies one through three in favor of the fifth strategy? The answer, I suggest, is that he implicitly recognizes that the first three approaches all employ further *additional stipulations*, making them vulnerable to the epicycle critique. The first two strategies, for instance, require denying that whenever we have conductivity, we have opacity. But in denying the closeness of these two properties, one must

[101] The first is a denial of the following: "If no opacity realizer had been instantiated, then no conductivity-realizer would have been instantiated" (Kroedel 2019: 70). The second reaches an identical conclusion by weakening the notion of events involved.
[102] Ibid., 73-74.

give a reason for their separability, which in this case is that since there are transparent conductors and opaque non-conductors, the two properties are in principle separable. Thus, worlds in which the ladder is composed of some "exotic" transparent conductor must be, *ex hypothesi*, just as close to the actual world as worlds in which the ladder is composed of a normal opaque conductor.[103] If someone uses the first or second strategy, this reason is the sort of explanation they have to endorse. But this reason, of course, itself counts as an *additional stipulation*, beyond the bare counterfactual condition. Thus, the idea that opacity and conductivity are separable relies on a further criterion. But then we are back again to the same problem: from whence does the epicycle come?

The third response attempts to complexify the counterfactuals involved. Instead of focusing on one antecedent, it requires that several events be "fixed" to ensure the outcome of the relevant counterfactual. Thus, the helicopter becomes airborne only if the engine starts *and* the blades are attached to the rotating mast. Essentially, a flight only counterfactually depends on engine motion when the event of the blades being attached is held fixed. Otherwise, of course, the engine could fire up quite successfully without producing consequent flight.

Kroedel says that the proper way to spell out this idea—holding fixed certain "off-path" events—depends on investigation of more complicated causal modelling approaches. Without getting into the technicalities of those causal models, which are composed of various "variables" representing different aspects of a system, note that a "path" is defined as a sequential linkage of these variables with an ordered dependence on each other.[104] When spelling these out later, Kroedel endorses what he claims to be the best response to the counterfactual formed from the

[103] E.g. Gordon (2000): ZnO, Indium Tin Oxide, and Carbon nanotubes are common transparent conductors. Kroedel writes: "it requires giving up the natural thought that the ladder would have been made of some middle-of-the-road transparent material if it had not instantiated any opacity-realizer" (2019: 72). Though Kroedel's stated rationale isn't false, mine will explain more.
[104] Kroedel (2019: 121).

Opacity Example: "[we] formulate what seems, overall, the best response to the problem of overlapping realizers, namely that the opacity-instance is a cause of the electrocution, but one that is typically considered explanatorily relevant."[105] Thus, in the Opacity Example, it is not that the best sufficient condition for causation eliminates opacity as a cause, but rather that its causal role is diminished because of its explanatory irrelevance. This irrelevance is due to its merely "off-path" influence on the effect, whereas conductivity has "on-path" influence.

Thus, the final and indeed most attractive response to these worries for Kroedel is to simply accept that opacity is a cause of electrocution. He thus writes, "We shall have to accept the result that the opacity-instance causes the electrocution, but we can try to explain away the implausibility of this result."[106] Our only hesitation for accepting this view, according to Kroedel, is that we think opacity has "little explanatory relevance in our context."[107]

Without wishing to conflate explanation and causation, surely a genuine cause is explanatorily relevant? Note that he is not overly worried about stipulating explanatory relevance as a criterion, because mental causation does not exhibit the kind of explanatory irrelevance found in the case of opacity and electrocution. He writes, one response "holds onto our [counterfactual] principle about causation but denies that the opacity-instance is a cause that is explanatorily relevant in our context, [and so] does not threaten the status of mental causes. For mental events do typically count as explanatorily relevant."[108] The difficulty of this conclusion for Kroedel is that it requires a crucial addendum to his counterfactual account. If "explanatory relevance" is a necessary *additional stipulation*, then Kroedel has admitted that a counterfactual condition, on its own, cannot capture causation.

---

[105] Ibid., 129.
[106] Ibid., 74.
[107] Ibid.
[108] Ibid., 75.

This final response falls prey, once again, to the epicycle critique. If explanatory relevance is read as being an essential part of the best definition of a condition of causation (something Kroedel wants to avoid), then we simply have as our finalized condition: counterfactuals + on-path causality (the latter being an *additional stipulation*). In that case, counterfactuals do not provide a sufficient condition for causation. If, on the other hand, the criterion of explanatory relevance is supposed only to make a given cause more important than other ones, the less relevant cause nevertheless still has causal power. Thus, when Kroedel bites the bullet in the fifth response, he holds that opacity is just as much a cause of electrocution as conductivity. On the one hand, this conclusion seems false. But more importantly, it becomes pressing to know why on-path causation is more explanatorily relevant. A good condition for causation should be able to express why explanation via on-path causes is more helpful than explanation via off-path causes. By stipulating a temporal ordering of events, eliminating backtracking, or relying on explanatory relevance, Kroedel falls prey in multiple ways to the epicycle critique.

3. A Powers Theory of Causation

But then I do not think that there are any cogent arguments against admitting all sorts of "non-Humean' features into the knowable empirical world in the first place, features like causings and powers, and exercises of powers and interferences with those exercises. – Nancy Cartwright (2009: 153)

Before defending a more tenable theory specific to mental causation, I advance a general powers or dispositional theory of causation. With this general theory of causation in hand, it will be possible, with great benefit, to apply it to mental causation in the final chapter. I contend that

not only is a powers view of causation intrinsically attractive, but that it offers the most feasible way of solving the problems of mental causation. More specifically, I contest that it has the resources to avoid the problems of *overdetermination* and *additional stipulations* that plague Kim and Kroedel. Nevertheless, I plan to take something of importance from each of those thinkers. I follow Kim in his endorsement of a "production" view of causation, even if I do not subscribe to his physicalism. From Kroedel, I borrow the argumentative strategy of accommodating a range of metaphysical possibilities.

Theories of causation can be divided into two main families.[109] The counterfactual or dependence family traces its genealogy from Hume through its prolific twentieth century advocate David Lewis.[110] This theory relies on counterfactual dependence to express causality. As elaborated last chapter, Kroedel is a recent advocate, using such a theory to provide a sufficient condition for mental causation.

A second family of causal theories are the so-called "production" theories of causation.[111] There are different views about what is produced, with possible candidates being energy, powers, or any kind of physical quantity. Speaking inclusively, all such accounts argue that the effect is *produced,* in some real way, by the cause. As Ned Hall contrasts it with the first family of causal accounts, "The second variety is rather more difficult to characterize, but we evoke it when we say of an event *c* that it helps to *generate* or *bring about* or *produce* another event *e*, and for that

[109] See Hall (2004), Kim (2005: 18 n. 12). It is likely that the two families are attempting to answer slightly different questions, the first providing a conceptual analysis of causation, the second attempting to explicate its ontology. My account falls in the second group, but that does not rule the possibility that a good conceptual analysis would describe this ontology well. Thus, the two approaches may not be entirely incompatible.
[110] Note that what I call "counterfactual" accounts might fall in a broader camp of "difference-making" accounts, which would include accounts like Mackie's INUS condition for causation (1965) and probabilistic accounts, Lewis (1986: 214, 217).
[111] Holger and Gunther (2020), though their concerns are somewhat epistemological. NB: Kroedel calls these "transference" views.

reason I call it 'production.'"[112] Kim's account falls within this branch of accounts and, in what follows, I defend another such production theory.

I will outline and defend a neo-Aristotelian powers view of causation in this chapter. Though not a dominant view in contemporary philosophy of causation, a powers view of causation has been defended by some.[113] In this chapter, I motivate and present this view of causation, starting from the account expounded by Stephen Mumford and Rani Anjum in their book *Getting Causes from Powers*.[114]

The plan for the chapter is as follows. First, I motivate the view with a challenge from Nancy Cartwright and a glance at "finks." Second, I sketch the Aristotelian background, before turning to and qualifying Mumford and Anjum's powers view of causation. In the third section, I advance an argument for preferring processes to events as an ontology of causation, for three reasons. These are the simultaneity in causation, the possibility of interruption of causation, and lastly, the need to distinguish between preexisting conditions and actual causes.

3.1 Cartwright's Challenge

Nancy Cartwright is well known for her skepticism that the laws of nature accurately represent the world. She expresses her challenge against the reliability of causal laws, "I have become gripped with despair about the general uselessness of our hard-won causal

[112] Hall (2004: 225).
[113] Interestingly, Paul and Hall, in their handbook of causation, focus on counterfactual theories, writing: "*Counterfactual dependence* seems sufficient for causation: if it's true that had *C* not happened, *E* would not have happened, then it seems that *C* must for that reason be among the causes of *E*" (2013: 4). Defenders of powers causation include Prior, Pargetter, and Jackson (1982); Mumford and Anjum (2011); Gibb (2015); Marmodoro (2017); Cartwright (2009, 2017); Mayr (2017). Interestingly, Choi and Fara (2018: §6) say those who *don't* grant causal efficacy to powers occupy a more unusual position than the alternative.
[114] Mumford and Anjum (2011).

knowledge."[115] She continues more optimistically, "*Capacities*, I claim, are a source of stable causal laws. When a capacity is properly triggered it will regularly exercise itself in a canonical way."[116] Stability and regularity, on her view, are necessary if causal laws are going to be *predictive*. Though Cartwright doubts that causal laws in isolation account for natural regularity, she thinks that instead of giving up hope, we should turn to capacities—or powers—for solutions. Cartwright's reasons for despairing of causal laws and her defense of real powers are an helpful segue to my own defense of powers. One of Cartwright's main reasons for defending powers is the inability of Humean laws to *explain the predictability* required for scientific investigation.

Cartwright, after she expounds what she means by "causal law," still finds the notion inadequate. Her view of causal laws satisfies several criteria. They are population-relative, prioritize singular causation, specify what is sufficient for causation, can accommodate producing the effect with a variety of frequencies (including failure), and pertain to *factual* effects.[117] She thinks that her view of causal laws is superior to views that don't specify the law's contents, are non-empirical, or don't track sufficiently well current scientific methodology. Then she turns to the difficult cases where causal laws seem to be violated. The question arises: Is a law invalidated by a violation?

She criticizes answers to this question that rely on adding qualifying clauses. She directs us to James Woodward's further specification, building on an account using systems of linear equations to capture causality:

> *Modularity*: there must be at least one way to change the other causal relations in a system that leaves any genuine causal relation invariant.[118]

---

[115] Cartwright (2009: 144).
[116] Ibid. I set aside her argument that nomological machines can play a similar role to that of capacities.
[117] Ibid., 129-130.
[118] James Woodward, *Making Things Happen*, 2003, op. cited (2009: 137).

The “change” here is similar to Lewis’s “small miracle”—a minimal intervention that interferes with an otherwise expected causal outcome. The trouble, Cartwright thinks, is that this criterion allows many relations between variables bearing the title “causal relations” to be manipulable. So, if Woodward stipulates invariance in relation to some causal relations, that means that many of them can, under the right conditions, fall short of invariance. She concludes, “if we want to use those laws for predicting what happens as we manipulate the causes, to think of causal laws as useful for policy prediction, we must *add on* invariance.”[119] Such a view won’t do for Cartwright, since causal laws are supposed to be predictive, without add-ons.

Cartwright contends that capacities can help us avoid adding on to a given causal law. For example: if Nancy is irritable—has a capacity to be angered—encountering an annoyance will activate her anger. Importantly, “the presence of a capacity guarantees that the matching causal law obtains.”[120] Further, she believes that we have relatively good ideas about when and when not capacities obtain. Of course, that does not mean that the law will obtain under any and all Woodward-style manipulations, but we can now explain the variability introduced via manipulations by assessing when the capacity obtains.[121] Cartwright concludes that we have the means, via capacities or powers, to provide successful outcome predictions. Thus, knowledge of causal laws is useful just when “causal laws are generated in the right way,” that is, based on our knowledge of capacities.

These considerations show that the failure of causal laws to predict outcomes can be fixed if we simply examine what underlying capacity blocked or overcame the expected outcome. Since a reliable law can sometimes fail to predict when powers succeed, the latter are

[119] Cartwright (2009: 138).
[120] Ibid., 145.
[121] Cartwright distinguishes between a power, its exercise, and the resulting ‘occurrent,’ contra Hume. (Ibid., 151).

more basic. Cartwright's finding offers a powerful reason that our best causal analysis might not be in terms of laws of regularity or counterfactual correlations, but instead in terms of powers.

Before concluding this section, I add a brief addendum on "finks." While I will not add much to this well-discussed issue, the moral of finks is largely equivalent to that drawn from Cartwright.[122] Finks illustrate that powers are more basic than counterfactual analyses and succeed in predicting outcomes when a nomic conditional analysis fails. C.B. Martin (1994) introduced the concept of an electro-fink, giving a definition in terms of a conditional:

(A) The wire is live.

(B) If the wire is touched by a conductor then electrical current flows from the wire to the conductor.[123]

Further, he stipulates that (A) is true just when (B) is. Now here's the difficulty: Martin's electro-fink is rigged so that whenever the conductor touches the wire, the current to the wire is cut. Thus, whenever the antecedent of the conditional in (B) obtains, the consequent doesn't. As a result, (A) is false. Thus, the conditional analysis is flawed.

But on a powers view of causation, the electro-fink is easily explicable once the relevant dispositions are identified. Since the conductor has the disposition to conduct electricity, we expect electricity to flow when it is in contact with another conductive object. In the electro-fink, however, the wire also has a disposition to inhibit conductivity when touched to a particular conductor. Thus, the outcome expected in light of just the first disposition does not obtain on contact, but by analyzing *all* the relevant powers, the outcome becomes unmysterious.[124]

For these reasons, then, we might question some aspects of what John Heil calls the "received view,"[125] which says that causation is asymmetrical, nonreflexive, transitive, and holds

[122] Martin (1994), (2008: Chp. 2), Choi and Fara (2018).
[123] Martin (1994: 2).
[124] One might add on *ceteris paribus* clauses. I agree with Martin (2008: 16-18) that such a project can't reduce powers.
[125] Heil (2012: 117).

between events. In particular, I will question that last point—that the proper relata of causation are a pair of events. Many endorsing the "received view" position it within a counterfactual framework, but I take the above considerations to demonstrate that this approach leaves central questions unexplained, the answering of which requires a more robust theory.

3.2 What Is Powers Causation?

One way of characterizing a powers view of causation is to look at its historical genesis, particularly in Aristotle's *efficient causation*. Another way is to examine the growing discussion of *dispositional properties* in the contemporary literature.[126] Interestingly, these approaches may coalesce, as many contemporary authors writing on dispositions classify their views as neo-Aristotelian.[127] Nevertheless, in this introductory section, I attempt each approach briefly to sketch the background of a powers view of causation.

3.2.1 The Historical Approach: Aristotle's Efficient Causation

Aristotle is the most easily identifiable historical antecedent of a causal powers view. It should be noted that what Aristotle calls causation may in fact fall under the contemporary notion of "explanation."[128] Explanation, though it might depend on certain metaphysical commitments, tends to be characterized as epistemological rather than metaphysical, so that Aristotle's theory might not cleanly parallel what contemporary theorists consider causation. Further, as Nathanael Stein notes, "it should go without saying that Aristotelian *aitiai* [causes]

[126] Ryle (2000 [1949]); Prior, Pargetter, and Jackson (1982); Mumford (1998); Bird (1998); Armstrong (2005); C.B. Martin (2008); Mumford and Anjum (2009), (2011); Ellis (2013); Maier (2020: 1.1.); Choi and Fara (2021).
[127] Mumford (1998), Mumford and Anjum (2011), Marmodoro (2017).
[128] Stein (2011).

are not *Humean* causes, and so if we thought that an account of causation must be given within a Humean framework, the four *aitiai* would indeed be disqualified from constituting a theory of that."[129] Aristotle famously proposes four causes,[130] four ways to answer the question *dia ti*—literally translated "through what?" or more idiomatically, "why?" He calls the four types of answers the *material*, *formal*, *efficient*, and *final* causes. After briefly explaining Aristotle's system, I focus on efficient causation.

The example of a statue helps elucidate the Aristotelian taxonomy of causes.[131] First and foremost, we might wonder what constitutes the statue. Answering "bronze" to the question "why?" gives what Aristotle calls the *material cause.*[132] The second answer to the question "why?" gives the organizational attributes of the statue. These include its height, hardness, and all it qualities.[133] The third cause is more controversial. "Why," one might ask, "does the statue exist?" One teleological answer says: "To commemorate a victory."[134] The fourth and final answer to the question "why?" for Aristotle responds to the question of how something is produced. In our example, we might suggest that the sculptor is this type of cause. This last explanation, the closest to contemporary notions, Aristotle terms *efficient causation*.

It is important to note that, for Aristotle, efficient causation is the cooperation of two different powers, an active and a passive one.[135] In his analysis of change, Aristotle says that the "mover" and the "moved" are really two parts of one complex whole. Further, Aristotle's account analyzes not just local motion, but any kind of change—like the cooling of coffee or the

---

[129] Stein (2011: 704).
[130] See *Physics* II 3 194b24-195a2, 195a15-26, II 7 198a14-22, and *Metaphysics* V 2 1013a24-37. Aristotle's Greek word for "cause," *aition*, also has connotations of "motive" and "responsibility."
[131] See Stein (2011: 701) for the example. I don't engage with exegetical debates on Aristotelian causation, particularly on final causation. E.g. Nussbaum (1979), Falcon (2023).
[132] Material causation is *potential* to the *formal* cause.
[133] For living things, the formal cause also explains capacities, like growth and perception.
[134] See *Physics* II 8-9. Unlike many contemporary philosophers, Aristotle believes that final causality applies to natural phenomena ranging from animal movement to rocks falling.
[135] *Physics* III 3 202a17-20; *Metaphysics* IX 1 1046a5-18; Marmodoro (2017).

melting of an ice cube. He says, “it is on the movable that it [the mover] is capable of acting. Hence there is a single actuality of both alike, just as one to two and two to one are the same interval…for these are one and the same, although their definitions are not one.”[136] Thus, for Aristotle, a change wrought by efficient causation is a single complex entity, involving a passive and active power.

I pause to consider and dismiss a possible application of Aristotle’s causal theory. Could an Aristotelian framework purchase a cheap solution to overdetermination in mental causation? If there are multiple types of causation, why not propose that in mental causation, physical causes operate as one species of causation and mental causes as another? Based on Aristotle’s *hylomorphic* view of mind and body, for instance, one could propose that efficient causation occurs at the physical level, but formal causation at the mental level.[137] Indeed, this approach seems favored by some, for example in the *factual causation* of the mental Lowe proposes.[138]

But this approach removes the kind of causal efficacy we want the mental to possess. If the mental is merely formal, though not drained of causal relevance, it does not *produce* the effect. In essence, this approach weakens the mental cause more than desirable—if we wish to preserve its ability to bring about real effects. To be truly helpful, Aristotle’s causal theory will have to be deployed differently.

3.2.2 A Dispositional Approach: Mumford and Anjum on Powers Causation

[136] *Physics* III 3 202a18-20; *Physics* II 5 196b10-11, 197a32; II 8 199b25.
[137] “Hylomorphism” is the view that natural things are composed of two principles, one *material* the other *formal*. It combines the words *hyle* (“matter”) and *morphē* (form). Dretske (1993), Mayr (2017: 87) give related contemporary suggestions.
[138] Lowe (2008). See Dretske (1993), Sider (2003) for defenses of non-Aristotelian causal pluralism.

The account here put forward builds on an Aristotelian foundation but attempts to position it within the contemporary literature on dispositions. I here advance key features of the account without yet positioning it against the accounts discussed in Chapters 1 and 2, a task reserved for later. In this section, I present and qualify Mumford and Anjum's general view of dispositions and how it applies to causation.

Dispositions are taken by many to be a kind of property, state, or condition. Enough authors, however, characterize them as properties that I too follow this practice. Dispositions—like Cartwright's capacities—are the ways in which something can manifest. Importantly, contra Hume, dispositions are distinct both from their manifestations and the results that their manifestations bring about.[139] Thus, glass has the *disposition* of fragility, which *manifests* when hit by a rock, with the *result* that shards of glass are strewn on the floor.

Dispositional properties are often contrasted with categorical properties.[140] Brian Ellis writes: "The dispositional properties are those whose identities depend on what they dispose their bearers to do, and the categorical ones are those whose identities depend on what they are—but not, apparently, on what they do."[141] The contrast has given rise to various camps,[142] but I remain agnostic on these debates, simply focusing on dispositional properties as the one's relevant to my project.[143]

Before advancing key elements of Mumford and Anjum's account, I note four main areas where I disagree with them. First, I don't endorse their defense of a third modality (beyond possibility and necessity) the so-called "dispositional modality." The fact that dispositions don't

[139] Cartwright (2017: 9).
[140] Armstrong (2005: 212-216).
[141] Ellis (2011: 14). Shape is frequently given as a paradigmatic categorical property example and fragility as a paradigmatic dispositional property.
[142] Mumford (1998: 19): "Arguable proponents [of pan-dispositionalism in which all properties are dispositions]: Popper; Mellor; and Shoemaker… Proponents of the [alternative] view [Categoricalism, which reduces all properties to categorical ones], though not explicitly, are arguably Quine, Mackie, and Armstrong."
[143] Ellis (2013); Maier (2020: 3 and 4).

necessarily manifest doesn't require anything more than *possibility*. Second, as mentioned, I see no need to endorse their pan-dispositionalism—the view that all properties are dispositional—to attain the benefits of a powers view of causation.[144] Third, Mumford and Anjum are happy to speak of isolated powers. I, on the other hand, endorse Marmodoro's view that partner-powers mutually manifest in the Aristotelian sense, described above in relation to efficient causation.[145] Thus, I defend a view of causation in which there is minimally two manifesting powers involved: an active power, the "doer," acting on a passive power, the "sufferer." Fourth, I deny their claim that properties are just *clusters* of causal powers.[146] Rather, I follow a common thread in identifying an individual disposition with an individual property.

Despite these caveats, I use Mumford and Anjum's view of powers causation as the basis for expositing my own. First, it should be noted that on their view "power" and "disposition" are synonymous terms.[147] A disposition, something like fragility, is a power that an object may manifest. Unlike some authors, they don't understand a disposition as something with a tendency towards a given outcome. It is simply a bare capacity.

Second, based on this view of a disposition, it is just as real when unmanifested as when manifested. Thus, the active manifestation of a disposition is not prerequisite for a disposition's existence. Rather, a disposition, like fragility, is a full-fledged disposition even if never manifested. They write, "The power to warm something is real, even if it is not manifesting itself in actually warming something else."[148]

Third, and of crucial importance, Mumford and Anjum maintain *causal primitivism*. This is the idea that causation is basic and therefore will never be successfully analyzed by means of

[144] Mumford and Anjum (2011: 3, 5, 6, 18, 45, 103, 125, 134). Also, Ellis (2013) for helpful discussion.
[145] Marmodoro (2017: 72-75).
[146] Mumford and Anjum (2011: 3). This sort of view follows from their pan-dispositionalism.
[147] Ibid., 4.
[148] Ibid., 6.

non-causal notions. They maintain this view against what they call "reductive" analyses of causation which try to spell out causation in non-causal concepts (in which regard they mention Hume).[149] One might immediately worry that their reasoning is circular by analyzing causation in terms of dispositions. Their response to escape this critique runs as follows: "the notions of power or disposition are already causally laden notions and it can thus reasonably be argued that unless one already has a grasp of causation, one cannot have a grasp of power."[150] Thus, they believe that powers are an intrinsically causal concept. Yet, they see this feature as a virtue of their account, since they don't want to reduce causal concepts to non-causal ones.

Fourth, Mumford and Anjum loosely endorse what they call the "total cause." This idea is based on Mill's contention that the effect comes about through "the sum total of the conditions" leading to the outcome.[151] They endorse this view since most cases of causation don't have a single cause, but rather "in virtually every case, a cause is complex, consisting in many different factors working together."[152] If this part of their view is taken to mean that powers cooperatively co-manifest when bringing about an effect, I am sympathetic. But if it is to follow Mill in believing that all preceding events are causal contributors to the effect, then I wish to resist this view. I will have reason to treat co-manifesting powers ("polygeny") at the end of this section and the possibility of an entire preceding light-cone being a cause at the end of the chapter.

Fifth, and finally, there are a couple of points on which these authors wish to remain metaphysically uncommitted. Far from being a vague imprecision, I take this strategy, like Kroedel's agnosticism between NRP and super-nomological dualism, to be an attractive move. On the one hand, Mumford and Anjum do not decide whether particular or general, or token or

[149] Also, Mayr (2017: 78).
[150] Ibid., 7.
[151] Cited in Ibid., 11. From Mill's *A System of Logic* (1842: 217).
[152] Ibid., 12.

type causal analyses are more metaphysically fundamental.[153] On the second hand, they are agnostic about the "level" at which causation occurs. The authors contest instead that perhaps there is no basic level, and further, causation may happen at multiple levels.[154] If it does not happen at the fundamental physical level, that does not preclude its arising at more macro-levels. Either way, then, dispositions are real and can serve as the basis for a powers analysis of causation.

Before concluding this section, it will be helpful to glance at a modelling method the authors introduce to deal with plural causes. The need for such a method arises because "in any real-world situation, there will be a great many relevant powers, which will dispose any particular object in all of many different directions."[155] For Mumford and Anjum, *vector diagrams*[156] are preferable to more standard *neuron diagrams* because the latter implicitly favor a counterfactual approach to causation, with which they disagree.[157] They write, "Neuron diagrams are conducive to a Humean ontology and, through a widespread and sometimes unquestioned use, they promote that ontology."[158] This claim refers to attributes of Humean theories like the constant conjunction of the causal relata. Thus, they aim to produce a model of causation that reflects their own ontology.

A vector model works by adding and subtracting the contributing causal powers. A well-produced diagram can model the differing strengths of various constitutive powers. Consider a very simple example of a vector diagram:

[153] Ibid., 14.
[154] Also, Schaffer (2003: 504-5).
[155] Kroedel (2019: 25).
[156] See especially, Ibid., chapter 2.
[157] Ibid., 19-22.
[158] Ibid., 20.

Figure 3.2: Vector Diagram

The starting disposition of a given object in a one-dimensional quality space is represented by the vertical line. *F* and *G* represent the extremes of a spectrum, say from cold to hot, for the object. So, for example, the above diagram could represent air temperature, with vectors *a* and *b* representing dispositions of the air towards each extreme. The length of the vector from the starting point represents the intensity of the disposition. Thus, in the above diagram the ratio of the lengths of vector *a* to vector *b* is 5:4. Mumford and Anjum propose a strategy of vector addition to explain what happens with the presence of these dispositions:

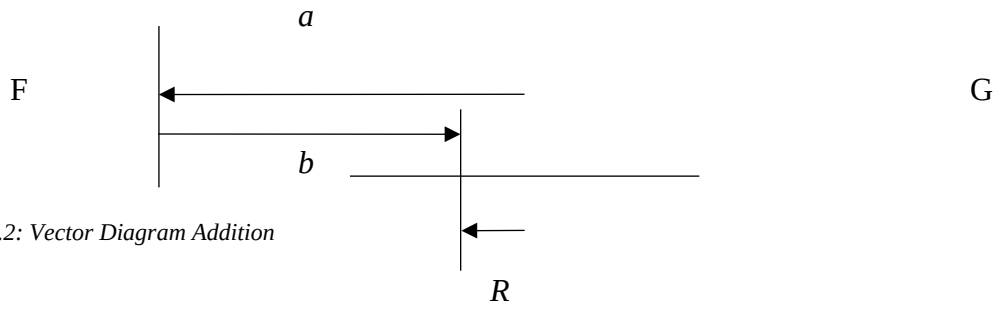

Figure 3.2: Vector Diagram Addition

Thus, two powers tending in the same direction reinforce a given effect, but two powers tending in different directions—as in this figure—cancel each other out. Thus, since *a* = 5, there are five units disposing towards *F*, whereas *b* = 4, so there are four units disposing towards *G*. Using vector addition, 5 + (-4) = 1, we get one unit disposing towards *F*. The resultant, *R*, is that the air cools slightly, given this overall disposition towards *F*. Vector addition can obviously complexify as more powers are integrated into the quality space. One important result of this

type of analysis is that causal powers can be operative even if no active outcome is observed, as in a stationary rope in a tug-of-war.[159]

Finally, note that Mumford and Anjum are aware that vector diagramming could lead to oversimplification. To respond to the worry that simple the addition of powers might not explain many cases of causation, they complexify the model by countenancing what they call *compositional pluralism*.[160] To see how this works, consider "overdose" cases. For instance, watering the plant is good for the organism's health up to a point beyond which it is dangerous.[161] There is therefore a non-linearity to the outcome produced by the power of water with respect to the plant. The authors suggest that the outcome can be explained by adopting a different function for the relationship between the various causal powers—the powers of water, soil, nutrients, sunlight, etc. The outcome can be "a genuine joint effort of two or more powers where the resultant is composed by some non-linear function."[162] This non-linear function will yield an "emergent" outcome but does so without undermining the basic intuition of vector diagramming. As they articulate more fully:

> Here we can make a relatively simple amendment to the theory: we allow compositional pluralism, which means that the mode of composition can be open. Addition is just one function – the simplest to grasp – among many other functions that may describe the compositions of powers that have magnitude. The non-additive modes of composition will be by non-linear functions.[163]

Thus, they believe that despite the difficulties presented by "overdose" cases, their model can be emended so that there is still *a* functional representation for how manifesting powers cause.

Thus, on Mumford and Anjum's view, powers are an intrinsically causal basic constituent of the world. By endorsing polygeny they allow that many dispositions can bring about a single

[159] Mumford and Anjum call such cases "equilibrium," ibid., 93.
[160] Ibid., 86.
[161] Ibid., 88.
[162] Ibid., 98-99.
[163] Ibid., 105.

effect. They propose a vector diagramming model to accommodate *compositional pluralism*: powers can combine in complex ways to cause a single effect.[164] This account, I contend, whatever its shortcomings, has radically attractive features that give it greater explanatory power than those already discussed.

3.3 Events-Pair vs. Single Process: Reconsidering the Humean Orthodoxy

Events are largely embraced as the basic constituent elements of causation, with the cause being one event and the effect another one. This commitment is sometimes explicit,[165] sometimes implicit,[166] and held by both counterfactual[167] and production[168] theories of causation. Davidson identifies this widespread conviction: "Much of what philosophers have said of causes and causal relations is intelligible only on the assumption (often enough implicit) that causes are individual events, and causal relations hold between events."[169] In this section, I question this general orthodoxy on three different grounds, all substantiated by a powers view of causation. Through these critiques, positive features of the powers view emerge.

Kim, Kroedel, and most causal theorists think of causation as occurring between two relata, which are commonly characterized as events. The powers view of causation reformulates this way of thinking. Instead of taking two entities and asking how they are related, a powers view of causation presents causation as a single entity—which might be a single event—constituted by operating powers. On this view, then, it is imperative to know how to characterize

[164] A note on terminology. Henceforward, "mutual manifestation" refers to: active + passive power(s); whereas "co-manifestation" refers to *either*: active + active power(s); *or*: passive + passive power(s).
[165] Kroedel (2019), Lowe (2008).
[166] Field (2003), Hall (2004).
[167] Lewis (1973a), (1979), (1986), (2000); Bernstien (2019).
[168] Kim (1998), (2005).
[169] Davidson (1967: 161).

this single entity. In this section, I argue that it is best conceived of as a *process*, for three reasons.

First, I argue that powers do much to account for the *simultaneity* of cause and effect between, at least some, causal relata. Second, I argue that the fact that causal happenings can be *interrupted* indicates that a process might better capture causation than a pair of events. Third, I argue that causation understood as a process of powers mutually manifesting allows us to avoid a bloated number of causes. Thus, the main conclusion from the following arguments is that causation is better understood as a *single process* than as the connection between *two events*.

3.3.1 Simultaneity

On standard views that rely on two events to explain causation, one might wonder when the causal event stops and the effectual event starts. Indeed, implicit in Kim and Kroedel's discussions in the previous chapters is a commitment to the view that causes occur before their effects. On this view, the cause must stop *before* the effect commences. In this section, I argue that cause and effect are simultaneous.

Many everyday examples of causation appear to elude easy explanation on the "received view" of causation. The broom pushes the dust, the cat catches the mouse, the wind hits the turbine's blades, the wave forces the surfer forward, and even, the first billiard ball hits the second.[170] In all these cases, it's unclear what causal event is supposed to predate the effectual event. Consider, in particular, the case in which a tablespoon of salt dissolves in a glass of water

[170] Mumford and Anjum (2011: 108) argue Hume's billiard ball example is really a case of simultaneous causation (focusing on the moment of contact). They write: "the problem is that while ever *a* is rolling along the metre of green baize, it has not yet met with ball *b*. How, then, can any causation, with respect to ball *b*, have yet occurred? Doesn't the causation occur at that place and time that balls *a* and *b* meet and not before?" On simultaneity, see also Aristotle *Physics* III 1 201a23, VII 1 242a57-62.

over the course of five minutes.[171] How many events are required to analyze the causation here occurring? Perhaps the first event is the event of the salt and water being mixed. This would be the cause. If so, then perhaps the second event (the effect) is the event of the salt and the water becoming a solution. But when exactly does the mixture of salt and water become a solution? Since the final state is only achieved after five minutes, one might reasonably wonder how we are to divide up the five minute period into events. Whether or not an event is supposed to be instantaneous or prolonged will affect our answer to the question.

Perhaps, we will need to analyze what takes place over the five minute period as the conjunction of a succession of snapshot event-causes, c_1, c_2… c_n, with a succession of snapshot event-effects, e_1, e_2… e_n. Then we can wonder how we should individuate events: How long should each event last? One second? Or, the length of time it takes a gram of sugar to dissolve? Further, there will be the question of which c_i should go with which e_i in the correct causal analysis. Should c_1 be paired with e_1, c_2 with e_2, and so on? Or should c_1 be paired with e_2, c_2 with e_3, and so on, to permit temporal ordering? I ask these questions not because I think there are sensible answers to them, but because they point to the difficulties of identifying which elements of the dissolution should count as discrete causes and effects. It seems unparsimonious to need a multitude of events to account for one continuous activity.

On the other hand, one might respond to these difficulties by saying that events are prolonged. On this tack, the event-cause and the event-effect might have some overlap, but they do not cover the same exact stretch of time. The event cause—the adding of the salt to the glass of water—takes more than an instant. Nevertheless, it is complete at, say, the second minute of the five minute period (the pouring takes awhile). The dissolution of the salt, on the other hand, does not begin immediately at the start of the five minute period. Rather, it begins at the first

[171] Example from Marmodoro (2017), Heil (2012: 118-122).

minute and ends only when all the salt is dissolved after five minutes. This approach also seems lacking, however, not least because it fails to keep the causal event before the effectual event—a stipulation to which most events-pair analyses are committed.[172] To see why, start with the part of the causal event that occurs at the second minute, call it $c_{@t2}$, and with the part of the effectual event that occurs at the first minute, call it $e_{@t1}$. Unfortunately for accounts that wish the causal event to pre-date the effectual event, since $e_{@t1}$ pre-dates $c_{@t2}$, there is a clear counterexample.

Since it is not plausible, on the given understanding of effectual and causal events, that the effect has not started before the completion of the cause, only one other possibility remains. Perhaps one might try to maintain temporal ordering by positing that both the relevant causal event and effectual event are limited by physical parts of the activity as described. Thus, perhaps the relevant causal event is the addition of one particle of sugar and the relevant effectual event this single particle's dissolution. Thus, though sugar particle no. 1 dissolves at t_1, sugar particle no. 355 only lands in the water and dissolves at t_5. We could, in other words, elaborate a different causal and effectual event for each sugar particle and thus maintain correct temporal ordering.

There are two types of response to this position. The first questions the validity of the thought experiment: when we seek a theory of causation, we want one that clarifies the macroscopic world. If a theory fails on this front, we should view it with skepticism. The theory's failure is not its falsification, of course, but it violates the desideratum of keeping appearances. But there is another response. Take sugar particle no. 1. The process of its dissolution is not instantaneous but requires a short stretch of time, say from t_1 to t_5. Now, let's grant that the causal event is happily prior to the effectual event(s).[173] Event individuation is again a question; in this case, for the effectual event. Is the effect one event running from t_1 to t_5?

[172] At least standardly within Humean accounts, setting aside retro causality.
[173] One might question this admission: perhaps the start of the effectual event (dissolution) is exactly cotemporaneous with the causal event (combination of sugar particle with water).

Or is it instead multiple sub-events, e_1, e_2… e_n, which together run from t_1 to t_5? In the latter case, the effectual event is constituted by a number of sub-events, e_1, e_2… e_5, each indexed to a time, t_1, t_2… t_5.

Let us use the second possibility since it is more fine-grained, while recognizing the conclusion to be drawn applies either way. Say that the causal event causes the sub-event e_1, because of temporal proximity. Now one might wonder how the cause can cause any sub-event occurring after t_2, once temporal proximity is lost. It might seem odd that I can *directly* fell the second domino in a chain without touching it any way.[174] But someone defending events as the proper relata of causation supersedes this worry by appealing to causal *transitivity*. On this response, the causal event causes e_1 which, in turn, causes e_2. But this seems highly implausible. How is a particle of sugar partially dissolving (e_1), supposed to *cause* the same particle to dissolve slightly further (e_2)? One way of seeing the difficulty of using transitivity as a solution in the current example is that the activity may well have started with a slightly smaller particle of sugar at t_2 such that e_2 would look identical, though e_1 did not occur in the adjusted example. Then it's no longer the case that e_2 causally depends on e_1. A similar overall response can be run against an objection using smaller sub-particles of sugar, because any physical quantity will take *some* amount of time to dissolve, no matter how microscopic.

A powers view of causation, however, can easily explain the simultaneity of such everyday examples of causation. It does so, however, *without relying on events as the fundamental causal relata*. Rather, it analyzes causation by understanding it as a *causal process constituted by mutually manifesting powers.* Anna Marmodoro provides another helpful example of simultaneity in causation:

[174] See Heil (2012: 123) for a similar worry.

> The reciprocity [of two powers] may be illustrated for example with a causal scenario where A causes B to become hotter. A's power to heat is activated; it manifests by heating B up. But A's heating up B can take place only if B's power to be heated is activated too, and manifests itself in B's getting hotter. A's heating and B's being heated are mutually dependent in a variety of ways, and each activity lasts as long as the other lasts.[175]

In the above example, A and B do not represent events but rather powers. B represents the power of water, for example, to be heated up. On the other hand, A represents, for example, the power of the stovetop to heat up the water. Setting aside possible confounding factors, so long as B is heating up, A is doing the heating. Similarly, so long as A is heating, it will be heating up B.

The aspect of this passage I would like to focus on is the final point that cause and effect persist for the same length of time. Marmodoro concludes the passage by endorsing the simultaneity of the active and the passive powers during causation: "each activity lasts as long as the other lasts." Importantly, however, she does not thereby sacrifice one kind of asymmetry between the cause and the effect. As Marmodoro puts it, "My account of causal agency is an alternative way, to the temporal one, of underpinning causal direction metaphysically."[176] Nevertheless, her view that causation always involves an active and a passive power doesn't imply that a power is *essentially* active or passive. Thus, the fact that something will (passively) cool down by means of another agent power also usually means that the same thing has a power to (actively) heat up something else, by means of that heat loss. Thus, her view establishes non-temporal *asymmetry* between powers causal relata, without essentializing active and passive powers.

But this view needs qualification. Consider the cases in which two playing cards lean on one another or two identical projectiles collide.[177] Presumably, such a collision of projectiles can

[175] Marmodoro (2017: 58).
[176] Marmodoro (2017: 72).
[177] As mentioned, Mumford and Anjum call this "equilibrium."

happen with such perfect symmetry that characterizing the active and passive power would be difficult. These cases illustrate that sometimes the distinction between active and passive powers is not immediately evident. Both causal relata manifest a causal power on the other, but with perfect symmetry. Importantly, then, for the powers theorist, no change needs to occur for causation to occur. Thus, though I use the language of active and passive powers, note that this language characterizes most but not all cases of causation. One strategy for making symmetric cases closer to standard ones is to say that one relatum can be both active and passive, though in different respects. For instance: Projectile_1 actively impacts Projectile_2 but at the same time is passively impacted by Projectile_2.

On a powers view of causation, two separate mutually manifesting powers operate simultaneously, so there is no need to posit a pair of temporally ordered events. Returning to the salt dissolution example, such an analysis seems to make much better sense of what is going on than an events-pair analysis. Instead of two temporally ordered events succeeding each other, we can now say that throughout the whole activity, two powers are present. One is the power of the water to dissolve (the active power). And the other is the power of the salt to be dissolved (the passive power). As long as the activity endures, both cause and effect mutually manifest, otherwise causation would cease. I now turn to the question of how best to characterize the "causal activity."

3.3.2 Interruptibility

If causation is not best conceived as the relationship between two events, how ought we to conceive of it? In a paper comparing events and processes, Rowland Stout writes, "I am quite attracted to the Aristotelian idea of identifying a *process* in terms of an underlying capacity for

change."[178] Similarly, near the conclusion of a recent paper, Helen Steward responds to the self-imposed question: what is the significance of a distinction between processes and events? She writes: "Questions both of causation and of explanation look rather different when one is thinking about processes from the way they look when one is thinking in terms of events—and these differences are I believe, crucially significant."[179] She continues, "it is natural to think of the cause of an event simply as another event, or perhaps a collection of pre-existing events and states which give being to the caused event simply by triggering it off." [180] But we might, she says, instead think of the cause of an effect as something formal, with a normal course of development, which can gain extra temporal parts without losing its identity. While I will not delve into all the sophisticated arguments that Steward marshals in defense of her "individual process" ontology, it is interesting to note the payoff she believes her arguments to have.[181] In this section, I present one strong reason for favoring a "processual" view of causation rather than an events-pair one. A key reason for endorsing this shift in mentality is the fact that processes, unlike a pair of events, can be interrupted without eliminating *some* causal happening.

Recent work has scrutinized and defended processes. In an influential article, Alexander Mourelatos builds on the work of Kenny and Vendler to give a grammatical analysis that he thinks ultimately tracks an underlying ontology.[182] He favors an event-process-state ontology, that distinguishes events from processes/states. He reaches this conclusion based, firstly, on the distinction between perfective and imperfective uses of verbs, the former indicating events, the

---

[178] Stout (1997: 26). My emphasis. *Metaphysics* Θ 6 distinguishes between *kinêsis* (movement) and *energeia* (activity) which might both be characterized as processes, one continuous and one completed. Barnes's (1984) translation favors this view at 1048b34. Also, Coope (forthcoming).
[179] Steward (2013: 809-810).
[180] Ibid., 810.
[181] I do not parse Steward's critiques of Stout's view of processes, for two reasons. First, her "individual processes" might have been termed "processes." Then her view would qualify pre-existing views and not perhaps needlessly introduce a new ontological type. Second, her arguments compare a single process and a *single* event, whereas mine compare a process with an *events-pair*.
[182] Mourelatos (1978).

latter processes/states. Second, he thinks that when a sentence is given a nominalization transformation, events will be classed as count nouns and processes/states as mass nouns.[183] On his view, "Water dissolves salt" is nominalized as "There is dissolving of salt by the water." He proposes that the absence of a definite or indefinite article before "dissolving" in the nominalization transformation indicates that processes are best characterized by mass, not count nouns.

Helen Steward, responding to the work of Rowland Stout, builds upon Mourelatos's theory to ultimately endorse a distinct ontological category she calls "individual processes." She argues that the description of an event is necessarily countable (like a count noun), whereas the description of an individual process is something that can be augmented without losing its identity (like a mass noun). Even if not decisive, these views offer some reasons to admit processes into a complete ontology. If so, then events are not the only ontological category we have for causal analyses.

Take again the example of salt and water from the previous section but imagine now that dissolution is not completed. After the water dissolves exactly half of the sugar, the remaining quantity is rapidly strained out of the solution. The causation is thus *interrupted*.[184] Has causation failed to take place? Since *some* of the sugar has dissolved, it would be odd to insist that causation has wholly failed. Rather, it seems that *some* causation has taken place, just not all that might have. On the original events-pair analysis of the example, the causal event is the combination of sugar and water, while the dissolution of sugar in water is the effectual event. In the modified example, the causal event so described is untainted, but the effectual event so

[183] See Mourelatos (1978: 424-431) for details.
[184] Aristotle similarly suggests that natural things exhibit tendencies that are realized "if there is no impediment" *Physics* II 8, 199b18. Also, *Metaphysics* IX 5 1048a17-20, IX 7 1049a5-12.

described *doesn't take place*.[185] As a result, it's not possible to spell out the relationship between the two original events relata, since one no longer exists.

Countenancing *interruption* gives another serious reason for questioning the reliability of an events-pair analysis of causation. In the given example, some sugar has dissolved, just not the original amount intended. The interruption eliminates one of the events that served as a causal relatum and thus eliminates one essential element of the events-pair analysis of causation. A process, on the other hand, does not lose its identity when interrupted. Thus, even if a process is incomplete, we can understand how *some causation* still took place in that part of the process that was completed.

One might here object that in the new case, we can simply redescribe the two events that serve as the causal relata. Thus, we can keep the causal event identical to how it was in the original example, while supposing that the new effectual event is the dissolution of a smaller quantity of sugar. Then, we could offer a fresh account of the relationship between these two new events. There are two worries with this reply. First, it fails to identify a strong identity between the original example of dissolution and the new example of dissolution with half as much sugar.[186] Though similar, we are nevertheless dealing with two distinct effectual events and thus two different analyses of the relationship between the causal event and effectual event. A processual view of both examples, however, permits identity to be *the very same* between the interrupted process and the original one until the point of interruption. In Steward's words, "we may think…of a process as a thing which can be made larger by the addition of further parts of the same sort."[187] Second, even if an events-pair analysis redescribed the process with a new

[185] An events-pair apologist might respond that his analysis *is* still possible on the micro-physical level. But it seems from my perspective that this micro-physical events-pair story could also be interrupted, so long as time elapses between the event-cause and event-effect.
[186] Appealing to a type-token distinction does not quite do the trick: though both the new and original occurrences are dissolution-types, they also share some token features (i.e. the first part of the process).
[187] Steward (2013: 805)

effectual event, one would have to worry about a regress. For if the original dissolution becomes a partially new instance of causation if interrupted at the half-way point, the interrupted causal occurrence might itself be interrupted at a new half-way point. Then, once again, this second modification would have need of yet another distinct effectual event, and thus of a third analysis of the relationship between the causal event and the new effectual event. That each of these causal relations needs novel—if similar—analyses on an events-pair view doesn't explain the identity of the process until the moment of interruption. But this outcome is exactly what a process ontology would expect.

Returning to the idea of causation as the mutual manifestation of active and passive powers, a process is more than an events-pair. Inasmuch as we analyze the case of water dissolving salt as the process of an active and passive power mutually manifesting, we don't need a prior and posterior event. Rather, the active power (water's power to dissolve) acts on the passive power (salt's power to be dissolved) in a single process.

Further, an instance of causation, as explained above, is the kind of entity that can be interrupted. Nevertheless, even if causation were terminated at t_3, we don't want to say that no causation occurred. If the causal relata are two events, we would be forced to draw this conclusion or to redescribe the situation. If, however, the relata are an active and passive power mutually manifesting in a process, another kind of answer is available. We can say rather that causation indeed took place from t_1 to t_3, but ceased when interrupted.

It must be noted that someone could respond by insisting that the whole occurrence of mutually manifesting powers might be itself categorized as a single event. This idea is not totally inimical to the powers of view of causation here presented. I could still maintain that two powers, an active and passive one, cooperate in causation, but that they constitute a single event,

not two temporally separated ones. And it is only the latter claim that I wish to contest. Admittedly, the ontological distinction between events and processes can be somewhat murky. Thus, if an opponent finds the above reasons for stating that the causal occurrence is a process unconvincing, I can allow that the whole activity be categorized as a *single* event without compromising the essential improvement of the powers view: the relevant relata of causation are temporally indistinct asymmetric active and passive powers.

Thus, a powers view of causation more easily analyzes causation in terms of processes than events, though this provisional conclusion need not banish events altogether from one's ontology. Consider the following statement from Mumford and Anjum: "Causation often, though it need not always, involves a change. When it does so, there is an event."[188] This conciliatory tone is mirrored by Lowe, who suggests that though his "substance causation"—similar to powers causation—is more fundamental than "events causation," the latter may still "be seen as capturing conceptual truths."[189] I support this sort of strategy that doesn't prevent events analyses from supervening on more fundamental powers analyses. Thus, though I prefer to call causation a process, events need not be eliminated wholesale.

3.3.3 Proliferation of Conditions: How Powers Narrow the Relevant Causes

In an oft-quoted passage, J.S. Mill suggests that the actual cause of an effect is the sum total of all the pre-existing conditions.[190] In contemporary terminology, one might articulate this

[188] Mumford and Anjum (2011: 23).
[189] Lowe (2008: 145).
[190] Discussed by Davidson (1967: 150), Lewis (1986: 216), Mumford and Anjum (2011: 11). It was suggested to me that a cross-section of the light-cone from the frame of reference (a Cauchy surface) plus laws is all you actually need to determine the effect on an events-pair analysis, because this cross-section would contain all relevant preceding events.

notion of a *total cause* as every event in the backwards light cone of the effect.[191] On a powers view of causation, we already have the necessary resources to counteract this overly-expansive view of causation. In short, by identifying the relevant active and passive powers, mere preexisting conditions are properly contextualized in their non-causal role.

One might worry that a Millian conception of causation is too broad, leading us to count an overabundance of causes. In response to this worry, Mumford and Anjum distinguish between a genuine cause and a mere "condition" (e.g. the sun's *illumination* permitted my shooting the apple off his head). Thus, on their view, only powers that dispose towards a specific outcome count as causal; every other factor is simply a *sine qua non* condition.[192] And properties that have powers irrelevant to the outcome may be ignored.

To illustrate this point, let us return briefly to the Opacity Example from the previous chapter. In that example, not only opacity, but rigidity, length, and width were all conceivably (on a counterfactual view) causes of electrocution by way of the conductive ladder. In fact, the instantiation of any quality of the ladder whatsoever could serve as a possible cause of the effect. But using powers, it is possible to restrict the scope of causally relevant entities. Thus, if I have the power to be electrocuted, it isn't in virtue of *any* power possessed by the ladder that my passive power manifests. Rather, it will only be by means of the ladder's very particular active power of *conductivity* that the effect is activated. While the ladder has myriads of powers that could manifest in different causal circumstances, only the active power of conductivity is relevant to activating my passive power for electrocution.

Several qualifications are needed. First, despite the simplicity of the foregoing example, that by no means requires that we restrict causation to a single active and single passive power.

---

[191] Mumford and Anjum (2011: 13).
[192] Ibid., 173. Gibb (2015: 79) distinguishes between genuine causes and "enablers," but her account still uses an events ontology. Also, Heil (2012: 123-124); Mayr (2017: 83-84).

As with Mumford and Anjum's *compositional pluralism* explained above, most examples of causation will involve the co-manifestation of many powers, active and passive. Many powers go into propelling a rocket booster, for example. Minimally, the passive powers require the active powers of spark plugs, fuel propellant, and pressure chambers all co-manifesting. But just as those authors expand vector modelling to include more complex cases with non-linear functions, so too we can acknowledge that often multiple factors are needed to correctly encapsulate what produces a given outcome. As Cartwright urged, effect and outcome must be disambiguated: the first is the passive power(s) and one of the causal relata, the outcome is the state of affairs at the causal process's conclusion.

A powers view of causation comes with many benefits. In this chapter, I have outlined a powers theory of causation—motivated by Cartwright's challenge—that is elaborated in terms of asymmetric mutually manifesting active and passive powers. I have argued that we should reconceptualize causation as a single process. When deployed with care, it accounts for the simultaneity of cause and effect, the interruptibility of a causal process, and narrows Mill's over-expansive total cause. And it may help solve the difficulties inherent in mental causation.

4. A Powers Picture of Mental Causation

> *"Our concern is with a restricted class of dispositional terms, namely those appropriate only to the characterization of human beings. Indeed, the class we are concerned with is narrower than that, since we are concerned only with those which are appropriate to a characterization of such stretches of human behaviour as exhibit qualities of intellect and character" – Gilbert Ryle (2000 [1949]: 121).*

The final task that remains is to apply the dispositional theory of causation developed in Chapter 3 to mental causation. The relevant class of dispositions is the right kind of human activity. Thus, I do not discuss manifesting non-human dispositions, nor manifesting human dispositions that fail to exhibit activity characteristic of rational agency. Thus, I will not, for example, discuss the causal processes involved in instinctual responses.

I will proceed in two main stages. First, I argue that from powers perspective on causation we can now more substantially respond to the difficulties that faced Kim and Kroedel than we could in Chapters 1 and 2. I argue, on the one hand, that powers explain why Kim's *Overdetermination* principle is out of place in mental causation arguments. On the other hand, I argue that a powers view of mental causation also explains why Kroedel depended on *additional stipulations*. An important payoff is a reply to counterarguments that depend on a "double-prevention" view of mental causation. Second, after this largely negative project, I conclude by mentioning several key theses for mental causation within the powers causation paradigm.

4.1 Contra Kim: *Exclusion* Revisited

Recall that Kim based his Master Argument for reduction on five principles, two of which I challenged. A powers view of causation gives a rationale for rejecting *Exclusion*. Once causation is understood as a constitutively pluralistic process, involving mutually manifesting active and passive powers, the principle becomes a non-starter. *Exclusion*, in addition to exhibiting the difficulties noted in Chapter 1, also fails to acknowledge the ubiquitous and standard co-manifestation of active causal powers in causation.

We can now reexamine Kim's principle on a new footing. Recall his articulation:

> *Exclusion.* No single event can have more than one sufficient cause occurring at any given time—unless it is a genuine case of causal overdetermination.

We can, in particular, question, qualify, or reject what Kim means by "event," "sufficient cause," and "causal overdetermination" from a powers perspective.

I begin with events. Since we have good reasons, based on the simultaneity and the possibility of interruption, for questioning a two-event analysis of causation, we can now reject

Kim's view that the effect of causation is a single event separate from the causing event. I do not reject the idea of causal relata wholesale of course. Rather, a powers view of causation simply rearticulates what the causal relata are, as schematized in the following diagram:

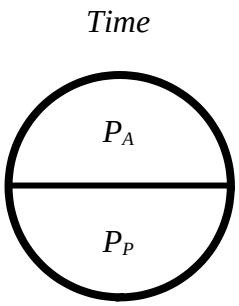

Figure 4.3: The Powers Relata of Causation

The relata, according to my version of powers causation, are an *active power* (P_A) and a *passive power* (P_P). The active power (like the power of water to dissolve) exercises its potency on a passive power (like the power of salt to be dissolved). Notice, however, that the relata on a powers view of causation are synchronically and not diachronically related to each other. Thus, so long as the active power manifests in the way relevant to the activation of the passive power, the passive power too manifests. Of course, the two manifesting powers *together* constitute a diachronic process, hence the *time* arrow included in the diagram. Thus, the relata on the new analysis are simply mutually manifesting powers, which are not temporally separated, even if together they are temporally extended.

The lesson from this view urged in Chapter 3 is that it doesn't make good sense to think of causation in terms of an events-pair. A pair of two events had difficulties explaining the simultaneity and interruptibility of causation. Further, even on the powers view of causation, we prefer labelling the whole a process rather than an event. Causation, on the new mutually manifesting powers analysis, might still be labelled as one event, but not as an events-pair.

Note that someone might wish to say that the *outcome* of the whole process is an event effect. In this sense, the salt's dissolving would not be the effect, but rather the completion of the process of dissolution.[193] This position is reasonable on the events-pair view: an outcome-event, called an "effect," depends on a causal-event. But I side with Cartwright in making a crucial distinction between a power's (1) *manifesting* and (2) *the manifest results*.[194] Causation governs how a power gets exercised, but the outcome is something else altogether. Once all the salt has dissolved in the water, the saline solution, for example, is an outcome but not an effect.[195] From a powers view, Kim would seem to falsely call an effect an event. If the second relata of causation is a manifesting passive power, it cannot be an event.

Next, what does Kim mean by "sufficient cause"? Recall first that a cause can be an aggregate of co-manifesting active powers. Thus, though phrased in the singular, many cases of causation (like the causal aggregate responsible for propelling a rocket booster) are the co-manifestation of a set of active powers on a set of passive powers.[196] In other words, to properly explicate a causal process such as a rocket's liftoff, we need many powers. First, we must understand the passive propensity of the oxidizer and propellant when combined and combusted to expand rapidly. Second, we need to understand the active powers of a sparkplug to ignite suitable material and of the propellant tanks to pressurize the reactants. Further component active and passive powers could be catalogued: tiny particles of dust might invade the rocket fuel or extra sparks aid ignition. Intimate knowledge of rocket boosters could produce a long list of

---

[193] Heil (2012: 120) considers such an objection. He allows that an outcome can be an effect but argues that saying so does not undermine the ontological priority manifesting powers constituting a "causing."
[194] Cartwright (2009: 151-152).
[195] A rumbling objection: the Humean might say that the *commencement* of an outcome is an event effect. But I would respond by saying that a static existent (e.g. saline solution) cannot be a stage of the causal process.
[196] Kim seems to think that the relevant set of physical realizers of a mental property is singular. Not all his interpreters phrase the view similarly, however. See Block (2003: 140): "The Exclusion Principle (that causally sufficient *properties* at one level exclude causally sufficient *properties* at another level) is to blame and should be rejected." My emphasis.

powers, active and passive, so this case of causation, like most, consists of a multitude of powers. A workable concept of "sufficient causation," then, must include multiple factors. Despite not emphasizing plurality, Kim might nevertheless be sympathetic to this idea, amending his "sufficient cause" to include plural constituents (setting aside momentarily the fact that Kim thinks the cause is an event).[197]

A primary implication of the fact that a sufficient cause can be composite is that there is no reason to preclude the possibility that both mental and physical components compose the cause. In mental causation, a pluralistically composed sufficient cause could without contradiction have both mental and physical parts. There might be good reasons, based on the interaction problem, for supposing that mental powers cannot contribute to the sufficient cause of bodily motion. But those reasons, if they exist, are not predicated upon the notion of sufficient causation *simpliciter*.

Turning to the last part of Kim's *Exclusion* principle, we can now reject his use of overdetermination on firmer ground. Recall that, in Chapter 1, I distinguished between two types of overdetermination. *Mental Overdetermination* with indistinct and dependent constituent causes was not too worrisome, but the more fearsome *Prototype Overdetermination* with distinct and independent causes can now be recognized as impossible. Recall its formulation:

> *Prototype Overdetermination*: A single event *can* have more than one sufficient—but distinct and independent—causes occurring at any given time when there is a case of genuine causal overdetermination.

Again, bracketing Kim's use of "event" in light of a powers view, consider what could be meant by an outcome having multiple *sufficient* causes. The active powers of two rocks to cause the pane of glass to manifest fragility does not imply that either is sufficient, even if distinct and independent. If Bill and Suzy, unbeknownst to each other, hurl their rocks at the window

[197] Kim discusses the possibility of composite sufficient causes in (1989).

simultaneously from opposite sides, the breaking that occurs is obviously different than the one that would have occurred by means of one rock only.[198] Of course, one rock alone could have sufficiently caused the window to manifest fragility, but that manifestation would have been different from that caused by both rocks. On the powers view, it is unclear that "genuine causal overdetermination" of either the *Prototype* or *Mental* variety can occur.

Perhaps a counterfactual approach can rescue overdetermination: a similar outcome *would have* obtained through the operation of one active power instead of through several. But then the outcomes would differ, on my view, since causation is constituted by the collectively mutually manifesting powers. These considerations show that the counterfactual formulation —"*If* X had obtained, Y *would have* occurred"—is not fundamental on a powers view. Nevertheless, a powers view maintains that though there be a sufficient cause, part of that sufficient cause on its own *could have* sufficiently caused a *very similar* outcome. Yet, the truth of the counterfactual does not mean that overdetermination ever actually occurs. For the cooperation of two *hypothetically* sufficient causes is still a single sufficient cause. Perhaps the counterfactualist could respond by adopting my approach and individuating outcomes by actually occurring origins. That strategy might work but requires appealing to a further principle and is thus not intrinsic to a counterfactual causal analysis. In any case, my main target is Kim.[199] On my powers view, overdetermination is mythical.

One way to see that several individually *hypothetically* sufficient active powers do not cause overdeterminately is to map out the various possible relationships between active and passive power(s). This case is one among several possible permutations of the causal relata:

---

[198] My idea is similar to Davidson's thought about two killings in "Thinking Causes." He writes, "Had the gun been equipped with a silencer, a quiet shot, if aimed as the fatal shot was, and otherwise relevantly similar, would no doubt have resulted in *a* death. But it would not have been the *same* shot as the fatal shot [without a silencer], nor could the death it caused have been the same death" (1993: 200).

[199] When discussing overdetermination in one place, Kim phrases his formulation counterfactually (1989: 91). Lewis is happy to depend on causal history for *causal explanations* (1986).

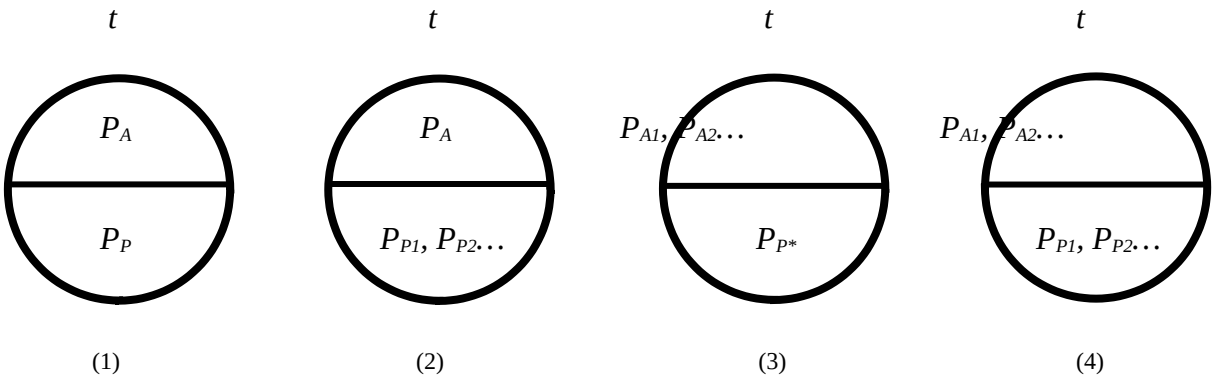

Figure 4.2: Possible Permutations of Powers Causal Relata

As in the previous diagram, in any case of causation, the composite process manifests diachronically. Now, in some cases a single power, P_A or P_P, has been replaced with a set of manifesting powers, (P_{A1}, P_{A2}…) or (P_{P1}, P_{P2}…). In (1), a single manifesting active power causes a single passive power to manifest—the water dissolves the salt. In (2), one active power causes co-manifesting passive powers—the fire burns wood and coal. In (3), co-manifesting active powers cause one passive power to manifest—two rocks break one window. In (4), multiple active powers co-manifest to cause multiple passive powers to co-manifest—sparkplug, astronaut, and the shape of the fuel tank all collaborate to cause propellant, binding agents, and oxidizer to react and expand, creating propulsion. In real life, we would expect cases like (4) to be ubiquitous.

In considering possible overdetermination, cases like (3) are particularly pertinent. Imagine, for a moment, that the passive power manifested in a case like (1) and in a case like (3) are the same. Say, for example, that the sun *or* several UV radiation lamps cause someone's sunburn. Further suppose that each lamp *would have been sufficient* to produce the effect on its own. In our hypothetical scenario, the sunburn caused is identical in both cases: the distribution of exposure, the likelihood of developing melanoma, and even the localized areas of discomfort

are identical in each case. If the sun is sufficient for producing the effect (= the manifesting of certain passive powers), surely the UV lamps overdetermine the effect?

Several responses are possible. First, as argued, the "sufficient cause" can be plural. Though multiple elements constitute the cause in (3), none of them (P_{A1}, P_{A2}…) is *sufficient on its own in this particular case of causation* for causing the passive power to manifest. Even if P_{A1} in separate circumstances, away from all its active power-partners *would be* sufficient, it is not sufficient when part of the composite active cause, (P_{A1}, P_{A2}…). Second, since the entire causal processes of (1) and (3) are different—now requiring the co-manifestation of a collection of active powers—the effect is in fact different. The effect is distinct, despite empirical equivalence, because a process of causation is composed by the unity of all its co-manifesting active and passive powers. The "*" next to P_P in (3) represents that the effect is a different token effect, even if empirically equivalent to P_p in (1).

Overdetermination, then, is impossible. Either the sufficient cause is singular, or it is pluralistically composed, in neither of which cases there is overdetermination. A counterfactual rescue also fails, since even if part of the pluralistically composed sufficient cause *would have brought about the same effect* the token effects are in fact distinct.[200]

4.2 Contra Kroedel: Double-Prevention and Transference Views

The powers view also has the resources to quell a potentially dangerous attack from Kroedel, not yet discussed. Recall that he gives a counterfactual sufficient condition for causation. His approach ultimately includes cases of causation that a powers view would not.

[200] Kim and I both wish to avoid overdetermination, but while I use powers to escape the problem, he urges reductivism. Nevertheless, I agree with him that "counterfactual accounts of causation continue to have difficulties with preemption and overdetermination, showing, in my opinion, that our core idea of causation is more intimately tied to generative/productive causation than to counterfactual dependence" (2005: 18 n. 12).

Kroedel argues, in particular, that cases of *double-prevention* are genuinely causal and that mental causation itself, because of a certain neuro-molecular mechanism, demands such cases to be causally efficacious. His stance on double-prevention thus bears weight for his argument, while also giving him fodder for an attack on powers causation. In this section, I argue that a powers view need not succumb to his attack.

A powers analysis can yield a more satisfactory reading of double-prevention cases than its counterfactual alternatives. Kroedel explains that cases of double-prevention "all have the following structure: something happens that would have been prevented by something else, which is itself prevented."[201] For example, if a rack supports a pillar, my removal of the rack is responsible for the pillar falling. I have prevented the preventer. Defending the genuine causation of double-prevention, he writes, "According to our counterfactual principle about causation, cases of double prevention are cases of causation."[202] The outcome counterfactually depends upon the removal of the preventer, which is sufficient for causation.

Kroedel's view on double-prevention has crucial implications for his overall argument. At the end of his second chapter, he argues that a particular neuro-molecular mechanism of muscle contraction uses double-prevention.[203] In brief, Kroedel claims that empirical facts about human physiology require believing double-prevention is causation—at the risk of undermining human agency. What are these facts? He writes,

> They [the myosin heads] would bind to actin filaments, move forward and thus make the muscle contract if it weren't for the obstruction of binding sites by tropomyosin molecules. If the muscle receives a nerve signal, calcium is released at the neuromuscular junction, which causes the tropomyosin to move away from the binding sites… an event (here the calcium release) prevents something from happening (the obstruction of the

[201] Kroedel (2019: 49-50).
[202] Ibid., 55.
[203] Ibid., 92-96.

> binding sites) which, unless prevented, prevents the another event (the muscle contraction) from happening [*sic*].[204]

The moral Kroedel draws is that muscle contraction—surely the relevant sort of effect of mental causation—comes about via double-prevention. Now, his argument goes, any reasonable view accepts that to preserve human agency, we must employ the available means of causation. The available means is a case of double-prevention. Therefore, we must accept that double-prevention is genuinely causal.

A first minor reservation is that the empirical views countenanced by Schaffer and Kroedel might turn out to be false. Perhaps, for example, there is an intervening step, not yet discovered, between the calcium release and the tropomyosin removal that makes the process look much more like a standard causal chain. If so, it would be unfortunate to ground mental causation on a falsifiable empirical detail.

A much stronger worry is that Kroedel must argue that double-prevention is the *only* available means of causation at hand. Perhaps there are *other* means, besides double-prevention, to explain some bodily movement. In that case, he could not maintain that *all* mental causation depends upon double-prevention. As a possible counterexample: take the inverse case where bodily movement comes about by muscle relaxation, not contraction. Imagine that I am lifting a heavy weight overhead. To keep the weight overhead, I actively *choose* (a mental cause) to keep certain muscles—biceps, or whatever—contracted. At some point I relax my arms and lower the bar. Did this bodily movement come about by double-prevention? Even on Kroedel's physiological picture, it is unclear that it did. All the time my muscles are contracting, a nerve signal causes calcium release at the neuromuscular junction, which likewise ensures that the

[204] Ibid., 92. He depends partially on Jonathan Schaffer, "Causation by Disconnection" and "Causes Need Not Be Physically Connected to Their Effects: The Case for Negative Causation." The details are simplified more than the anatomist would desire.

tropomyosin stays away from the binding sites, so that myosin and actin can bind, which binding constitutes the contraction of the muscle fiber. When the nerve signal causing contraction ceases because I *decide* (a mental cause) to lower the bar, calcium release at the junction stops and tropomyosin re-bonds to the binding sites. Then the myosin and actin stop binding, so the muscle fiber relaxes. I have caused my muscles to relax by mental causation. We can thus question whether Kroedel's position actually requires that every case of bodily movement depends on the highlighted mechanism. Therefore, Kroedel's conclusion, "it is an empirical fact that in humans bodily movements are caused by double prevention"[205] appears stronger than desirable.

But in a revealing counter-move, Kroedel might suggest that I have simply avoided double-prevention by choosing a specific description of the phenomena. In fact, when I lower the bar, it's because the absence of a nerve signal *prevents* calcium release that tropomyosin can return to the bonding sites and *prevent* actin and myosin from binding. But if a redescription is possible at all, a deeper worry emerges: perhaps double-prevention only arises as a specific description of otherwise standard causal processes. I turn to this more substantial critique from a powers perspective in the next section.

Kroedel, however, has reasons for resisting a powers view of causation, which he conceives of in a very particular way. He places powers causation among a broader class he labels "transference views." He writes, "A rival [to a counterfactual] view of causation requires that a certain physical quantity be transferred from cause to effect."[206] He further clarifies, "Let us call theories of causation according to which causation requires transfer of causal powers *powers transference views*."[207] For Kroedel, transference views—including ones that involve powers—require that a physical entity is handed over from cause to effect. While certain cases

[205] Ibid., 96.
[206] Ibid., 49.
[207] Ibid., 58.

seem to support this view, it fails from his perspective to accommodate cases of causation like double-prevention. When the wooden rack is removed and the pillar falls, Kroedel's version of transference views states that no causation takes place, since no physical entity is transferred from rack to pillar.[208] Since the pillar's falling counterfactually depends on the rack's removal, Kroedel says his condition fares better. As mentioned, a failure of the transference view has important implications for mental causation for Kroedel.[209] He thinks counterfactual accounts are superior to transference views because they can explain causation when nothing is transferred.[210]

But Kroedel mischaracterizes at least Mumford and Anjum's view. Powers theorists can consistently deny that cases of double-prevention represent causation *qua* an omission[211] (i.e. the rack's absence). But it is unfortunate that these cases are his preferred counterargument, since powers theorists need not deny all causation in double-prevention. Though a powers theorist can't analyze double-prevention as the counterfactualist does, surely causal powers manifest. Thus, the very powers theorists whom Kroedel engages, Mumford and Anjum, have a preliminary response to the accusation that the pillar falling isn't causation.[212] They say that the exercise of causal powers explains the state of affairs prior to the removal of the rack. Since the rack exercises its causal powers to keep the pillar supported, removing the rack also removes manifesting causal powers. Since double-prevention is intelligible on the powers view, Kroedel too quickly dismisses his opponent.

4.2.1 Double-Prevention Re-Analyzed

---

[208] Ibid., 55.
[209] Ibid., 92-96.
[210] Ibid., 49-59.
[211] Bernstein discusses omissions as causes (2019: 167-168). Heil (2012: 126-130) gives an overview of Lewis's treatment of the "void" as causally efficacious. Also, Paul and Hall (2013: 176-182).
[212] Mumford and Anjum (2009: 288-289).

Even granting Kroedel's empirical claims, however, it's possible to offer a more sophisticated analysis of the phenomenon he describes. In fact, a powers view of causation can rigorously explain double-prevention in muscle contraction by means of mental causes. In other words, there is another possible analysis of the above empirical details that doesn't make double-prevention causal *in the way* counterfactualists assert. Because using the empirical details of muscle contraction introduces complications unnecessary to make the philosophical point, I stick with the simpler rack and pillar example. Further, this method shows more generally that double-prevention shouldn't worry the powers theorist. I begin by labelling and separating the relevant active and passive powers, at least in a rough manner. For the rack and pillar example, we might produce the following chart:

Active Powers	**Passive Powers**
Active power of rack to support objects	Passive power of pillar to be supported
Active power of person to remove rack	Passive power of the rack to be removed
Active power of gravity to attract objects	Passive power of the pillar to be moved by gravity
Active power of earth to support objects	Passive power of objects to be supported (by the earth)
All negligible active powers exercised by objects in the setup	*All negligible passive powers exercised by objects in the setup*

Figure 4.3: Active and Passive Powers in Double-Prevention

The chart inventories salient powers in our example of causation. Note that I focus on an idealized case, leaving aside the italicized elements in the sixth row of the chart, which represent the *negligible* contributing powers, like friction, air resistance, and the like. With this caveat in place, I formulate a diagram of the double-prevention case stylized like those foregoing:

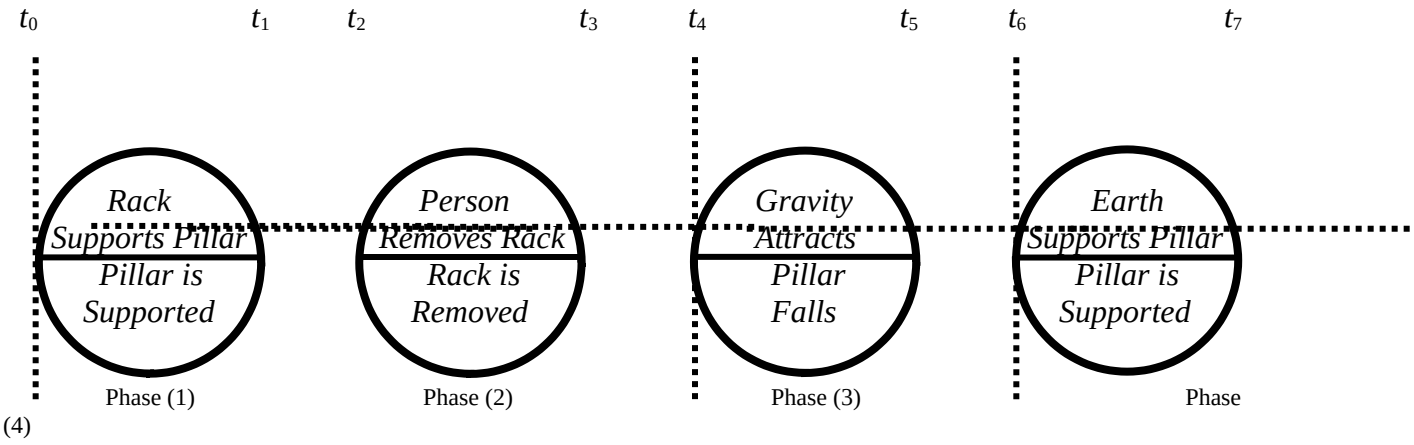

Figure 4.4: The Phases Constituting a Case of Double-Prevention

First, notice that my powers view analyzes the case of double-prevention as several "phases" of causation. Thus, from a powers perspective, there are several causal processes that go into a single double-prevention case. These constituent processes run, respectively, from t_0 to t_1, from t_2 to t_3, from t_4 to t_5, and from t_6 to t_7. Second, notice that each of these phases has its own active and passive causal powers. For the sake of simplicity, I label only one for each, though a more thorough representation could represent co-manifestation of multiple powers above and below the hemispheric lines. Third, the diagram permits inter-causal temporal gaps that might elapse between one phase and the next, indicated from t_1 to t_2, from t_3 to t_4, and from t_5 to t_6.

So what is wrong with Kroedel's analysis of double-prevention? It's not categorically false—quite plausibly the outcome depends on earlier phases in most cases of double-prevention. Rather, the powers view has unveiled a deeper interpretation of what is taking place. Powers co-

manifest in several distinct phases that each contributes to the later ones and the final outcome. I am thus sympathetic to Martin's comment that "'Causal' counterfactuals have a place, of course, but only as clumsy and inexact linguistic gestures to dispositions and they should be kept in that place."[213] On the powers view, causation is not dependent on a counterfactual analysis of double-prevention.

Turning to Kroedel's specific claim about double-prevention's role in mental causation, I now have a direct reply. Rather than possessing causality in virtue of a counterfactual analysis, cases of double-prevention are causal in virtue of constituent processes of active and passive powers. The outcome predictably follows the co-manifestation of constituent active and passive causal powers. As a final reply, recall Kroedel's fourth premise from the argument in Chapter 2:

(4) If *P** is instantiated later than *M*, and *P** would not have been instantiated if *M* had not been instantiated, then the instance of *M* causes the instance of *P**.

Fortunately for us, the consequent of this counterfactual is true. Yet crucially, it is not true *in virtue of the antecedent*. On a powers view, there are always exceptions to the most carefully crafted counterfactual. And we can save ourselves the trouble of endless *ceteris paribus* clauses if we turn instead to the underlying causal explainers—the dispositions, active and passive, that make up the causal process.

4.3 Powers in Mental Causation

A fortunate feature of active causal powers is that they must, on my view, always be paired with passive causal powers.[214] In real-life causation, the active power exercises only if it has a passive power on which to act, and vice versa. Also, in most real-life cases, however,

[213] Martin (1994: 8).
[214] Though in symmetrical cases I allow it is hard to identify which is which.

multiple active and passive powers co-manifest. These general principles provide enough background against which to state some constructive abductive conclusions about mental causation. In other words, if we observe the manifestation of a passive power (i.e. the mechanisms of bodily movement), we can already begin a causal analysis without articulating the ground of the mutually manifesting active power or powers. This brief section outlines several positive theses we can attain from applying the powers view of causation to mental causation.[215]

Consider these attributes of mental causation on the powers view, starting with the obvious. First, the mental and bodily powers are *asymmetrically* related in mental causation—the former active, the latter passive. Thus, properly human bodily movement depends on an asymmetric causal relationship between mental intentionality and the body's passive powers.

Second, and in a way that does differentiate a powers view from its alternatives, this asymmetry is *not temporal*. For, the mental power and bodily power together mutually manifest diachronically. The causal relata thus endure for exactly the same stretch of time.

Third, this diachronic mutual manifestation, as argued against double-prevention cases, is also quite permissibly the *co-manifestation* of multiple active (and passive) powers. A crucial corollary of this third thesis regarding mental causation is that co-manifestation explains how sufficient causes, including mental causes, can be complex without risking overdetermination. Thus, the active powers in a case of mental causation could without contradiction include *both physical and mental powers* as part of a composite sufficient cause. This thesis helps bolster my metaphysical agnosticism. Depending on one's sympathies, the manifesting active mental cause could be (1) physical, (2) mental, or (3) composite, permitting adoption of physicalism, dualism,

[215] For the sake of streamlining the conclusions, I have ignored the case in which mental powers are passive. Inversion is of course also possible. When I stub my toe, the body manifests active powers through sending nerve signals and the mental manifests passive powers through experiencing pain. But that's not mental causation.

or a hybrid position.[216] No doubt, empirical examination of the bodily passive power helps play into what metaphysics is the most coherent.

Fifth, and most crucially, since the causal relation can only be understood as the *pairing of an active and passive power(s),* it is inconceivable that we could observe just bodily movement absent the active mental power's mutual manifestation. We have, in short, a compelling powers-based account of how mental powers can, in the right circumstances cause bodily movement. Using the word "can" avoids the charge that active mental powers and passive bodily powers must *necessarily* operate together. Thwarted intentions and bodily paralysis obviously inhibit otherwise expected mind-body causation. Another way to capture this point would be to rephrase the modality as necessary, but to add the clause "in the right circumstances." That approach would rely on Aristotle's idea that "when the agent [power] and patient [power] meet in the way appropriate to the potentiality in question, the one must act and the other be acted on."[217] But if active mental powers and passive bodily powers meet in the appropriate way, one can expect successful mental causation.

The strength of the powers view is not its quotidian conclusion that mental causation occurs. Rather the strength is its ability to justify this quotidian conclusion from anodyne and immediate knowledge of the passive powers—even if we cannot view their mutually manifesting partners directly. We are allowed to do this because causation always involves the mutual manifestation of both active and passive powers.

[216] The dualist, of course, still faces the interaction problem.
[217] *Metaphysics* IX 5 1048a6-8.

5. Conclusion

What is novel in a powers account of mental causation is not its ingredients, but their combination. Thus, the problems of mental causation and dispositional properties are separately fairly well discussed. The true novelty of my project is that it uses the latter to clarify the former.

My proposal leads to the following conclusions. First, I suggest that a powers analysis of causation is more fundamental than an events analysis. Thus, if we speak about events as causes or effects, we do so by shorthand, but do not get at the root of the causal explanation. My tentative suggestion, based on the arguments from simultaneity and interruptibility, is that we instead focus on *processes* of mutually manifesting powers as the basic causal ontological category. The powers approach is explanatorily superior to the events-pair outlook, on my view, because it understands causal relata as composing a *single* processual entity.

Second, my proposal provides an explanation of mental causation without endorsing a particular view of mind. No doubt the metaphysics of mind presents arduous questions that should be pursued with precision and alacrity. Yet, I have argued that powers can explain much about mental causation without answering that separate question. When we observe bodily movement, we are a good way towards observing the whole instance of causation.

We do, in fact, observe the passive causal powers—the mechanisms of bodily movement—at work. Based on their manifestations being activated, we can make strides towards conjecturing at the kind of active powers which activate them. This ability is predicated on the fact that active and passive powers, when constituting causation, compose a single process. This sort of analysis is impossible on an events-pair view of causation.

I thus preliminarily conclude that, during a case of mental causation, the mind manifests active power(s) and the body passive one(s), the mind and body are related as non-temporally discrete asymmetric powers, that each is composed of co-manifesting sets of powers, that these powers together manifest diachronically, that mental causation *could* consistently occur by means of active causal powers composed of both physical and mental powers, and that, finally, the right circumstances necessitate this kind of interaction.

More can be said and in greater detail. But a powers hypothesis of mental causation coheres with a broader powers analysis of causation while explaining more than our other options. I leave the task of developing a full-fledged theory to future endeavors, hopefully undertaken by my betters.

Bibliography

Ackerill, Edward and Bernard Keating. “Does Interactionism Violate a Law of Classical Physics?” *Mind* Vol. 40 no. 357 (1981): 102-107.

Aimar, Simona. “Counterfactuals, Overdetermination, and Mental Causation.” *Proceedings of the Aristotelian Society* 111 (2011): 469-477.

Aristotle. *The Complete Works of Aristotle*. Trans. Jonathan Barnes. Princeton: Princeton University Press, 1984. 2 Vols.
———. *Physics* (vol. 1)
——— . *On Generation and Corruption* (vol. 1)
———. *Metaphysics* (vol. 2)

Aristotle. *Metaphysica: Pars Prior*. Hermann Bonitz ed. Leipzig: In Aedibus B. G. Teubneri, 1848.

Armstrong “Four Disputes about Properties.” *Synthese* Vol. 144 no. 3 (2005): 309-320.

Bach, Kent. “Actions Are not Events.” *Mind* Vol. 89 no. 353 (1980): 114-120.

Baker, Lynne Rudder. “Metaphysics and Mental Causation.” In *Mental Causation* John Heil and Alfred Mele eds. Oxford: Clarendon Press, 1993.

Bird, Alexander. “Dispositions and Antidotes.” *The Philosophical Quarterly*, Vol. 48 no. 191 (1998): 227-234.

Bernstein, Sara. “David Lewis’ Theories of Causation and their Influence.” In *Cambridge History of Philosophy*, ed. Kelly Becker. Cambridge: Cambridge University Press, 2019.

Bickle, John. “Multiple Realizability.” *The Stanford Encyclopedia of Philosophy* (2020), Edward N. Zalta ed., URL = <https://plato.stanford.edu/archives/sum2020/entries/multiple-realizability/>

Block, Ned. “Can the Mind Change the World?” In *Meaning and Method: Essays in Honor of Hilary Putnam*, ed. George Boolos. Cambridge: Cambridge University Press, 1990.

Block, Ned. “Do Causal Powers Drain Away?” *Philosophy and Phenomenological Research* Vol. 67 (2003): 133-150.

Cartwright, Nancy. “Causal Laws, Policy Predictions, and the Need for Genuine Powers.” In *Dispositions and Causes* ed. Toby Handfield. Oxford: Clarendon Press, 2009.

Cartwright, Nancy. “Causal Powers: Why Humeans Can’t Event Be Instrumentalists.” In *Causal Powers* ed. Jonathan Jacobs. Oxford: Oxford University Press, 2017.

Casati, Roberto and Achille Varzi. “Events.” *The Stanford Encyclopedia of Philosophy* (2020), Edward N. Zalta ed., URL = <https://plato.stanford.edu/archives/sum2020/entries/events/>

Charles, David. *The Undivided Self: Aristotle and the Mind-Body Problem*. Oxford: Oxford University Press, 2021.

Choi, Sungho and Michael Fara. “Dispositions.” *The Stanford Encyclopedia of Philosophy* (2018), Edward N. Zalta ed.
URL = https://plato.stanford.edu/archives/spr2021/entries/dispositions/

Coope, Ursula. “Aristotle on Movement, Incompleteness, and the Now.” Unpublished manuscript, forthcoming (2023).

Crane, Tim. *Dispositions: A Debate* ed. Tim Crane. London: Routledge, 1996.

Crane, Tim and Seinvör Thöll Árnadóttir “There is No Exclusion Problem.” In *Mental Causation and Ontology* S.C. Gibb, E.J. Lowe, and R.D. Ingthorsson eds. Oxford: Oxford University Press, 2013.

Crisp, Thomas and Ted A. Warfield. “Review: Kim's Master Argument.” *Noûs* Vol. 35 no. 2 (2001): 304-316.

Davidson, Donald. “Causal Relations.” In *Essays on Actions and Events.* Oxford: Clarendon Press, 1980 [1967].

Davidson, Donald. “Events as Particulars.” In *Essays on Actions and Events.* Oxford: Clarendon Press, 1980 [1970].

Davidson, Donald. “Thinking Causes.” In *Mental Causation*, John Heil and Alfred Mele eds. Oxford: Clarendon Press, 1993.

Dretske, Fred. “Mental Events as Structuring Causes of Behaviour.” In *Mental Causation* John Heil and Alfred Mele eds. Oxford: Clarendon Press, 1993.

Ellis, Brian. “The Categorical Dimensions of the Causal Powers.” In *Properties, Powers and Structures: Issues in the Metaphysics of Realism.* Alexander Bird, Brian Ellis, and Howard Sankey eds. New York: Routledge, 2012: 11-26.

Falcon, Andrea. “Aristotle on Causality.” *The Stanford Encyclopedia of Philosophy* (2023), Edward N. Zalta & Uri Nodelman eds. URL = <https://plato.stanford.edu/archives/spr2023/entries/aristotle-causality/>

Field, Hartry. “Causation in a Physical World.” In *The Oxford Handbook of Metaphysics* Michael Loux and Dean Zimmerman eds. Oxford: Oxford University Press, 2003.

Fodor, Jeremy. “Making Mind Matter More.” *Philosophical Topics* 17 (1989): 59–79.

Gibb, S.C., E.J. Lowe, and R.D. Ingthorsson eds. *Mental Causation and Ontology* Oxford: Oxford University Press, 2013.

Gibb, S.C. “Mental Causation and Double Prevention.” In *Mental Causation and Ontology*. S.C. Gibb, E.J. Lowe, and R.D. Ingthorsson eds. Oxford: Oxford University Press, 2013.

Gibb, S.C. “Physical Determinability.” *Humana.Mente Journal of Philosophical Studies* Vol. 29 (2015): 69-90.

Gordon, Roy. “Criteria for Choosing Transparent Conductors.” *MRS Bulletin* Vol. 25 (2000): 52–57.

Hall, Ned. “Two Concepts of Causation.” In *Causation and Counterfactuals.* John Collins, Ned Hall, and L.A. Paul eds. Cambridge: The MIT Press (2004): 225-276.

Heil, John. *The Universe As We Find It.* Oxford: Oxford University Press, 2012.

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

Holger, Andreas and Mario Gunther. “Causation in Terms of Production.” *Philosophical Studies* 177 (2020): 1565-1591.

Hornsby, Jennifer. “Which Physical Events are Mental Events?” *Proceedings of the Aristotelian Society* New Series Vol. 81 (1981): 73-92.

Jackson, Frank. “Mental Causation.” *Mind* Vol.105 no. 419 (1996): 377-413.

Kim, Jaegwon. “Events as Property Exemplifications.” In M. Brand and De. Walton eds. *Action Theory.* Dordrecht: Reidel, (1976): 159-177.

Kim, Jaegwon. “Concepts of Supervenience.” *Philosophy and Phenomenological Research* 45 no. 2 (1984): 153-176.

Kim, Jaegwon. “Mechanism, Purpose, and Explanatory Exclusion.” *Philosophical Perspectives* Vol. 3 (1989): 77-108.

Kim, Jaegwon. “Multiple Realization and the Metaphysics of Reduction.” *Philosophy and Phenomenological Research* Vol. 52 no. 1 (1992): 1-26.

Kim, Jaegwon. *Mind in a Physical World*. Cambridge: MIT Press, 1998.

Kim, Jaegwon. “Reviewed Work(s): Mind in a Physical World by Jaegwon Kim.” *Philosophy and Phenomenological Research* Vol. 65 no. 3 (2002): 671-680.

Kim, Jaegwon. *Physicalism, Or Something Near Enough*. Princeton: Princeton University Press, 2005.

Kroedel, Thomas. *Mental Causation: A Counterfactual Theory*. Cambridge: Cambridge University Press, 2019.

LePore, Ernest, and Barry Loewer. “Mind Matters.” *Journal of Philosophy* Vol. 84 (1987): 630–642.

Lewis, David. “Causation.” *Journal of Philosophy*, 70 (1973a): 556–67.

Lewis, David. *Counterfactuals*. Oxford: Blackwell, 1973b.

Lewis, David. “Counterfactual Dependence and Time’s Arrow.” *Noûs* Vol. 13 no. 4 (1979): 455-476.

Lewis, David. “Causal Explanation.” *Philosophical Papers*, Volume II, Oxford: Oxford University Press, (1986): 214-240.

Lewis, David. “Causation as Influence.” *The Journal of Philosophy*, Vol. 97 no. 4, (2000): 182-197.

Loewer, Barry. “Mental Causation, or Something Near Enough.” In *Meaning, Mind, and Matter: Philosophical Essays*. Ernest Lepore and Barry Loewer eds. Oxford: Oxford University Press, 2011.

Loewer, Barry. “Reviewed Work(s): Mind in a Physical World by Jaegwon Kim.” *Philosophy and Phenomenological Research* Vol. 65 no. 3 (2002): 655-662.

Lowe, E. J. “Physical Causal Closure and the Invisibility of Mental Causation.” In *Physicalism and Mental Causation: The Metaphysics of Mind and Action.* Sven Walter and Heinz Dieter Heckmann eds. Exeter: Imprint Academic, 2003.

Lowe, E. J. *Personal Agency: The Metaphysics of Mind and Action*. Oxford: Oxford University Press, 2008.

Mackie, J. L. “Causes and Conditions.” *American Philosophical Quarterly*, Vol. 2 no. 4 (1965): 245-264.

Maier, John. “Abilities” in *The Stanford Encyclopedia of Philosophy* (2022), Edward N. Zalta & Uri Nodelman eds. URL = <https://plato.stanford.edu/archives/fall2022/entries/abilities/>

Marmodoro, Anna. “Aristotelian Powers at Work: Reciprocity without Symmetry in Causation.” In *Causal Powers* ed. Jonathan Jacobs. Oxford: Oxford University Press, 2017.

Martin, C.B. “Dispositions and Conditionals.” *The Philosophical Quarterly* Vol. 44 no. 174 (1994): 1-8.

Martin, C.B. *The Mind in Nature*. New York: Oxford University Press, 2008.

Mayr, Erasmus. "Powers and Downward Causation." In *Philosophical and Scientific Perspectives on Downward Causation.* Michelle Paolini Paoletti and Francesco Orilia eds. New York: Routledge (2017): 76-91.

McLaughlin, Brian. "The Rise and Fall of British Emergentism." In *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*, A. Beckermann, H. Flohr, and J. Kim eds. New York: de Gruyter, 1992.

McLauglin, Brian. "The Varieties of Supervenience." In *Supervenience: New Essays*, Elias Savellos and Umit Yalcin eds. Cambridge: Cambridge University Press, 1995.

Menzies, Peter. "The Causal Efficacy of Mental States." In *Physicalism and Mental Causation: The Metaphysics of Mind and Action*, Sven Walter and Heinz Dieter Heckmann eds. Exeter: Imprint Academic, 2003.

Menzies, Peter. "Mental Causation in the Physical World." In *Mental Causation and Ontology* S.C. Gibb, E.J. Lowe, and R.D. Ingthorsson eds. Oxford: Oxford University Press, 2013.

Menzies, Peter and Helen Beebee. "Counterfactual Theories of Causation." *The Stanford Encyclopedia of Philosophy* (2020), Edward N. Zalta ed. URL = <https://plato.stanford.edu/archives/win2020/entries/causation-counterfactual/>

Mourelatos, Alexander. "Events, Processes, and States." *Linguistics and Philosophy* Vol. 2 no. 3 (1978): 415-434.

Mumford, Stephen. *Dispositions*. New York: Oxford University Press, 1998.

Mumford, S. and R. L. Anjum. "Double Prevention and Powers." *Journal of Critical Realism* Vol. 8 (2009): 277-293.

Mumford, Stephen, and Rani Lill Anjum. *Getting Causes from Powers*. Oxford: Oxford University Press, 2011.

Nussbaum, Martha. "Essay 1." In *Aristotle's De Motu Animalium.* Princeton: Princeton University Press, 1979.

Papineau, John. "The Rise of Physicalism." In *Physicalism and Its Discontents*, Carl Gillet, & Barry Loewer eds. Cambridge: Cambridge University Press: 2001, 3-36.

Paul, L.A., and Ned Hall. *Causation: A User's Guide*. Oxford: Oxford University Press, 2013.

Place, U.T. "Conceptualism and the Ontological Independence of Cause and Effect." In *Dispositions: A Debate* ed. Tim Crane. London: Routledge, 1996.

Prior, Elizabeth, Robert Pargetter, and Frank Jackson. “Three Theses about Dispositions.” *American Philosophical Quarterly* Vol. 19 (1982): 251-258.

Putnam, Hilary. “Psychological Predicates.” In *Art, Mind, and Religion* W.H. Capitan and D.D. Merrill eds. Pittsburgh: University of Pittsburgh Press, (1967): 37–48.

Robb, David, John Heil, and Sophie Gibb. “Mental Causation.” *The Stanford Encyclopedia of Philosophy* (2023), Edward N. Zalta & Uri Nodelman eds.
URL = https://plato.stanford.edu/archives/spr2023/entries/mental-causation/

Russell, Bertrand. “On the Notion of Cause.” *Proceedings of the Aristotelian Society* Vol. 13 (1912-1913): 1-26.

Ryle, Gilbert. *The Concept of Mind.* Harmondsworth: Penguin, 2000 [1949].

Schaffer, Jonathan. “Is There a Fundamental Level?” *Nous*, Vol. 37 no. 3 (2003): 498-517.

Shoemaker, Sydney. *Physical Realization*. Oxford: Oxford University Press, 2007.

Sider, Ted. “What’s so Bad about Overdetermination?” *Philosophy and Phenomenological Research* Vol. 67 (2003): 719–26.

Sosa, Ernest. “Mind–Body Interaction and Supervenient Causation.” *Midwest Studies in Philosophy* Vol. 9 no. 1 (1984): 271-281.

Stalnaker, Robert. “A Theory of Conditionals.” *Knowledge and Conditionals: Essays on the Structure of Inquiry.* Oxford: Oxford University Press, 2019 [1968].

Stein, Nathanael. “Causation and Explanation in Aristotle.” *Philosophy Compass* Vol. 6 no. 10 (2011): 699-707.

Steward, Helen. “Chapter 8: Agency and Causation.” In *A Metaphysics for Freedom*. Oxford: Oxford University Press, 2012.

Steward, Helen. *The Ontology of Mind: Events, Processes, and States*. Oxford: Clarendon, 1997.

Steward, Helen. “Processes, Continuants, and Individuals.” *Mind* Vol. 122 no. 487 (2013), 781-812.

Stout, Rowland. “Processes.” *Philosophy* 72 (1997): 19-27.

Tanney, Julia. “Gilbert Ryle.” *The Stanford Encyclopedia of Philosophy* (2022), Edward N. Zalta ed. URL = <https://plato.stanford.edu/archives/sum2022/entries/ryle/>

Williamson, Timothy. *Suppose and Tell: The Semantics and Heuristics of Counterfactuals.* Oxford: Oxford University Press, 2020.

Yablo, Stephen. “Mental Causation.” *The Philosophical Review*, Vol. 101 no. 2 (1992): 245-280.