



# Mental Causation

Karen Bennett\*

Princeton University

---

## Abstract

Concerns about ‘mental causation’ are concerns about how it is possible for mental states to cause anything to happen. How does what we believe, want, see, feel, hope, or dread manage to cause us to act? Certain positions on the mind–body problem – including some forms of physicalism – make such causation look highly problematic. This entry sketches several of the main reasons to worry, and raises some questions for further investigation.

---

## 1. Introduction

You are sitting at your computer, reading an online philosophy article. You become aware that you are thirsty, and would like something cold to drink. You go to the kitchen and get yourself a glass of ice water. That is – you *want* some water, you *believe* you can get some in the kitchen, and that belief and desire together cause your arms and legs to move in various complicated ways. Questions about mental causation are questions about how this is possible. How is it that our beliefs, desires, and other mental states manage to guide our action? Does what we think, want, feel, and perceive really cause our bodies to move?

Most people take it as a datum that the answer is ‘yes’. Jerry Fodor puts the point in his characteristically gripping way:

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. (‘Making Mind Matter More’ 156)

So if a view about the nature of mind does not allow for mental causation, that is a very large strike against it. It is not surprising, then, that some version of this charge has been levied at a large number of views: substance dualism, property dualism, and nonreductive physicalism of various forms, including in particular functionalism and anomalous monism.

‘Some version of this charge’, indeed. By my count, there are at least seven problems of mental causation, each depending upon different assumptions, and causing trouble for different views:

1. The traditional problem of interaction, for substance dualism.
2. Kim's new 'pairing problem', for substance dualism.
3. The exclusion problem – and the 'supervenience argument' that extends its reach – for anyone who denies type identity.
4. The causal inheritance problem, for all physicalists.
5. The problem of metaphysically necessitated effects, for role functionalists.
6. The anomalism problem, for anomalous monists like Davidson.
7. The extrinsicness problem, for anyone who believes in broad content.

I shall discuss only the first five here. I am not going to discuss either the extrinsicness problem or the anomalism problem; the reader can get a full introduction to these from any of a number of fine survey articles, introductory books, and anthologies (for example, Heil and Mele; Kim, *Philosophy of Mind*; Robb and Heil). Of course, the same could be said about the exclusion problem and the traditional problem of interaction. So why am I focusing on the five that I am focusing on? Because each of them meets a disjunctive criterion that seems appropriate for *Philosophy Compass*: each is either new, or neglected, or an issue about which I have something to say.

I will assume that the reader is familiar with the basic positions on the mind-body problem, and with the motivations for those positions. Two points seem worth explicitly noting at the outset, however.

First, I want to be clear about my usage of the term 'physicalism'. Physicalists believe that the mental (and everything else) supervenes with metaphysical necessity upon the physical. They believe that every minimal physical duplicate of the actual world is a mental duplicate as well (Jackson, *From Metaphysics to Ethics*; see Lewis, 'New Work'; Chalmers for discussion of similar definitions). Those who deny physicalism – i.e. dualists – deny that the mental supervenes on the physical in this way. Some dualists think that the mental supervenes upon the physical with merely nomological necessity, and some dualists do not think that it supervenes at all, but all dualists deny that the mental supervenes upon the physical with metaphysical necessity. All physicalists, in contrast, endorse that claim.<sup>1</sup> Physicalists do of course disagree amongst themselves on other matters. In particular, some physicalists are 'reductive physicalists', and some are 'nonreductive physicalists'; I will explain these labels in section 4. All that matters at the moment is that both kinds of physicalist endorse the metaphysically necessary supervenience claim.<sup>2</sup>

Second, much of what follows assumes that it makes good sense to talk about the causal efficacy of properties, to say that *c* (your thirst, say) caused *e* (your movement towards the kitchen) 'in virtue of' or 'qua' some but not other of its properties. This idea – and the worry that establishing the causal efficacy of particular ('token') mental events, states, or processes<sup>3</sup> does not guarantee the efficacy of mental event-*types* or properties – attracted attention in the wake of Davidson's defense of anomalous monism. Davidson himself resists all talk of the causal efficacy of properties ('Thinking Causes' 13), but

it really is a quite natural idea. Suppose you place an apple on a scale, causing the display to read '5 oz'. It seems clear that the *mass* of the apple is relevant to the effect, and that the *color* of the apple has nothing to do with it. (For more in-depth discussion, see several of the essays in Heil and Mele; especially McLaughlin.)

The rest of this article discusses the first five of the seven problems just listed, in order. Two questions will surface repeatedly in what follows, though the article is not organized around them. Both deserve further attention. First, to what extent do any or all of the problems depend upon particular views about the nature of causation? (Presumably, one should let one's theory of causation drive one's views about mental causation rather than the other way around.) Second, to what extent can concerns about the possibility of mental causation be wielded as a weapon against dualism? Physicalists frequently try to do so. I do think it is a mistake to treat nonreductive physicalism as though it faces exactly the same problems with mental causation as property dualism, simply because both views deny type identity (see section 5, and 'Exclusion Again'). However, for reasons that will emerge below, I am not so sure that arguments for physicalism from mental causation can really get very far.

## 2. Substance Dualism I: The Traditional Problem of Interaction

As the name suggests, substance dualists think that there are two radically different kinds of substances: mental and physical. Mental substance is essentially thinking, and is not extended in space; physical substance is essentially extended in space, and does not think. Minds and bodies are distinct, and can exist apart from each other.

Many substance dualists – most famously Descartes – are moved by the thoughts that began this article, and want to claim that minds and bodies causally interact. However, it is not clear that this position is coherent. Princess Elisabeth of Bohemia was the first to press this charge. She wrote to Descartes and asked him to please explain

how the mind of a human being can determine the bodily spirit in producing voluntary actions, being only a thinking substance. For it appears that all determination of movement is produced by the pushing of the thing being moved, by the manner in which it is pushed by that which moves it, or else by the qualification and figure of the surface of the latter. Contact is required for the first two conditions, and extension for the third. [But] you entirely exclude the latter from the notion you have of body, and the former seems incompatible with an immaterial thing. (Princess Elisabeth to Descartes, May 6/16, 1643)

The Princess' concern is that Cartesian minds are just not the right sort of thing to cause anything physical. They do not exist in space, they have no chemical composition, they carry no electric charge, they are not subject to the forces of gravity or magnetism. It therefore looks as though minds cannot push or pull bodies in any of the ways in which *bodies* can push and

pull other bodies. So Descartes owes us an account of how exactly minds can cause anything physical.

Descartes replied to the Princess by saying that the causal interaction between minds and bodies must be taken as primitive, and that she was wrong to think that it could be understood on the model of physical causation. Indeed, each of the three notions in play – body, mind, and their union – must be taken as primitive, and

we go wrong if we try to explain one of these notions by another, for since they are primitive notions, each of them can only be understood by itself . . . [it is a mistake] to conceive the way in which the mind moves the body after the manner in which one body is moved by another. (Descartes to Princess Elisabeth, May 21, 1643)<sup>4</sup>

Unsurprisingly, Princess Elisabeth was not impressed. She replied that ‘it would be easier for me to concede matter and extension to the mind than it would be for me to concede the capacity to move a body and be moved by one to an immaterial thing’ (Princess Elisabeth to Descartes, June 10/20, 1643).

But although Princess Elisabeth’s attitude is shared by many contemporary physicalists, whether it ultimately succeeds appears to depend upon what view about the nature of causation is correct. In particular, it looks as though the Princess is assuming a *process* view of causation, according to which cause and effect must be linked by some sort of process of production or generation (e.g. Salmon; Dowe). But there are other views about the nature of causation on which it is not clear that Descartes faces any real difficulty at all. Two quick examples. First, on a Humean regularity theory, causation is just a matter of constant conjunction – one event causes another just in case events of the first type are reliably followed by events of the second type. But presumably even the Cartesian dualist can say that water desires are reliably followed by water-seeking behavior, that pains are reliably followed by stimulus-avoidance behavior, and so forth (see Loeb). Second, on a counterfactual dependence theory, one event causes another if the second event would not have occurred if the first had not occurred (see Lewis, ‘Causation’; ‘Causation as Influence’). Here too, it looks as though the Cartesian dualist might be able to claim that there are causal links between the mental and the physical. After all, if I had not wanted water, I would not have gone looking for some; if I had not been in pain, I would not have jerked my hand away from the hot stove. (The only tricky question is going to be deciding whether those counterfactuals are back-tracking.<sup>5</sup>) Crucially, in neither case would the Cartesian dualist be claiming that mental-physical causal interaction is in any way special. He would be treating it entirely on a par with purely physical causation.

It is therefore not clear that Princess Elisabeth’s objection has the force that we usually treat it as having when we teach undergraduate philosophy of mind courses. Settling the issue would appear to require settling what the

correct account of causation is. (See the *Philosophy Compass* article on causation: Hall's 'Philosophy of Causation'.)

### 3. Substance Dualism II: The 'Pairing Problem'

Jaegwon Kim has recently provided an updated version of the interaction problem, partly to avoid the concern just raised (*Physicalism* ch. 3; *Philosophy of Mind* 44–50). Kim's claim is that the substance dualist cannot explain how or why one immaterial mind rather than another counts as engaging in some particular causal interaction with the physical. Why is it that my mind controls *my* body, rather than yours? More precisely, Kim suggests that the Cartesian dualist should think that it is metaphysically possible for there to be two intrinsically indiscernible Cartesian minds, such that

- a) they simultaneously perform indiscernible mental acts  $m_1$  and  $m_2$ , and
- b)  $m_1$  causes some particular physical effect  $p$  that  $m_2$  does not cause.

But then we can ask what the relation is that pairs the first soul's action, but not the second's, with the physical effect (*Philosophy of Mind* 46). If we could appeal to spatial relations, this question would be easy to answer. Consider, for example, two intrinsically indiscernible rock-throwings. The fact that one results in the smashing of a window and the other does not can be fully explained by the fact that the first was *aimed at a window* and the second took place on an empty beach. But, of course, the Cartesian dualist cannot tell this sort of story about the mental activities of souls, for she denies that they occur in space at all. So how is she to explain what pairs mental causes with their physical effects?

Notice that simply appealing to something like Humean constant conjunction, subsuming laws, or counterfactual dependence does not really answer the question (Kim does not emphasize this point, but see *Physicalism* 83n15, 84). If  $m_1$  and  $p$  fall under types that are constantly conjoined, or if  $p$  counterfactually depends on  $m_1$ , we can always ask *why* – and our answers are tightly constrained by the fact that  $m_2$  and  $p$  are *not* so connected. Why does  $p$  counterfactually depend on  $m_1$ , given that it does not counterfactually depend on  $m_2$ ? Why are  $m_1$ -type events reliably followed by  $p$ -type events, while  $m_2$ -type events are not? The answers to these questions must not invoke  $m_1$ 's intrinsic properties nor any spatial relations between it and  $p$ . The point of the pairing problem is that the pairing relation does not supervene on intrinsic properties plus spatial relations.

How should the Cartesian dualist best respond? She has a couple of options, but in all cases her first move should be to argue that the problem also arises within the purely physical domain, and therefore has nothing specifically to do with the nature of Cartesian minds. Let us start, then, with that claim. Here are two ways the pairing problem might arise for physical events.

*Case 1.* Imagine a world containing only two intrinsically indiscernible particles, each of which undergoes an intrinsically indiscernible process that has a 50% chance of generating a daughter particle. Imagine that one daughter particle is generated equidistant between the two originals. What makes it the effect of one of the original particles rather than the other? (This sort of case is due to Tooley *Causation*; ‘Causation: Reductionism vs. Realism’)

*Case 2.* Many people believe in spatio-temporally coincident objects and events (for the classic modal argument, see Gibbard; for nonmodal versions, see Fine). Consider a particular statue, Goliath, and the lump of clay, Lump1, of which it is made. Anyone who thinks that these are distinct objects probably also thinks that they are involved in distinct events – that Lump1’s falling on my foot is not identical to Goliath’s falling on my foot. Which of those two events caused my toe to break? I will not survey the possible answers here, aside from noting that the issue is closely related to (but not the same as) the exclusion problem discussed in section 5. All I wish to point out is that anyone who says that only *one* of the events caused my toe to break faces the pairing problem. What makes (say) *Lump1*’s falling on my foot the cause of my injury, if Goliath’s falling on my foot is not? The two events are intrinsically and spatio-temporally indiscernible.

The claim that the pairing problem can arise within the physical domain defuses it as an objection to dualism in particular. However, it does not quite suffice as a solution. After all, we have not been told what to say about the problem, just that it arises more broadly than we might have thought.<sup>6</sup> To parlay this into an actual response, then, the Cartesian dualist should do one of two things.

First, she could argue that she can avail herself of substantive responses that have been proposed for these other cases. This does not seem promising, however. The main response to examples like case 1 above has been to say that there simply is no fact of the matter about which was the cause (Lewis, ‘Postscripts to “Causation”’ 180–1; Schaffer 8). However, it is unlikely that a Cartesian dualist will want to allow that there might be no fact of the matter about which mind causes what. Another option might be to try to adopt Stephen Yablo’s treatment of causal relations involving spatio-temporally coincident objects (‘Cause and Essence’; ‘Mental Causation’), but there is no particular reason to think that his ‘proportionality constraint’ will do the Cartesian dualist any good.

Second, the dualist could simply insist that the pairing relation is primitive. This line is made more plausible by the fact that the pairing problem can arise from assumptions other than Cartesian dualism. The claim would just be that *causation* is primitive. Whether the causal relation holds between two events is just a matter of brute fact. It holds between one of the original particles and the creation of the new particle, and does not hold between the other and the creation of the new particle. Similarly, it holds between  $m_1$  and  $p$ , and does not hold between  $m_2$  and  $p$ . In neither case is there anything further to be said on the matter.

This response has two virtues. First, it echoes Descartes's response to Elisabeth, and thus has at least some claim to historical plausibility. Second, it is a view about the nature of causation that a variety of people hold for reasons having little to do with dualism. Michael Tooley (*Causation; 'Causation: Reductionism vs. Realism'*), David Armstrong (e.g. 'Going Through the Open Door Again'), and John Carroll all defend 'realism' or 'antireductionism' about causation: the view that the causal facts do not supervene on the noncausal facts and noncausal laws. Fix the noncausal laws of nature, as well as *all* the noncausal features of two events – including their intrinsic properties and spatio-temporal locations – and you still have not settled whether there is a causal relation between them. Indeed, Case 1 above is a case of Tooley's. Note, then, that the force of the pairing problem rests on the thought that this view is not plausible. Kim's argument only shows, and is really only intended to show, that Cartesian dualists are committed to anti-reductionism about causation. It does not show that this commitment constitutes an objection to Cartesian dualism. I myself am no fan of anti-reductionism about causation, but I nonetheless think it is important to see that further work is required to make the case against Cartesian dualism in this way.

The pairing problem is interesting and under-studied; it deserves further attention.<sup>7</sup>

#### 4. Interlude: Nonreductive Physicalism

Of course, Cartesian dualism is not the only view that has trouble accommodating mental causation. Property dualists like Chalmers do too, as do many physicalists. Which physicalists? All those who deny type (property) identity (and arguably *only* those, though there are definitely some complications here).<sup>8</sup>

There are various reasons to deny type identity, some compatible with physicalism and some not. One reason to deny type identity is the Davidsonian concern about the anomalism of the mental, which I have already said I will not discuss here. Another reason, one that moves nonphysicalists, is the thought that conscious experience is deeply mysterious, and cannot be physically explained. The primary reason why many *physicalists* deny type identity is simply that type identity conflicts with the claim that mental states are 'multiply realizable' – that different creatures can possess the same mental state-type in virtue of different physical state-types (see the *Philosophy Compass* article on multiple realization: Funkhouser's 'Multiple Realizability').

Those that respond to multiple realization by denying type identity are generally called 'nonreductive physicalists'. They are the physicalists that have trouble accounting for mental causation. It should be noted, however, that not every physicalist responds to the putative multiple realizability of mental state-types in this way. Lewis has long argued that the proper response

is to claim that mental predicates non-rigidly designate different physical state-types in different kinds of creature ('Argument for the Identity Theory'; 'Review of *Art, Mind and Religion*'; 'Reduction of Mind'). Perhaps 'pain', when applied to humans, picks out *C-fiber stimulation*, but when applied to Martians picks out *M-fiber stimulation*. This enables him to endorse 'local', species-specific type-identities consistently with the appearance of multiple realizability. Kim agrees with the local reductions, though he, unlike Lewis, explicitly denies the existence of cross-species multiply realizable mental properties ('Multiple Realization').<sup>9</sup> So some physicalists think predicates like 'pain' refer to genuinely multiple realizable properties, and hence to properties that cannot straightforwardly be considered physical. Others resist this move. The former are nonreductive physicalists, and the latter are reductive physicalists.

Functionalists come in both nonreductive and reductive varieties. The former sort are 'role functionalists'; the latter sort are 'realizer functionalists'.<sup>10</sup> Role functionalism is the version that comes from Putnam ('Psychological Predicates'; 'Philosophy and Our Mental Life'); realizer functionalism is the version that comes from Lewis ('Argument for the Identity Theory'; 'Review of *Art, Mind and Religion*'; 'Psychophysical and Theoretical Identifications') and Armstrong (*Materialist Theory of the Mind*). Both take mental state-types to be in some sense associated with causal roles – to give an oversimplified example, the causal role associated with pain is *being typically caused by tissue damage, and typically causing aversive reactions*. Realizer functionalists identify the mental state-type *pain* with whatever occupies this causal role, while role functionalists identify it with the role itself. More precisely, role functionalism says that mental state-types are second-order, relational properties. Pain is the property of having some property or other that meets a certain functional specification, or 'plays the pain role'. Realizer functionalism, in contrast, says that pain is the *first-order* property that actually plays the role in some population – in humans, for example, pain might be C-fiber stimulation. Role functionalists take mental predicates like 'pain' to rigidly designate second-order functional properties, while realizer functionalists take them to nonrigidly designate first-order physical properties. (For more detail, see Block 'What is Functionalism?'; Lewis 'Argument for the Identity Theory'; 'Review of *Art, Mind and Religion*'; 'Psychophysical and Theoretical Identifications'; especially 'Reduction of Mind').<sup>11</sup>

So there certainly are physicalists who endorse a version of the claim that mental state-types are identical to physical state-types, even in the light of concerns about multiple realization. Those who do so are largely motivated by some package of the concerns about the causal efficacy of mental properties that I shall discuss in the rest of this article.

It should be noted, however, that not everyone is convinced. Many think that such views are not plausible all things considered, or even that they miss the point as a way of securing the causal efficacy of mental properties

(e.g. Block 'Can the Mind Change the Word?' 146; Jackson and Pettit, 'Functionalism' 387–8; 'Causation' 199–200; Hiddleston. See Fodor 'Special Sciences: Or, the Disunity'; 'Special Sciences: Still Autonomous'; Block, 'Anti-Reductionism Slaps Back' for the background view about the autonomy of psychology). Discussing their concerns in any detail is beyond the scope of this article, but I would like to at least note what might be called the *dilemma of reductionism*. On the one hand, those who go for 'local reductions' – claiming, for example, that human pain is identical to C-fiber stimulation – have little trouble accommodating the relevance of mental properties to particular occurrences, but arguably cannot make sense of causal generalizations and *ceteris paribus* laws. Since pain is realized differently in other creatures, it is hard to see how local reductionists can account for claims like 'pains cause aversive reactions'. (Perhaps they should refuse to do so.) On the other hand, those who resist reductionism by instead identifying mental properties with functional properties have trouble – as we shall see – accommodating the causal efficacy of mental properties.

The rest of this article basically focuses on the second half of this dilemma. Just what *are* the challenges to the causal efficacy of mental properties that nonreductionists face, and that motivate local reductionists and realizer functionalists? As I have already said, I am not going to discuss every challenge; I am going to ignore the problem from anomalism and the problem from extrinsicness. Instead, I propose to spend the remainder of the article focusing on problems 3, 4, and 5, listed above. While these problems are often mentioned in the same breath, they are in clear need of disentangling.

The first of these is the familiar and widely discussed exclusion problem. It at least purports to be a problem for both dualists and nonreductive physicalists. The second is an argument from what Kim calls 'the causal inheritance principle', and which is only a problem for nonreductive physicalists. The third is what I call 'the problem of metaphysically necessitated effects',<sup>12</sup> and is only a problem for role functionalists. I will begin with a fairly extensive discussion of the exclusion problem. I will briefly discuss the other two, with a particular eye towards the question of whether they are really free-standing problems in their own right, or whether they merely piggyback upon the exclusion problem.

### 5. The Exclusion Problem

The exclusion problem is also known as the overdetermination argument. (Kim's recent 'supervenience argument' is an extension, and will be discussed below.) It is important to see that this problem is somewhat different from the other problems about mental causation. The other problems – including the extrinsicness problem, the anomalism problem, and the two problems yet to be discussed – turn on the claim that the mental is somehow *by its nature* unsuited to causing anything. It is not up for causing anything

because it is not spatially extended, or is not invoked in the requisite sort of strict laws, or is somehow inappropriately relational, or inappropriately second-order . . . The exclusion problem does not turn on any such claim. It is best thought of as arising in the *wake* of these other worries. We can pretend that we have both established that, and explained how, the mental can count as causally efficacious. The question remains – how does it manage to be *nonoverdeterministically* causally efficacious? How does it manage to do anything but overdetermine effects that can already be fully accounted for in physical terms?

To see the distinction here, take the ‘causal work’ metaphor seriously for a moment. There are two hurdles to becoming gainfully employed. First, one must have the requisite skills; second, there must be employment opportunities available. It’s one thing to be fit for work, and quite another to actually find a job! The exclusion problem comes in at the second stage; the other arguments enter at the first. It is worth insisting upon this distinction, because it follows that responding to the exclusion problem requires less than is sometimes supposed. Responding to the exclusion problem does not itself require positively establishing that the mental *is* by its nature perfectly apt for entering into causal relations. That is required by a full defense of the causal efficacy of the mental from all challengers, but not by a defense from the exclusion problem in particular.

As far as I am aware, the exclusion problem is originally due to Norman Malcolm, in a paper called ‘The Conceivability of Mechanism’. Malcolm himself concluded that ‘mechanism’ could not be true – that neuroscience is not ‘adequate to explain and predict all movements of human bodies except those caused by outside forces’ (45). Others, like Kim (‘Mechanism’; ‘Myth of Nonreductive Physicalism’; ‘Nonreductivist’s Troubles’; ‘Postscripts on Mental Causation’; *Mind in a Physical World*), use a version of the exclusion problem to argue for the type identity theory. The exclusion problem can be spun in both directions, because at heart the issue is that the following five claims form an inconsistent set:

1. Distinctness: Mental properties are distinct from physical properties.<sup>13</sup>
2. Completeness: Every physical occurrence has a sufficient physical cause (or at least has its probability fixed by purely physical antecedents).
3. Efficacy: Mental events sometimes cause physical ones, and sometimes do so in virtue of their mental properties.
4. Nonoverdetermination: The effects of mental causes are not systematically overdetermined.
5. Exclusion: no effect has more than one sufficient cause unless it is overdetermined.

Everyone has to reject one of these five claims; the only question is which. Here are the options, in numerical order.

To reject #1 is to give up on anti-reductionism, and to return to type-identity claims. The problem only arises for those who accept the first premise.

To reject #2 is to claim, with Malcolm, that physics is causally incomplete, that the full explanation of some physical occurrences has to reach beyond the physical. However, no physicalist can take this route, and it is not particularly attractive even to dualists (Chalmers 150). Nonetheless, it must be admitted that it is not obvious how to satisfactorily argue that even dualists should accept the completeness of physics. Papineau has argued that we have empirical reason to accept it ('Arguments'; 'Rise of Physicalism'). However, this is only true if counterfactual reasoning and appeals to simplicity count as empirical – the completeness of physics certainly cannot be read off of the purely *observational* data.

There are at least two ways to reject #3. The obvious one is to endorse epiphenomenalism, the claim that the mental never causes anything at all. This is even less attractive than rejecting the completeness of physics (see Jackson, 'Epiphenomenal Qualia'; Bieri; Chalmers 150–60; Robinson for defenses). But note that rejecting #3 does not in fact require claiming that the mental never causes anything at *all*, just that the mental never causes anything *physical*. An alternative to epiphenomenalism, then, is to endorse a 'dual-explanandum' strategy, according to which mental effects have mental causes, and physical effects have physical causes (e.g. Dretske, 'Reasons and Causes'; 'Mental Events'; Hornsby). Properly sorting this out requires characterizing what makes an effect count as 'mental' or 'physical', but the basic idea is that the two sorts of causes bring about different effects – mere bodily movements vs. actions, for example. This sort of approach faces various questions, particularly about the size of its ontology. The only issue I shall pursue here, however, is that it requires denying that the mental supervenes on the physical. That means that it not only requires rejecting physicalism, but requires rejecting most of the more sensible forms of dualism as well. (Physicalists, recall, think that the mental supervenes on the physical with metaphysical necessity; some dualists, like Chalmers, think it supervenes with merely nomological necessity.)

Why does the dual explanandum strategy require denying that the mental supervenes on the physical? Because the supervenience claim extends the exclusion argument to problematize mental to mental causation as well. When the exclusion argument is supplemented in this way, Kim calls it the 'supervenience argument' (*Mind in a Physical World* 38–47; *Physicalism* ch. 2).<sup>14</sup> Here is how it goes. Suppose some mental occurrence  $M_2$  has a mental cause  $M_1$ . Either  $M_1$  causes  $M_2$  directly, or it does so by causing  $M_2$ 's physical supervenience base  $P_2$  to be instantiated. If it causes  $M_2$  directly, we have a tension between the two 'determiners' of  $M_2$ :  $M_1$  and  $P_2$ . The tension is not exactly *causal*, and hence would only be ruled out by a *modified* version of the exclusion principle, but the point seems clear.  $P_2$  itself guarantees  $M_2$ , so what need of  $M_1$ ?<sup>15</sup> Alternatively,  $M_1$  could cause  $M_2$  by causing its supervenience base  $P_2$  to be instantiated. It is independently plausible to think that this is the only way to cause a supervening property to be instantiated. As Kim points out, drawing on an analogy with aesthetic

properties, 'there is no *direct* way to make your painting more beautiful . . . you must change it physically if you want to change it aesthetically – there is no other way' (*Mind in a Physical World* 43; italics mine). But if  $M_1$  causes  $M_2$  by causing  $P_2$ , then we are back at the original exclusion problem.

Consequently, it does not look promising to respond to the exclusion argument by endorsing a dual explanandum strategy. Doing so requires either rejecting the supervenience of the mental on the physical, or else claiming that the supervenience argument goes wrong in some other way – but figuring out where *else* the supervenience argument goes wrong basically amounts to figuring out where the original exclusion argument goes wrong. Let us therefore return to the exclusion argument proper.

We have reached premise #4, Nonoverdetermination. To deny #4 is to say that the effects of mental causes are always overdetermined. There are two ways to make this claim. One way to do so is to simply embrace the idea that the effects of mental causes are always overdetermined in precisely the same way as when a death is caused by several members of a firing squad simultaneously shooting their guns. Mental causation is precisely analogous to the deaths of firing squad victims; this is what happens whenever we move our bodies. There is, however, an alternate way to deny premise #4, which is really just a terminological variant of denying premise #5. This is to say that although the effects of mental causes are indeed overdetermined – they do, after all, have two sufficient causes – they are importantly *different* from the death of the firing squad victim, and that this difference is precisely what makes it unproblematic to claim all of our actions are overdetermined.

Why is this just a terminological variant of denying #5? To deny #5 is to say that some effects have more than one sufficient cause without thereby counting as overdetermined. The second way of denying #4 basically says that some effects have more than one sufficient cause without thereby counting as overdetermined *in the standard sense*. In both cases, though, the claim is that some cases of 'double-causing' are very different than others. It makes no difference whether we characterize this as a distinction between 'good' and 'bad' overdetermination, or as a distinction between overdetermination and not. I shall arbitrarily choose to speak in the second way, and characterize this move as a rejection of #5, the exclusion principle. Call views that respond to the exclusion problem this way *compatibilist*, because they say that the mental and physical causes of the same effect are perfectly compatible with each other.

Various versions of compatibilism have been proposed recently (e.g. Yablo, 'Mental Causation'; Shoemaker, 'Realization and Mental Causation'; Pereboom; Bennett, 'Why the Exclusion Problem Seems Intractable'). The details differ, but all of them turn on the idea that the relation between the mental and physical causes of an effect is much more intimate than the relation between two shooters on a firing squad. Perhaps the physical cause constitutes, realizes, or determines the mental cause; perhaps the mental

cause simply supervenes with metaphysical necessity on the physical one. The two tasks here, for any compatibilist, are a) to characterize the relation between the mental and the physical; and b) to explain *why* the fact that that relation holds defuses the threat of overdetermination. Different compatibilists may provide different answers to these questions. But they are united in rejecting Kim's assertion that

the exclusion problem doesn't go away when we recognize the two purported causes as in some way related to each other. As long as they are recognized as distinct events, each claiming to be the full cause of a single event, the problem remains. (*Mind in a Physical World* 53)

I think compatibilism is the best way out of the exclusion problem. I also think that it is quite important to notice that the relations I appealed to in sketching the position are all paradigmatically *physicalist* relations. Indeed, I do not think that dualists, who deny that the supervenience claim holds with metaphysical necessity, can properly motivate compatibilism (see Bennett, 'Exclusion Again').

However, I also think that it is not fully clear that dualists should let the exclusion problem bother them very much. Whether or not they should depends entirely on whether or not they are obligated to accept the completeness of physics. (Descartes certainly would not accept it; Chalmers, in contrast, is inclined to.) In the end, then, perhaps it is not fully clear whether *anyone* should be too distressed. Dualists can deny completeness, and physicalists can adopt compatibilism as a cost-free solution – pending serious objections, of course, which have not yet been forthcoming.

One final point. Earlier, I suggested that the traditional problem of interaction appears to depend upon a particular model of causation. The same question arises here. Does the exclusion problem similarly depend upon a production or connecting-process view of causation? (See Loewer, 'Comments on Jagwon Kim's *Mind*'; 'Mental Causation'; Bennett, 'Exclusion Again' for discussion.)

### 6. *The Argument from Causal Inheritance*

In a couple of places, Kim has appealed to a claim that he calls the 'causal inheritance principle':

If mental property *M* is realized in a system at *t* in virtue of physical realization base *P*, the causal powers of *this instance* of *M* are identical with [or are a subset of]<sup>16</sup> the causal powers of *P* ('Multiple Realization' 326; also *Mind in a Physical World* 54).

This principle plays a central role in Kim's argument for the claim that there are no multiply realized mental properties, that 'each mental kind is sundered into as many kinds as there are physical realization bases for it' ('Multiple Realization' 327). But he *also* seems to use it to directly argue against the causal efficacy of mental properties (*Mind in a Physical World* 54–5, 115–16).

How exactly might such an argument go? And does it stand alone, or does it rely upon some version of the exclusion problem?

The way to put the causal inheritance principle to use here is to conjoin it with the idea that standard talk of the causal efficacy of properties is really talk of the causal efficacy of property *instances*. Surely what might do causal work is *M* as instantiated on a particular occasion, not *M* the *property itself* – the abstract entity, or set of instances, or whatever your view about the nature of properties dictates. So all we could reasonably want from an account of the causal efficacy of mental properties is an account of the causal efficacy of mental property instances. If this thought is conjoined with the Causal Inheritance Principle, it follows that mental properties are never independently causally efficacious. They never make any contribution of their own.

So is this a freestanding argument? That depends upon how the Causal Inheritance Principle is justified. One way to justify it would be by saying that if a particular instance of *M* were to have causal powers in *addition* to those of its realizer *P*, then any effects for which *M* is responsible would be overdetermined in a problematic way . . . etc., *à la* the preceding section. If that is the only way to justify the principle, then clearly the argument from causal inheritance is simply a variant of the exclusion problem.

But perhaps the principle can be justified differently. Kim suggests that the principle follows directly from physicalism. To deny it is ‘tantamount to belief in magic’ (*Mind in a Physical World* 55), and to accept causal powers that ‘magically emerge at a higher level and of which there is no accounting in terms of lower-level properties and their causal powers and nomic connections’ (‘Multiple Realization’ 326). The thought is that physicalism itself directly entails that there simply cannot *be* any non-physical causes. A commitment to physicalism requires shunning emergent causal powers. Because admitting any non-physical causes would violate physicalism, everything that happens has only physical causes.

Notice that this claim is much stronger than the premises in the exclusion argument. In particular, it is stronger than the completeness of physics. Contrast the following two claims:

Completeness: Every physical occurrence has a sufficient physical cause (or at least has its probability fixed by purely physical antecedents)

Closure: every physical occurrence has *only* physical causes (only physical antecedents affect its probability)

My use of the labels is idiosyncratic – they tend to be used interchangeably in the literature – but the distinction between the claims is important. While the former roughly says that everything has physical causes, the latter roughly says that nothing has non-physical causes. With the latter principle in hand, we can bypass exclusion-style reasoning; it *directly entails* that the mental is either physical or epiphenomenal.

What should we think of this argument? That depends upon two questions. One is whether physicalism really does entail the Causal

Inheritance Principle. It certainly appears to, but perhaps it could be argued that instances of mental properties can have causal powers not possessed by their physical realizers, as long as those causal powers supervene with metaphysical necessity on the causal powers of the physical realizer. (Getting anywhere with this thought obviously requires having a substantive account both of the realization relation, and of what causal powers are. See Shoemaker, 'Realization and Mental Causation' for such an account, though not one that would support this line of resistance.)

The other question is whether anyone should be bothered by this argument. After all, the conclusion is only that physicalists cannot think that the mental is *independently* causally efficacious, not that they cannot think it is efficacious at all. But why should they be bothered by the claim that the efficacy of the mental fully derives from the efficacy of the physical? That, surely, is just what it means to be a physicalist in the first place. I myself take this argument to nicely illustrate the physicalist world view, rather than pose any genuine problem for it. Dualists will have to do further work to convince the physicalist not to bite the bullet here.

### 7. Functionalism and the Problem of Metaphysically Necessitated Effects

Functionalists are sometimes accused of facing a special problem accounting for the causal efficacy of mental properties (e.g. Block, 'Can the Mind Change the World?' 155–60). The problem is specific to those functionalists who identify mental state-types with functional state-types – i.e. to role functionalists rather than realizer functionalists (see section 4).

The difficulty arises from the fact that role functionalists say that to be in pain is to have some first-order property or other that 'plays the pain role' – a role which, *inter alia*, involves causing an avoidance of the painful stimulus. Thus to say that the pain caused me to have an aversive reaction (to yank my hand off the hot stove, for example) is to say that my having some property or other that causes aversive reactions caused my aversive reaction. To a role functionalist, attributing causal power to mental states looks like attributing causal power to dormativity. But, the story goes, sleeping pills do not cause sleep in virtue of being dormative. If anything, it is the other way around – the pills count as dormative in virtue of causing sleep! To be dormative just is to have some property or other that causes sleep. So surely what makes a pill cause sleep is whatever first-order property realizes dormativity in that instance, not dormativity itself.

That is a purposely rather vague characterization of the argument; I shall spend the rest of this section trying to sort it out. But whatever exactly the problem is supposed to be here, note that realizer functionalists do not face it. They do not think that mental predicates like 'pain' pick out second-order properties in the way that 'dormativity' does, but rather pick out first-order properties like C-fiber stimulation (though they pick out different first-order properties in different contexts). Consequently, they

think that to say that my pain caused my aversive reaction is to say that my firing C-fibers caused my aversive reaction. Because realizer functionalists deny that mental properties are second-order properties, they dodge the threat here. Indeed, this issue provides a strong motivation for realizer functionalism over role functionalism, at least among Australians and honorary Australians.<sup>17</sup> It is a motivation for Lewis (perhaps ‘Argument for the Identity Theory’ 103; ‘Reduction of Mind’) as well as for Jackson (‘Essentialism’ 267n17; ‘Casual Roles’; see *From Metaphysics to Ethics* 90–5 for the parallel claim about color; see also Prior, Pargetter, and Jackson for the parallel claim about dispositions). Even Jackson’s earlier self, who, with Pettit, seemed torn about role and realizer functionalism, accepted that epiphenomenalism was the price of role functionalism. That is why they propose a theory of ‘program explanation’, which allows role states to function in good explanations, despite being epiphenomenal (‘Functionalism’; ‘Causation in the Philosophy of Mind’; ‘Program Explanation’).

So what *exactly* is the problem for role functionalism supposed to be? The literature is rather unclear on this point. I want to begin by clarifying two preliminary issues, and then, with those clarifications in hand, explore how the argument might be fleshed out.

First clarification: people often talk as though the problem has something to do with *analyticity* (e.g. Antony and Levine 91–3; Kim, *Mind in a Physical World* 51–5). But this is not correct. The real problem – if there is a problem – has to do with *necessity*. The problem is not that certain causal claims come out analytic, but rather that certain causal relations hold necessarily. Continue to suppose that causing aversive reactions is part of the constitutive functional role of pain – that part of what it is to be a pain is to cause aversive reactions. Then it is metaphysically necessary that pains cause aversive reactions. But it does not follow that ‘pains cause aversive reactions’ is *analytic*. That depends upon what the proper semantics of ‘pain’ is. Some functionalists think that descriptive content that specifies the functional role is part of the meaning of the term, but many others do not. (Indeed, those that *do* think it is part of the meaning of the term often think that mental terms are nonrigid designators of first-order properties – i.e. are realizer functionalists!) Notice also that so-called ‘psychofunctionalists’ or ‘empirical functionalists’, who think that the functional role characteristic of pain must be discovered through scientific research rather than reflection on the folk platitudes, can deny that claims like ‘pains cause aversive reactions’ are analytic. But like all role functionalists, they are claiming that the property is second-order; it is the property of having some property or other that plays a certain (empirically discovered) causal role.

In short, it need not be the case that a specification of the causal role gives the meaning of ‘pain’ for it to be the case that the causal role is the nature of *pain*. The problem at issue is not about the analyticity of certain causal statements. It is rather about the necessity of certain causal relations. If role

properties can be causes, they would appear to metaphysically necessitate their effects.

Second clarification: the problem here is, in the first instance, a problem about the causal efficacy of properties or types, not particulars. This is because the necessary connections are, in the first instance, between properties, not particular events that possess them. In other words, the primary problem is that role functionalists have to say that claims like ‘pains cause aversive reactions’ are necessary *de dicto*. But they are not obviously committed to saying that such claims are necessary *de re*. The fact that it is necessary that *E*-causers cause *Es* does not itself entail that it is necessary that any particular state that is in fact an *E*-causing state cause an *E*. Now, that *de re* claim does follow from the conjunction of role functionalism with the claim that each token mental state or event belongs to its mental type essentially – e.g. that each particular pain is essentially a pain. That would entail that a particular pain essentially cause aversive reactions, and thus would generate the problem for tokens as well as types. But although this essentialist thesis is quite plausible, particularly on some views about the nature of events or states, it is nonetheless a further question whether it should be adopted.

OK, so what *is* the problem here? What is wrong with saying that ‘pains cause aversive reactions’ is metaphysically necessary? I think there are a couple of threads of argument to be found in the literature, which deserve to be untangled.

One way to flesh out the difficulty, due to Ned Block, is basically a variant on the exclusion problem. If there is a necessary connection between two properties, there is no need to also posit a causal relation between their instances (Block, ‘Can the Mind Change the World?’ 157–8). Wouldn’t that be overkill? An aversive reaction is perfectly well accounted for by a) the fact that it was preceded by a pain; and b) the fact that that pain’s being a pain metaphysically necessitates its being followed by an aversive reaction. Why posit a causal relation as well? Doing so would seem to lead to a ‘strange sort of overdetermination’ (158).

Clearly, this version of the argument implicitly relies on an exclusion principle that is a cousin of the one in play in the exclusion problem. Consequently, the fate of this version of the argument is likely tied to the fate of compatibilist strategies for dealing with the exclusion problem. But there is a more important, and perhaps more obvious, maneuver the role functionalist can make to avoid the charge of overdetermination. She need not say that the two relations do not compete; she can simply say that there are not two distinct relations at all. It is not that there is a necessary relation between the pain and the aversive reaction, and then a contingent causal relation as an additional extra. Rather, there is one relation that is both causal and metaphysically necessary.<sup>18</sup> It would appear, then, that the case against the role functionalist requires denying that this is possible. It requires claiming that causal relations cannot be metaphysically necessary.

A variety of questions immediately present themselves, the most pressing of which is simply, ‘why not?’ Perhaps the answer is some version of Hume’s

dictum: there can be no necessary connections between distinct existences. However, this is not yet to fully answer the question. After all, why should we believe Hume's dictum? It is highly controversial. In this context, the most salient deniers are those who claim that the laws of nature are metaphysically necessary (Shoemaker, 'Causality and Properties'; Swoyer; see Rupert 258–60 for a discussion of their views in connection with the problem of metaphysically necessitated effects). But even setting them aside, the principle is widely flouted in contemporary philosophy. In particular, it is not clear what 'distinct' is supposed to mean. Many, many philosophers are committed to denying Hume's dictum if 'distinct' simply means 'numerically distinct'. I leave the reader to find examples in various areas of the field (see Stoljar for some relevant discussion).

It is also worth noticing that if the problem of metaphysically necessitated effects must detour through Hume's dictum, it has nothing in particular to do with mental causation. Forget about whether the role functionalist can accommodate the causal efficacy of mental properties. Causal efficacy is neither here nor there. The real issue would appear to be that the role functionalist postulates necessary connections between distinct existences. She cannot avoid violating Hume's dictum by biting the bullet and accepting epiphenomenalism.

Overall, how seriously should we take the problem of metaphysically necessitated effects? Kim has suggested that we should not fret about it too much, and should instead focus our attention on the exclusion problem (*Mind in a Physical World* 51–6). This is because the problem of metaphysically necessitated effects is only a problem for role functionalists, and thus has a narrower target than the exclusion problem. Thus worrying about it is like worrying about mosquitoes when there is a tiger nearby. However, to say that other problems are more pressing is not to solve it. Further, it is not clear whether the most promising solutions to the exclusion problem will help with the problem of metaphysically necessitated effects (see Rupert). This issue remains open; we may need that insect repellent after all.

### *Short Biography*

Karen Bennett's research is located at the intersection of metaphysics and philosophy of mind. She has published papers in these areas in *The Philosophical Review*, *Philosophical Studies*, *Noûs*, *Philosophy and Phenomenological Research*, *The Stanford Encyclopedia of Philosophy*, and a number of edited collections. She is currently working on the surprisingly difficult question of what exactly is wrong with dualism, as well as on metaontological issues about the status of certain debates in metaphysics. She has held a postdoctoral fellowship in the Research School of Social Sciences at the Australian National University, and currently teaches at Princeton University. She holds a B.A. from Oberlin College and a Ph.D. from the University of Michigan.

## Notes

\* Correspondence address: Department of Philosophy, 221 1879 Hall, Princeton University, Princeton, NJ, USA. Email: kbennett@princeton.edu.

<sup>1</sup> Some people think that this is not sufficient for physicalism, and appeal to a more robust notion of 'realization' or the like is required (Melnik; Wilson). I am not fully convinced by such arguments, though I shall not argue the point here (see 'Exclusion Again'). Notice, however, that no one denies that supervenience with metaphysical necessity is *required* for physicalism.

<sup>2</sup> Note that it recently become clear that this is *not* how Jaegwon Kim uses the labels. His 'nonreductive physicalist' only endorses supervenience with nomological necessity (*Physicalism* 49). This must be kept firmly in mind when evaluating his challenges to what he calls 'nonreductive physicalism'.

<sup>3</sup> I will typically speak of events, but those who prefer can replace 'event' with 'state' or 'process' without a problem. I will also not say anything about what exactly events *are*. That is beyond the scope of this article, and I will largely be focusing on the causal efficacy of types or properties anyway.

<sup>4</sup> It is not clear that this is what Descartes really *should* have said. Daniel Garber argues persuasively that, in light of Descartes's view about how (putatively) purely physical causation works, he should have said that mind-body causation is the basic notion of causation, in terms of which we understand *physical* causation. This would mean that 'to challenge the intelligibility of mind-body causal interaction is to challenge the whole enterprise of causal explanation' (188). However, the merits of *this* reply turn on the merits of Descartes's views about the role of God in causal interaction between bodies, and I will not take it up here.

<sup>5</sup> Note that these counterfactuals remain non-backtrackingly true on the assumption that my water-seeking-motion of my arms and legs is directly caused by physical activity in my brain. On Descartes's view, the brain activity is a link in a causal chain from my wanting to my bodily movement. Had I not wanted water, I would not have undergone the relevant pattern of neural activity.

<sup>6</sup> C.f. Kim, *Mind in a Physical World* 77–80, on whether the exclusion problem generalizes.

<sup>7</sup> Thanks to Paul Audi for discussion of this section.

<sup>8</sup> Here are three complications that I am bracketing. First, the problem from externalism about content does not obviously require denying type identity. Second, there are concerns about whether the type identity theorist has trouble accommodating causal generalizations. See the 'dilemma of reductionism' briefly discussed below. Third, many people – including some physicalists – deny *token* identity as well as type. However, this poses no challenge to the claim that only people who deny type identity have difficulties accounting for mental causation. No one denies token identity while accepting type identity. Further, I do not know of any concerns about mental causation that are truly specific to those who deny token identities. Some of the problems – notably the exclusion problem and the argument from externalism – can be formulated for the token case, but they are basically the same problem as in the type case.

<sup>9</sup> Note that Lewis's claim is weaker than Kim's. Claiming that there indeed *are* lots of second-order, multiply realizable properties is consistent with denying that natural language mental predicates pick them out. This combination of claims in fact seems to be Lewis's own view; see 'Reduction of Mind' 307).

<sup>10</sup> These views are sometimes called 'functional state identity theory' and 'functional specification theory' respectively (Block, 'What is Functionalism').

<sup>11</sup> It is at least arguable that good old-fashioned identity theorists like U. T. Place and J. J. C. Smart had something like realizer functionalism in mind. Place and Smart certainly said things like 'pain is identical to C-fiber stimulation', and Putnam's multiple realization objection was certainly taken to refute them. However, it is more charitable to interpret them as intending something like Lewis's view. For example, Smart insisted that the identity between pain and C-fiber firings was merely contingent (147). We could take this as an unfortunate pre-Kripkean failure to appreciate the necessity of identity, but we can also take it as an indication that they were not using 'pain' as a rigid designator.

<sup>12</sup> Robert Rupert calls it 'the problem of metaphysically necessary effects', but this is a little bit misleading. No one is committed to the claim that the effects are metaphysically necessary; the worry is rather that some functionalists are committed to the claim that the effects are metaphysically

necessary *conditional on the cause*. My yanking my hand off the hot stove is not metaphysically necessary, even if it is metaphysically necessitated by my pain.

<sup>13</sup> The exclusion problem can also be formulated in terms of particular events, to cause trouble for those who deny token identity. See note 8.

<sup>14</sup> Kim is not always clear about the relationship between the exclusion problem and the supervenience argument. For the record: the supervenience argument relies upon, and extends the reach of, the exclusion problem. The exclusion problem at best establishes that mental events and properties never cause anything physical. (The completeness of physics only says anything about the causes of *physical* effects.) The supervenience argument supplements the argument with the claim that physicalism is true, and thereby aims at the conclusion that mental events and properties never cause anything at all.

<sup>15</sup> In the latest version of this argument, Kim explicitly refuses to formulate any exclusion-like principle here (*Physicalism* 41n8).

<sup>16</sup> Kim adds this clarification in the 1998 formulation. The modification makes the principle much more plausible, but undermines the 1992 argument against multiply realized properties (which might well be called the 'fragmentation argument'). That argument rests upon the assumption that each instance of M have *exactly* the causal powers of its realizer P – an assumption that is not justified by the physicalist scruples to which Kim appeals. Presumably a physicalist could allow that an instance of M have *fewer* causal powers than its realizer (see Shoemaker, 'Realization and Mental Causation'). In what follows, I will only discuss the version with the 'subset' modification. The change does not affect the direct argument against the causal efficacy of the mental in quite the same way as it affects the fragmentation argument.

<sup>17</sup> Kim, in contrast, is primarily moved by the exclusion problem.

<sup>18</sup> Thanks to Brie Gertler for discussion of this point.

### Works Cited

- Antony, Louise and Joseph Levine. 'Reduction with Autonomy'. *Philosophical Perspectives* 11 (1997): 83–105.
- Armstrong, David. 'Going through the Open Door Again: Counterfactual vs. Singularist Theories of Causation'. *Causation and Counterfactuals*. Eds. J. Collins, N. Hall, and L. A. Paul. Cambridge, MA: Bradford, 2004. 445–57.
- . *A Materialist Theory of the Mind*. London: Routledge and Kegan Paul, 1968.
- Bennett, Karen. 'Exclusion Again'. *Being Reduced*. Eds. J. Hohwy and J. Kallestrup. Oxford: Oxford UP, forthcoming.
- . 'Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract It'. *Notus* 37 (2003): 471–97.
- Bieri, Peter. 'Trying out Epiphenomenalism'. *Erkenntnis* 36 (1992): 283–309.
- Block, Ned. 'Anti-Reductionism Slaps Back'. *Philosophical Perspectives* 11 (1997): 107–32.
- . 'Can the Mind Change the World?' *Meaning and Method: Essays in Honor of Hilary Putnam*. Ed. G. Boolos. Cambridge: Cambridge UP, 1990. 137–70.
- . 'What is Functionalism?' *Readings in Philosophy of Psychology, Vol. 1*. Ed. N. Block. Cambridge, MA: Harvard UP, 1980. 171–84.
- Carroll, John. 'Anti-Reductionism'. *Oxford Handbook on Causation*. Eds. H. Beebe, C. Hitchcock, and P. Menzies. Oxford: Oxford UP, forthcoming.
- Chalmers, David. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford UP, 1996.
- Davidson, Donald. 'Mental Events'. *Essays on Actions and Events*. Oxford: Clarendon Press, 1980. 207–25.
- . 'Thinking Causes'. *Mental Causation*. Eds. J. Heil and A. Mele. Oxford: Oxford UP, 1993. 3–18.
- Dowe, Phil. *Physical Causation*. Cambridge: Cambridge UP, 2000.
- Dretske, Fred. 'Mental Events as Structuring Causes of Behavior'. *Mental Causation*. Eds. J. Heil and A. Mele. Oxford: Clarendon Press, 1993. 121–36.
- . 'Reasons and Causes'. *Philosophical Perspectives* 3 (1989): 1–15.
- Fine, Kit. 'The Non-Identity of a Material Thing and Its Matter'. *Mind* 112 (2003): 195–234.

- Fodor, Jerry. 'Making Mind Matter More' (1989). *A Theory of Content and Other Essays*. Cambridge: MIT Press, 1989. 137–59.
- . 'Special Sciences: Or, the Disunity of Science as a Working Hypothesis'. *Synthese* 28 (1974): 97–115.
- . 'Special Sciences: Still Autonomous After All These Years'. *Philosophical Perspectives* 11 (1997): 149–63.
- Garber, Daniel. 'Understanding Interaction: What Descartes Should Have Told Elisabeth' (1983). *Descartes Embodied: Reading Cartesian Philosophy Through Cartesian Science*. Cambridge: Cambridge UP, 2001. 168–88.
- Gibbard, Allan. 'Contingent Identity'. *Journal of Philosophical Logic* 4 (1975): 187–221.
- Goldman, Alvin. 'The Compatibility of Mechanism and Purpose'. *The Philosophical Review* 78 (1969): 468–82.
- Hall, Ned. 'Philosophy of Causation: Blind Alleys Exposed; Promising Directions Highlighted'. *Philosophy Compass* 1 (2006): 86–94.
- Heil, John and Alfred Mele, eds. *Mental Causation*. Oxford: Clarendon Press, 1993.
- Hiddleston, Eric. 'The Reductivists' Troubles with Mental Causation'. Unpublished.
- Hornsby, Jennifer. *Simple Mindedness: In Defense of Naïve Naturalism in the Philosophy of Mind*. Cambridge, MA: Harvard UP, 1997.
- Jackson, Frank. 'Causal Roles and Higher-Order Properties'. *Philosophy and Phenomenological Research* 58 (1998): 657–61.
- . 'Epiphenomenal Qualia'. *Philosophical Quarterly* 32 (1982): 127–36.
- . 'Essentialism, Mental Properties, and Causation'. *Proceedings of the Aristotelian Society* 95 (1995): 253–68.
- . *From Metaphysics to Ethics*. Oxford: Oxford UP, 1998.
- Jackson, Frank and Philip Pettit. 'Causation in the Philosophy of Mind'. *Philosophy and Phenomenological Research* 50 (1990): 195–214.
- . 'Functionalism and Broad Content'. *Mind* 97 (1988): 381–400.
- . 'Program Explanation: A General Perspective'. *Analysis* 50 (1990): 107–17.
- Kim, Jaegwon. 'Mechanism, Purpose, and Explanatory Exclusion' (1989). *Supervenience and Mind*. Cambridge: Cambridge UP, 1993. 237–64.
- . *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: Bradford, 1998.
- . 'Multiple Realization and the Metaphysics of Reduction' (1992). *Supervenience and Mind*. Cambridge: Cambridge UP, 1993. 309–35.
- . 'The Myth of Nonreductive Physicalism' (1989). *Supervenience and Mind*. Cambridge: Cambridge UP, 1993. 265–84.
- . 'The Nonreductivist's Troubles with Mental Causation' (1993). *Supervenience and Mind*. Cambridge: Cambridge UP, 1993. 336–57.
- . *Philosophy of Mind*. 2nd ed. Cambridge, MA: Westview Press, 2006.
- . *Physicalism, or Something Near Enough*. Princeton, NJ: Princeton UP, 2005.
- . 'Postscripts on Mental Causation'. *Supervenience and Mind*. Cambridge: Cambridge UP, 1993. 358–67.
- Lepore, Ernest and Barry Loewer. 'Mind Matters'. *The Journal of Philosophy* 84 (1987): 630–42.
- Lewis, David. 'An Argument for the Identity Theory' (1966). *Philosophical Papers Volume I*. Oxford: Oxford UP, 1983. 99–107.
- . 'Causation' (1973). *Philosophical Papers Volume II*. New York, NY: Oxford UP, 1986. 159–72.
- . 'Causation as Influence'. *The Journal of Philosophy* 97 (2000): 182–97.
- . 'New Work for a Theory of Universals'. *The Australasian Journal of Philosophy* 61 (1983): 343–377.
- . 'Postscripts to "Causation"'. *Philosophical Papers Volume II*. New York, NY: Oxford UP, 1986. 172–213.
- . 'Psychophysical and Theoretical Identifications' (1972). *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge UP, 1999. 248–61.
- . 'Reduction of Mind' (1994). *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge UP, 1999. 291–324.
- . 'Review of *Art, Mind, and Religion*'. *The Journal of Philosophy* 66 (1969): 22–7.

- Loeb, Louis. *From Descartes to Hume*. Ithaca, NY: Cornell UP, 1981.
- Loewer, Barry. 'Comments on Jaegwon Kim's *Mind in a Physical World*'. *Philosophy and Phenomenological Research* 65 (2002): 655–63.
- . 'Mental Causation, or Something near Enough'. *Debates of Philosophers: Mental Causation*. Forthcoming.
- McLaughlin, Brian. 'On Davidson's Response to the Charge of Epiphenomenalism'. *Mental Causation*. Eds. J. Heil and A. Mele. Oxford: Clarendon Press, 1993. 27–40.
- Malcolm, Norman. 'The Conceivability of Mechanism'. *The Philosophical Review* 77 (1968): 45–72.
- Melnyk, Andrew. *A Physicalist Manifesto: Thoroughly Modern Materialism*. Cambridge: Cambridge UP, 2003.
- Papineau, David. 'Arguments for Supervenience and Physical Realization'. *Supervenience: New Essays*. Eds. E. E. Savellos and U. D. Yalçin. Cambridge: Cambridge UP, 1995. 226–43.
- . 'The Rise of Physicalism'. *Physicalism and Its Discontents*. Eds. C. Gillett and B. Loewer. Cambridge: Cambridge UP, 2001. 3–36.
- Pereboom, Derk. 'Robust Nonreductive Physicalism'. *Journal of Philosophy* 99 (2002): 499–531.
- Place, U. T. 'Is Consciousness a Brain Process?' *British Journal of Psychology* 47 (1956): 44–50.
- Prior, Elisabeth, Robert Pargetter, and Frank Jackson. 'Three Theses about Dispositions'. *American Philosophical Quarterly* 19 (1982): 251–7.
- Putnam, Hilary. 'Philosophy and Our Mental Life'. *Mind, Language, and Reality*. London: Cambridge UP, 1975.
- . 'Psychological Predicates'. *Art, Mind, and Religion*. Eds. W. H. Capitan and D. D. Merrill. Pittsburgh: U of Pittsburgh P, 1973. 37–48.
- Robb, David and John Heil. 'Mental Causation'. *The Stanford Encyclopedia of Philosophy*. Spring 2005 ed. Ed. Edward N. Zalta. 2005. <http://plato.stanford.edu/cite.html>
- Robinson, William. 'Epiphenomenalism'. *The Stanford Encyclopedia of Philosophy*. Spring 2003 ed. Ed. Edward N. Zalta. 2003. <http://plato.stanford.edu/cite.html>
- Rupert, Robert. 'Functionalism, Mental Causation, and the Problem of Metaphysically Necessary Effects'. *Noûs* 40 (2006): 256–83.
- Salmon, Wesley. *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton UP, 1984.
- Schaffer, Jonathan. 'Causation and Laws of Nature: A Reductionist Approach'. *Contemporary Debates in Metaphysics*. Eds. T. Sider, D. Zimmerman, and J. Hawthorne. Oxford: Blackwell, Forthcoming.
- Shoemaker, Sydney. 'Causality and Properties' (1980). *Identity, Cause, and Mind*. Oxford: Oxford UP, 2003. 206–33.
- . 'Realization and Mental Causation'. *Physicalism and Its Discontents*. Eds. C. Gillett and B. Loewer. Cambridge: Cambridge UP, 2001. 74–98.
- Smart, J. J. C. 'Sensations and Brain Processes'. *The Philosophical Review* 68 (1959): 141–56.
- Stoljar, Daniel. 'Distinctions in Distinction'. *Being Reduced*. Eds. J. Hohwy and J. Kallestrup. Oxford: Oxford UP, Forthcoming.
- Swoyer, Chris. 'The Nature of Natural Laws'. *Australasian Journal of Philosophy* 60 (1982): 203–23.
- Tooley, Michael. *Causation: A Realist Approach*. Oxford: Clarendon Press, 1987.
- . 'Causation: Reductionism vs. Realism' (1990). *Causation*. Eds. E. Sosa and M. Tooley. Oxford: Oxford UP, 1993. 172–92.
- Wilson, Jessica. 'Supervenience-Based Characterizations of Physicalism'. *Noûs* 29 (2005): 426–59.
- Yablo, Stephen. 'Cause and Essence'. *Synthese* 93 (1992): 403–49.
- . 'Mental Causation'. *The Philosophical Review* 101 (1992): 245–80.
- . 'Wide Causation'. Ed. James Tomberlin. *Philosophical Perspectives* 11 (1997): 251–81.