

Exclusion Excluded*

Brad Weslake[†]

July 30, 2009

*Earlier versions of this paper were presented to a Mellon Mental Causation Workshop at Syracuse University, Friday 30 November–Sunday 2 December 2007, and to the CUNY Cognitive Science Symposium, 14 November 2008. I am grateful to both audiences, and especially to Karen Bennett, Tony Dardis and David Rosenthal, for feedback on those occasions. I also thank Arif Ahmed, Ben Blumson, David Braddon-Mitchell, Michael Brent, John Campbell, Brandon Carey, Sommer Hodson, Adam Kay, David Macarthur, Kevin McCain, Huw Price, Nandi Theunissen and especially Peter Menzies for discussion or feedback on earlier versions.

[†]Department of Philosophy
University of Rochester
Box 270078
Rochester, NY 14627-0078
bradley.weslake@rochester.edu
<http://mail.rochester.edu/~bweslake/>

I Introduction

I take the exclusion problem to be the problem of providing a principled reason to reject at least one of the following inconsistent claims:

Nonreductionism Mental properties are distinct from physical properties.

Completeness Every event has a complete causal explanation in terms of physical properties.

Exclusion If an event has a complete causal explanation in terms of one set of properties then it has no causal explanation in terms of any other properties, unless it is causally overdetermined.

Mental Causation There exist causal explanations of events in terms of mental properties that are not cases of overdetermination¹.

By a principled reason, I mean a reason that does not simply consist in finding one of the claims to be the weakest of the group. That is, I take it as a constraint on an acceptable solution to the problem that it provide independent reason to deny one of the claims. Understanding the problem in this way immediately eliminates much of the voluminous literature on the exclusion problem as simply irrelevant, since so much of that literature consists in arguments from *Exclusion* to the rejection of *Nonreductionism* by way of *Completeness* and *Mental Causation*—and *vice versa*. It is tempting to describe this as an interminable oscillation, or at least a very tedious game of modus ponens *versus* modus tollens. What is required is a reason to abandon this dialectic altogether.

This dialectic does, however, serve to identify the two claims that are most likely to receive independent reasons for rejection. *Completeness* is both an extremely well confirmed empirical claim and a regulative ideal of fundamental physical enquiry, while *Mental Causation* is arguably an ineliminable component of our world-view,

¹Nothing here turns on my formulation of the issue in terms of explanation, properties and events. Instead of “complete causal explanation” we could substitute some appropriate causal closure principle, while instead of properties and events we could substitute whatever we think the relata are for explanation or causation. I choose to formulate the issue in this way because: i) talking about causal explanation renders the issue independent of any particular conception of causation; ii) many philosophers have worried about the causal relevance of properties in particular; iii) it is events such as actions that are the paradigmatic effects of mental causes. I defer discussion of completeness and overdetermination to §2. For the duration of the paper I will use “explanation” as a factive term, and assume that in a causal explanation all explanans properties are causally relevant to the explanandum.

one required to sustain our very conception of ourselves and others as agents². *Nonreductionism* and *Exclusion* are by comparison far more philosophically loaded claims, and are thereby the most plausible candidates for rejection.

Some principled reasons have been offered for rejecting *Nonreductionism*—reasons that go independently of the advantages the rejection would bring to the exclusion problem. Perhaps irreducible properties would be nomological danglers (Feigl 1958), or would figure in correlations that suffer at the hands of an inference to the best explanation (Hill 1991), or are not required to do the work in explanations of phenomenal consciousness that anti-reductionist arguments allege (Loar 1990; Papineau 2002; 2007). But it is uncommon to hear the claim that reductionism is independently plausible and therefore a reason to reject the exclusion problem—after all, the very bite of the problem is that it threatens to attack a non-reductionist position that many philosophers of mind would like to retain. Kim (1998, p. 60) concludes that we “might as well [...] pay the price” of adopting a reductionist metaphysics of mind—not that the price has already been paid, or that we are misled in thinking there is a price. Even the supporters of reductionism see themselves as biting hard metaphysical bullets, not as diagnosing a mistaken assumption.

For this reason it is clear that it is *Exclusion* that we would like to have independent reason to reject. Following Bennett (2003, p. 473), I will refer to the view which rejects *Exclusion* as *causal compatibilism*. I am far from the first to urge that causal compatibilism is the most promising strategy for resolving the exclusion problem³. Nor am I the first to argue—as I will—that attention to causal explanation in science promises to provide the resources appropriate to the task⁴. However, compatibilist claims have typically been made in the absence of a well-motivated framework for thinking about causation and causal explanation, and to that extent have yet to be made good. In particular, compatibilists have yet to provide a principled reason for thinking of overdetermination in a way that justifies the rejection of *Exclusion*⁵. I will be arguing that reasons of exactly this character can be found.

²I say “arguably” because of the appearance of overdetermination in the way I have formulated the claim. That there are causal explanations of events in terms of mental properties is certainly an ineliminable component of our world-view; that these explanations are not always cases of overdetermination is less obviously so.

³Bennett (2003) cites Goldman (1969), Blackburn (1991), Pereboom and Kornblith (1991), Yablo (1992), Burge (1993), Mellor (1995, pp. 103–104), Horgan (1997), Noordhof (1997) and Yablo (1997), to which we can add van Gulick (1992), Baker (1993), Menzies (2003), Ross and Spurrett (2005), Shapiro and Sober (2007) (see also Shapiro *forthcoming*) and Woodward (2008).

⁴Closest to the strategy I pursue are Shapiro and Sober (2007) and Woodward (2008).

⁵Here I agree with Bennett (2003, §2).

2 Interventionism Introduced

The framework I will employ is the justly influential theory of causation and causal explanation developed by James Woodward (2003), which I will refer to as *interventionism*. Central to the interventionist framework is the notion of a causal model. A causal model is a representational device for encoding counterfactual relationships between variables. The variables in a causal model represent the causal relata. Variables must represent particulars capable of being set to different values by interventions, but there is no further metaphysical constraint on what may count as a variable (pp. III–II4). An intervention is here to be understood as an exogenous change to the value of a variable in a model, in the sense that the values of the other variables in the model are not themselves causes or effects of the change⁶. Moreover, it is required that interventions be *surgical*, in the sense that the usual causes of the variable in question are suspended, so that the value of the variable depends only on the intervention (p. 98). For a given causal model \mathcal{M} , a necessary and sufficient condition for a variable X to be a (type-level) cause of a variable Y is that there be some state of \mathcal{M} for which an intervention on X would change the value of Y (p. 15). The particular value of X in turn figures in the (token-level) causal explanation for the particular value of Y *iff* X figures in the answer to at least one interventionist *what-if-things-had-been-different-question* (*w-question*) with respect to Y (Chapter 5)⁷.

This brief sketch is already sufficient to exhibit some of the key features of interventionism. First, we have here not an analysis or reduction of causation but rather an explication of causal claims in terms of interventions. The concept of an intervention is itself clearly causal in character, and in the interventionist framework it is explicitly defined in causal terms. What is important for present purposes is that the truth of causal claims can be established independently of any such analysis or reduction—it is whether or not it is true that mental properties sometimes causally explain physical events that is at issue in the exclusion problem, not whether these explanations can be grounded in a reductionist account of causation. Second, this is a kind of counterfactual account of causation—causal claims involve what *would* happen given some particular intervention, not what *actually* or *will* happen. Third, causal claims are model-relative in the sense that they are only well-defined with respect to the variables in a particular model. Note that this is not a version of causal anti-realism. Causal claims are not made true or false by causal models, they are made true by the counterfactuals regarding experimental interventions that are represented

⁶For a formal definition see Woodward (2003, p. 98).

⁷The precise form of the *w-questions* relevant to causation and explanation will be elaborated below.

by those models⁸. Because the counterfactuals are explicitly formulated in terms of interventions, it is transparent how these can be tested empirically, which is to say, evaluated with respect to the actual world⁹. What is important from the perspective of this paper is that these counterfactuals can be straightforwardly tested, *via* the corresponding interventions, using the ordinary procedures of the sciences. The relation different causal models bear to one another may be an interesting philosophical question in many cases—this one, for example—but it is not on this framework one which has the potential to impugn the vast array of causal knowledge acquired in the course of scientific history. Fourthly, the requirements on what can count as a variable do not place any constraint on how that variable can be set, apart from its being causally independent of other variables in the model in a sense to be explicated further below.

3 Exclusion Reformulated

In the interventionist setting, the exclusion problem can be formulated more precisely as follows:

Nonreductionism_i Mental variables are distinct from physical variables.

Completeness_i There exists a complete causal model \mathcal{M}_P containing only physical variables which specifies an explanation for every event.

Exclusion_i If there exists a complete causal model specifying an explanation for an event, there exists no other causal model containing distinct variables specifying an explanation for that event, unless it is a model in which the event is causally overdetermined.

Mental Causation_i There exist causal models $\mathcal{M}_{M1}, \mathcal{M}_{M2}, \dots, \mathcal{M}_{Mn}$ containing mental variables specifying explanations of events in which those events are not causally overdetermined.

⁸This may seem obvious, but the following mistake is routinely made: “A model is in the mind. As a consequence, causality is in the mind” (Heckman 2005, p. 2).

⁹Indeed, Woodward takes this to be so transparent that he dedicates no space at all to the move from actual experiments to counterfactuals. While it is certainly the case that this reflects the attention typically given to this question by practicing scientists, this does not mean there might not be interesting questions to ask about how we do this, especially in cases where the relevant experiments are impracticable. Indeed, in my view attention to this issue has the potential to further illuminate the nature of causation (Weslake 2006).

A few words about this formulation are in order.

First, since variables are best understood as features of causal models rather than as features of the world, framing the problem in terms of variables may appear to have trivialised the issue. In order for *Nonreductionism*_i to occupy the proper role in the exclusion dialectic it needs to express a claim about the world, not about our ways of representing the world. I will suppose then that variables are distinct only if they represent distinct properties¹⁰.

Second, *Exclusion*_i requires clarification of the notion of an event being causally overdetermined in a model. I will work up to a definition of this notion by way of considering a similar proposal made by Karen Bennett (2003). Bennett (§4) proposes a necessary condition (NC) on overdetermination, which specifies that *e* is overdetermined by *c*₁ and *c*₂ only if the following counterfactuals are non-vacuously true:

$$(c_1 \& \neg c_2) \square \rightarrow e \quad (O_1)$$

$$(c_2 \& \neg c_1) \square \rightarrow e \quad (O_2)$$

This is a natural and intuitive criterion for overdetermination. As Bennett notes, the *prima facie* appeal of the two parts of this condition is “simply that they capture the reasoning we engage in when we want to distinguish cases of genuine overdetermination from cases of joint causation” (p. 477). Be that as it may, Thomson-Jones (2007) has raised a convincing counterexample to NC that he calls the *staggered firing squad*. Suppose Billy and Suzy both fire at Victim, though Suzy fires first from a long distance away (*c*₁) and Billy fires second from a short distance away (*c*₂). The bullets arrive simultaneously at the same place with the same momentum and kill Victim (*e*). This is as paradigmatic a case of overdetermination as there is. Suppose however that it is also the case that Billy has an unmanifested squeamish (or subservient) disposition, so that had Suzy not fired, Billy would have fired inaccurately and Victim would not have died. But then (*O*₂) is false—if Billy had fired and Suzy had not, Victim would not have died. However it is plausible that we still have a case of overdetermination—the presence of this unmanifested disposition can hardly make a difference to whether there is overdetermination. So NC is false¹¹.

¹⁰This is a necessary condition on distinct variables. It is not a sufficient condition, since it is also necessary that it be possible to independently intervene on the variables. It may be that the first condition follows from the second on certain natural assumptions, but it does no harm to introduce it explicitly. I am neutral on whether the conditions are jointly sufficient.

¹¹Note that the case is explicitly designed to not require a backtracking reading of any counterfactuals (cf. Bennett 2003, pp. 477–478).

Here is how to represent the situation in a causal model. Call our variables B for Billy’s firing, S for Suzy’s firing, and V for Victim’s death. Let B be 0 if Billy does not fire, 1 if he fires inaccurately, and 2 if he fires accurately. Let S be 0 if Suzy does not fire and 1 if Suzy does fire. Let V be 1 if Victim dies and 0 if Victim does not die. The equations specifying the relationships between variables in the model are¹²:

$$V := S \vee (B = 2) \tag{1}$$

$$B := S + 1 \tag{2}$$

In the actual world, $B = 2$, $S = 1$ and $V = 1$. According to the definition of *actual cause* which I will assume, the procedure for identifying actual causes is as follows¹³. First we need the notion of a *direct cause* (Woodward 2003, p. 55):

(DC) A necessary and sufficient condition for X to be a direct cause of Y with respect to some variable set V is that there be a possible intervention on X that will change Y (or the probability distribution of Y)¹⁴ when all other variables in V besides X and Y are held fixed at some value by interventions.

Second we need the notion of a *directed path* (Woodward 2003, p. 42):

(P) A *directed graph* is an ordered pair $\langle V, E \rangle$ where V is a set of vertices that serve as the variables representing the relata of the causal relation and E a set of directed edges connecting these vertices. A directed edge from vertex or variable X to vertex or variable Y means that X directly causes Y . A sequence of variables $\{V_1 \dots V_n\}$ is a *directed path* from V_1 to V_n if and only if for all $i(1 \leq i < n)$ there is a directed edge from V_i to V_{i+1} .

From here on, *path* should be read as equivalent to *directed path*. Third we need to define the notion of a *redundancy range* (Woodward 2003, p. 83–84):

¹²Note that these equations are not to be interpreted symmetrically. “:=” here stands in for an assignment of values to variables on the left hand side in the manner specified on the right hand side. “=” here is a function returning 1 if the two sides are equal and 0 if otherwise.

¹³This definition is endorsed by Hitchcock (2001), Woodward (2003) and Halpern and Pearl (2005a; 2005b), to whom the idea can be credited. See Woodward (2003, §2.7) for a nice developmental story of the motivation for the definition. As Oisín Deery reminded me, Woodward (p. 84, fn. 46) is ambivalent on the question of whether a more complicated definition is required in order to handle cases of trumping; but these details will not impact the role the condition will play in this paper.

¹⁴For the remainder of the paper, all talk of actual or counterfactual changes to variables should be interpreted with this qualification to handle the probabilistic case. Matters are simplified by leaving it implicit.

(RR) For a path P from X to Y in a causal model, define $V_1 \dots V_n$ as all variables that are not on P . Values $v_1 \dots v_n$ are on the redundancy range for V_i with respect to P if no intervention on $v_1 \dots v_n$ while holding X fixed would result in a change to the actual value of Y .

Now we can define the following necessary and jointly sufficient conditions on an *actual cause* $X = x$ of $Y = y$ relative to a causal model:

(AC₁) The actual value of $X = x$ and the actual value of $Y = y$.

(AC₂) For each directed path P from X to Y , fix by interventions all direct causes Z_i of Y that do not lie along P at some combinations of values within their redundancy range. Then, for each path from X to Y and for each possible combination of values for the direct causes Z_i of Y that are not on this route and that are in the redundancy range of Z_i , determine whether there is an intervention on X that will change the value of Y . If there is at least one such intervention, (AC₂) is satisfied.

The definition appears at first glance a complicated one, but the intuitive idea is simple: an actual cause along a path is determined by whether there is an intervention along the path that would change the effect in background conditions which are irrelevant to the effect in the circumstances. Now, in the case of Billy and Suzy, we need to apply these conditions to each potential actual cause. Starting with Suzy, it is clear that all possible values of B are in the redundancy range of S , since no intervention on B while holding S fixed would result in a change to V ¹⁵. So we are permitted to test for the efficacy of S by setting $B = 0$ or $B = 1$. But if $B = 0$ or $B = 1$, setting $S = 0$ would result in $V = 0$. So $B = 2$ is an actual cause of $V = 1$. Turning to Billy, we see that the situation is exactly symmetrical. It is clear that all possible values of S are in the redundancy range of B , since no intervention on S while holding B fixed would result in a change to V . So we are permitted to test for the efficacy of B by setting $S = 0$. But if $S = 0$, setting $B = 0$ would result in $V = 0$. So $S = 1$ is an actual cause of $V = 1$.

Given this definition of *actual cause*, the natural way to define overdetermination should now be obvious. Intuitively, an actual cause is sufficient for an effect just in case any other actual or potential causes of the effect did not make or could not have made any further difference to the effect than the difference made by the sufficient actual cause in question. This is to say that an actual cause is sufficient for an effect

¹⁵On the assumption, which is built into my specification of the model, that if Suzy fires and Billy fires inaccurately, Victim dies.

just in case the effect would have occurred regardless of the other values any other actual or potential causes of the effect could have taken. Let us say then that an actual cause along path P is a *sufficient actual cause* for an effect so long as all possible values $v_1 \dots v_n$ for all variables $V_1 \dots V_n$ not on P are on the redundancy range for V_i with respect to P¹⁶. Overdetermination is in turn defined as occurring when there are two sufficient actual causes of an effect in a model.

Note that according to this criterion S and B overdetermine V. We have already seen how it is that they are adjudged actual causes, so all that remains is to show that they are both sufficient. Recall our observation that all possible values of each variable are on the redundancy range of the other. By our definition of sufficiency then, each is sufficient. Since S and B are both sufficient actual causes of V, they overdetermine V. The notion of overdetermination in a causal model succeeds where Bennett's NC does not¹⁷.

Third, *Completeness_i* should not be interpreted as being equivalent to the claim that every event can be explained in terms of some fundamental physical theory. This is because:

- i) It is an open question whether fundamental physical theories provide causal explanations, in which case *Completeness_i* should be interpreted as referring to a complete causal model that is not simply identical to some fundamental physical theory¹⁸; and
- ii) Even if fundamental physical theories do provide causal explanations, it is not the case that the structure of fundamental physical theories is identical to the structure of interventionist causal models¹⁹.

¹⁶The generalisation to jointly sufficient causes is obvious.

¹⁷I might add that it also succeeds rather more elegantly than the ultimately unsuccessful proposals canvassed by Thomson-Jones (2007) in the course of his extremely helpful analysis of the prospects for compatibilism. The key to the solution is that the interventionist model provides us with a rationale for considering a counterfactual where we explicitly have Billy shoot accurately without Suzy shooting even though Billy would not have shot accurately had Suzy not shot. The rationale for considering this counterfactual is provided by the very notion of an intervention, which allows us to hypothetically manipulate a variable independently of how that variable would ordinarily be caused (Woodward 2008, pp. 240–241 also notes the importance of this feature of interventionism in this context).

¹⁸See Russell (1912–1913), Field (2003) and the essays in Price and Corry (2007).

¹⁹One reason is that causal models do not allow the representation of continuous processes (Strevens 2007, pp. 242–244). Strevens puts the point by saying that interventionist causal models “represent less of causal reality than is actually out there” (p. 243), which presupposes an answer to i) that it is free for an interventionist to reject. Moreover an interventionist may consistently claim both that every interventionist model omits some causal truth, and that all causal truths are represented by some interventionist model or other.

In making these points, I do not mean to weaken the support that causal completeness assumptions rightly draw from the promise of complete explanations of events in terms of fundamental physical theories. My point is simply that there is an inference involved from the success of fundamental physics to the existence of a complete causal model in the sense required to formulate the exclusion problem.

Having clarified what *Completeness_i* does not say, I should say something about what it does say. The exclusion problem is often put in terms of causal sufficiency rather than completeness, so any defensible notion of completeness must bear some close relationship to a notion of causal sufficiency. There are at least four different notions of sufficiency that can be discriminated with the interventionist framework, only one of which, I will argue, is suitable for figuring in formulations of the exclusion problem. These notions are, in order of strength:

Sufficiency in the Circumstances in a Model A cause is sufficient in the circumstances for an effect in a model if it is a non-overdetermining actual cause of the effect in that model.

Sufficiency in a Model A cause is sufficient *simpliciter* for an effect in a model if it is an actual cause of that effect and would remain so no matter how the other variables in the model were set. I intend this to be equivalent to the notion of a *sufficient actual cause* defined above.

Sufficiency in an Effectively Closed Model Call a model \mathcal{M}_F framed in variables characterised by vocabulary F *effectively closed* with respect to variables characterised by vocabulary G with respect to an effect *iff* the range of *w-questions* answered by \mathcal{M}_F for the effect cannot be increased by adding variables from G²⁰. A cause is sufficient for an effect in an effectively closed model \mathcal{M}_F with respect to G if it is sufficient *simpliciter* for the effect in \mathcal{M}_F .

Nomological Sufficiency An event A is nomologically sufficient for an event B if the occurrence of A and the laws of nature together guarantee that B will occur, or fix a probability for B such that there are no further events conditioning on which would change the probability of B²¹.

²⁰To keep our emphasis on metaphysical questions, let us suppose that different vocabularies correspond to different types of properties. It is a standard presupposition here that there exists some relatively clear distinction between “physical properties” and “mental properties”. I make no stand on how these property types are to be demarcated.

²¹Note that the relevant notion of event here must be liberal enough to allow events involving all physical properties instantiated across the entire cross-section of a lightcone in spacetime, if any events are going to turn out to be nomologically sufficient for any others (Field 2003). This is a strong reason

It is important to note a difference between the first two definitions and the second two. The first two are defined in a wholly model-internal manner, while the second two involve model-external facts. Indeed, the final notion of *Nomological Sufficiency* makes no reference to causal models at all. Since I am proceeding under the interventionist assumption that causation is to be defined with respect to models, this notion is therefore a non-starter for formulating the exclusion problem. The first two model-internal notions of sufficiency on the other hand are non-starters because they do not even pose a *prima facie* problem for mental causation. Suppose that we have a model \mathcal{M}_P framed in variables characterised by physical vocabulary P which is not effectively closed with respect to variables characterised by mental vocabulary M . Suppose further that \mathcal{M}_P specifies a cause that is sufficient *simpliciter* for some action. Since \mathcal{M}_P is not effectively closed, there is simply no problem in supposing that there exists some expansion of \mathcal{M}_P to incorporate variables characterised by M so that the M -variables are sufficient in the circumstances for the action. After all, by definition these variables provide answers to at least one additional *w-question*, and therefore in that sense make some non-overdetermining actual difference to the explanation of the action. I conclude that if the exclusion problem is to arise in the interventionist framework, the relevant closure principle must be sufficiency in an effectively closed model.

Understanding the problem in this way justifies the great variability in the way completeness assumptions are formulated. Sometimes the worrying complete or sufficient explanation is supposed to be provided by physics, sometimes by biology, sometimes by neuroscience, sometimes by (at least the “syntactical” explanations appearing within) cognitive science. In my view the exclusion argument can be posed in terms of these different sciences precisely because it is reasonable to believe that there exist sufficient causes, in models framed in variables drawn from the vocabularies of each of these sciences, which are effectively closed with respect to M -variables. If I have all of the physical information about you my predictions would not be better enabled by knowing any further mental information about you—and likewise if I have all of the biological information, or all of the neuroscientific information, or all of the (“syntactic”) cognitive scientific information. Moreover once we understand completeness in the way I have suggested, it can be seen that the exclusion argument generalises—the physical causal model is effectively closed with respect to the variables of the biological causal model, the biological causal model is effectively closed

to think of nomological sufficiency as completely irrelevant to the exclusion problem, since none of the causes we identify in any sciences apart from the more theoretical arms of fundamental physics will turn out to be events in this sense.

with respect to the variables of the neuroscientific causal model, and so on up the hierarchy of the sciences and never *vice versa*²². And so if the exclusion argument works against mental variables it also works against any variables not appearing in some maximally effectively closed causal model²³.

Finally, note that my definitions of *Exclusion_i* and *Mental Causation_i* do not require that in order for mental variables to causally explain, they must be sufficient *simpliciter* for their effects. It is unclear to me why the exclusion problem is often framed so that mental causes must be sufficient for their effects in a stronger sense than sufficiency in the circumstances. Bennett (2003, §5) thinks that anything less than sufficiency would endanger the “full-fledged causal efficacy of the mental” (p. 481), granting it merely “a derivative efficacy” (p. 482). I cannot see the motivation for claims of this form if sufficiency is supposed to be stronger than circumstantial sufficiency—especially given the metaphors that are often used to characterise the exclusion problem²⁴. If an event has a complete physical cause, mental causes are often said to have “no work left to do” (Kim 1998, p. 35, 37, 54, 110, 126 n. 6), “no gaps left to fill” (Menzies 2003), no opportunity to “inject themselves” into the causal order (Kim 1998, p. 41; 2005, p. 16); if there is no lowest level of causation, we are supposed to worry that causal powers will “drain away” (Block 2003; Kim 2003). If there is work left to do, a gap to be filled, an injection to be provided or a drain to be plugged, presumably the context is almost sufficient, and the additional impetus will render the context wholly sufficient. The work, filler, injection or plug will not itself be wholly sufficient, but rather will be sufficient in the circumstances. Now perhaps these are all just poor metaphors for what is supposed to be at issue here; but metaphors aside, the claim in question would be that any actual causes that are not sufficient causes *simpliciter* must have merely derivative efficacy. Given that sufficiency in the circumstances is the sort of efficacy most causes have in most sciences, I say that derivative causes in this idiosyncratic sense would be causes enough for mental causation²⁵.

4 Compatibilism Motivated

My final formulation of the exclusion problem in the interventionist setting now has this shape:

²²It is a *hierarchy* in part *because* this relation is asymmetric in this way.

²³Here I side with Bontly (2002) against Kim (1997; 1998).

²⁴I do not suggest Bennett endorses the position I here criticise.

²⁵That said, the position I will be defending does not turn on taking a stand on this point, since it will work equally well against the exclusion problem formulated in terms of either notion of sufficiency.

*Nonreductionism*_j Mental variables are distinct from physical variables in the sense that they are drawn from distinct vocabularies M and P.

*Completeness*_j There exists an effectively closed causal model \mathcal{M}_P with respect to M which specifies a sufficient actual cause for every event.

*Exclusion*_j If there exists an effectively closed causal model \mathcal{M}_F with respect to variables characterised by vocabulary G which specifies a sufficient actual cause for an event, there exists no other causal model \mathcal{M}_G specifying an actual cause for that event in terms of variables characterised by vocabulary G, unless it is a model in which the event is causally overdetermined.

*Mental Causation*_j There exist causal models $\mathcal{M}_{M1}, \mathcal{M}_{M2}, \dots, \mathcal{M}_{Mn}$ containing mental variables specifying actual causes of events in which those events are not causally overdetermined.

The solution to the exclusion problem now afforded by the interventionist framework is very simple, but will require an extended defence. The solution is simply that *Exclusion*_j is not a principle that is entailed by the interventionist theory of causation. Indeed, I will argue below that interventionism, when combined with a certain constraint on the connection between G properties and F properties, entails the rejection of *Exclusion*_j. So if the interventionist theory of causation is correct, we have no argument from the nature of causation to the exclusion premise. If we have an independently attractive account of causation in which the exclusion premise is not motivated, we have a principled reason to maintain the truth of causal compatibilism. So interventionism provides a principled way to retain mental causation in the face of the exclusion problem—exactly the solution we set out to find. In the remainder of this section I defend these claims.

As I suggested above, it is reasonable to believe that *Completeness*_j can be justified in terms of variables from a number of different vocabularies, including physical, biological, neuroscientific and cognitive scientific vocabularies. The justification for *Mental Causation*_j I provided was that it is required for us to sustain a conception of ourselves and others as agents. In addition, it is highly plausible that it is quite generally a presumption of our folk-psychological explanatory practice. Now there are issues concerning whether this practice can reasonably be described as involving a commitment to an interventionist causal model of the kind specified by *Mental Causation*_j, and if so whether the model is veridical, but I will not defend or rely on these claims here²⁶. Let us suppose for the sake of argument, however, that \mathcal{M}_{M1} is

²⁶Part of the key is to see that there are no special issues arising in the case of folk psychological explanation that do not arise quite generally for understanding our explanatory practices in interven-

a model specified by folk psychology. Now suppose that there is no overlap in the variables appearing in \mathcal{M}_P and the variables appearing in \mathcal{M}_{M_I} , apart from a set of variables A representing a range of possible actions²⁷. Suppose also, as is reasonable, that the range of *w-questions* concerning A answered by \mathcal{M}_P is greater than the range answered by \mathcal{M}_{M_I} . The important point is that none of these suppositions is inconsistent with supposing that both \mathcal{M}_P and \mathcal{M}_{M_I} specify non-overdetermining actual causes for A . Intuitively, the threat of overdetermination is the threat that physical and mental causes overdetermine actions. Overdetermination however, as I have defined it in the interventionist framework, is a model-internal notion. So the fact that there are two models providing actual causes for some event, even if both provide sufficient actual causes for that event, does not entail that the event is overdetermined. And there is no principle in the interventionist framework disallowing the possibility of two such models.

I see two ways in which this picture might be challenged, and the exclusion argument resurrected. The first is to deny that it is possible for there to be causal models which stand in this relationship to each other. It might be argued that only \mathcal{M}_P *really* specifies causes, and that \mathcal{M}_{M_I} merely specifies *explanations*, or some other weak cousin of causation. This might be because only \mathcal{M}_P promises to be maximally predictively accurate and therefore maximally effectively closed (Davidson, 1963; 1967; 1970; 1995), or because \mathcal{M}_P is a truthmaker for \mathcal{M}_{M_I} (Robb and Heil 2003, §5.4)²⁸, or for more recherché metaphysical reasons (Jackson and Pettit, 1988; 1990a; 1990b). I think that the arguments for these claims are universally unsound, but for present purposes want simply to note that if they succeed they are arguments against interventionism in general and therefore should be addressed as independent claims about the nature of causation and causal explanation. Proceeding under the presumption of the truth of interventionism, I here leave them to one side²⁹.

The second challenge is to concede that there exist models standing in this relationship to each other, but to argue that a further constraint must be satisfied in

tionist terms. See Godfrey-Smith (2005) for a general discussion of the idea of folk psychological competence as involving facility with a model.

²⁷This supposition again shows that \mathcal{M}_P ought not to be understood as provided by some fundamental physical theory, since these say nothing about actions as described in the vocabulary of folk psychology. Some have wished to press this point into a defence of mental causation (see for example Gibbons 2006), but I find the strategy unpromising and assume that there exists some such model \mathcal{M}_P . If you are suspicious of actions, substitute behaviours.

²⁸Something like this worry seems to haunt Kim's discussion in many places, and I have encountered it repeatedly in discussion, but arguments for the claim do not often appear explicitly in the literature.

²⁹See Burge (2007b) and Woodward (2008, pp. 244–249) for arguments against some of these lines of objection. The weakness in each case is that the relevant metaphysical assumption is ill-supported.

order for \mathcal{M}_{M_I} to specify genuine non-overdetermining causes. Call the constraint the *non-embedding* constraint. The non-embedding constraint is that there must not be a causal model \mathcal{M}_{PM} containing variables drawn from both \mathcal{M}_P and \mathcal{M}_{M_I} in which the P-variables and M-variables are overdetermining causes in \mathcal{M}_{PM} , or in which the M-variables are not causes in \mathcal{M}_{PM} . Surely, the challenge suggests, the existence of such a model would undermine the claims of \mathcal{M}_{M_I} to have identified non-overdetermining causes. I agree—this is a plausible constraint, and I accept it. I will argue however that in the case of \mathcal{M}_P and \mathcal{M}_{M_I} the constraint is met, since under the assumption that a certain form of supervenience, to be further specified below, obtains, *there is no such causal model* \mathcal{M}_{PM} ³⁰.

Recall the two necessary conditions on the existence of distinct variables in a causal model introduced in §2 and §3. These were:

Property Distinctness Two variables are distinct in a model only if they represent distinct properties.

Independent Manipulability Two variables are distinct in a model only if they can be independently set to different values by interventions.

The crucial condition here is *Independent Manipulability*. This condition is required in order to determine the relationships of direct causation in a model. To determine whether X is a direct cause of Y requires, by the definition of (DC) from §3, that there be a possible intervention on X that will change Y when all other variables are held fixed at some value by interventions³¹. For there to be a possible intervention of this kind, it must be possible to directly cause X without directly causing Y, with all other variables held fixed. This is the sense of independence required for *Independent Manipulability*.

Now suppose for *reductio* that we incorporate all variables drawn from P and M into the same model \mathcal{M}_{PM} . Assume moreover that we add to *Nonreductionism*₁ the claim that the mental variables at least metaphysically supervene on the physical variables, as is standardly taken to be definitive of non-reductive physicalism³². Take some physical variable P_x and some mental variable M_x that we wish to test for overdetermination. Since the potential overdetermining causes that the exclusion argument

³⁰In this sense my solution to the exclusion problem is analogous to maintaining that Bennett's (O₁) and (O₂) are vacuous. This paper can be seen as an initial step towards reformulating and defending Bennett's form of compatibilism in the context of the interventionist theory of causation.

³¹In interpreting the condition in this way I agree with Baumgartner (2009).

³²While I will assume the relation is one of supervenience, what is crucial to the argument is the fact that it is a relation of metaphysical necessity.

aims to identify are typically the realised and realiser properties, or the supervening and subvening properties, invoked by some theory of the mind, I will assume further that the values of P_x are at least metaphysically sufficient for the values of M_x , as asserted by the supervenience claim. The question of whether these variables can be incorporated into the same causal model now turns on whether they are independent in the sense required by *Independent Manipulability*. This is to say that in order for P_x and M_x to appear in a single causal model, it must be possible to directly cause P_x without directly causing M_x and *vice versa*³³. And here is the crux of the matter. Since the relevant notion of interventional possibility is metaphysical possibility, this will only be the case if the correlations between the physical variables and mental variables obtain with *less* than metaphysical necessity. However we assumed *at least* the metaphysical supervenience of the mental variables on the physical variables. So any metaphysics of mind endorsing at least metaphysical supervenience will be in a position to deny that the mental and physical variables are independently manipulable. They will do so because there will be no independent interventions on M_x that are not also interventions on P_x ³⁴. So assuming the form of supervenience definitive of the non-reductive physicalist, there is no such causal model \mathcal{M}_{PM} ³⁵. Non-reductive physicalism and causal compatibilism are, on the interventionist framework, *made for each other*³⁶.

³³No doubt there are many physical and mental variables that can be manipulated independently of each other in this way. I can change my mind about whether to shoot or pass the ball independently of whether my arm is moving this way or that. But it is not in general these sorts of variables that we take to be the potential overdetermining causes at stake in the exclusion problem. The variables we take to be the potential overdetermining causes are precisely the supervening and subvening properties, the realised and realiser properties, and so on.

³⁴Note that there may well be interventions on P_x that are not interventions on M_x , since it is implausible that every fine grained change at the subvening physical level makes for some difference at the supervening mental level. It is enough for the argument that the reverse principle is true.

³⁵Here I agree with Shapiro and Sober (2007), who write: “It is not relevant, or even coherent, to ask what will happen if one wiggles X while holding fixed the micro-supervenience base of X”. See also Woodward (2008) and Shapiro (forthcoming).

³⁶Many have supposed that the exclusion argument has variable force depending on the form of non-reductionism one adopts, with the dualist in a dire situation and the non-reductive physicalist in a precarious if ultimately tenable one (Stoljar 2008; Bennett 2008). The argument I have given agrees that the non-reductive physicalist has an attractive position available. However, for reasons I do not have the space to defend here, I believe that the dualist who accepts nomological supervenience also has an attractive option. All such a dualist needs to defend, in order to adopt the solution I have motivated, is the view that interventions must be nomologically possible rather than metaphysically possible. While Woodward’s own view supports the non-reductive physicalist (Woodward 2003, pp. 127–133) there are reasons to think that he is mistaken (Weslake 2006 and Price and Weslake forthcoming). I leave this line of argument for another time.

Still, we might wonder whether it is always reasonable to assume that the sufficient actual causes of A identified by \mathcal{M}_P will always be precisely those that necessitate the actual causes of A identified by \mathcal{M}_{M_I} . So let us suppose that P_x is not metaphysically sufficient for M_x , say because M_x supervenes not only on P_x but on P_x and P_y together. This may be the situation if for example P_x is a variable representing the total internal physical state of a person, P_y a variable representing the total external physical state of the world, and content externalism is true. Is there a possible causal model that embeds all of these variables? The *Independent Manipulability* condition requires that there be a possible intervention on each of these variables when *all* other variables are held fixed. But if P_y and P_x are held fixed, there will be no independent intervention on M_x , since to suppose otherwise would be inconsistent with the supervenience of M_x on P_y and P_x . The upshot is the following necessary condition on variable distinctness in a model, which I have argued follows from *Independent Manipulability*:

Non-Supervenience A causal model cannot contain a variable with a possible value that supervenes on a possible combination of values for any other variables in the model.

Interventionism requires that the variables in a model satisfy *Independent Manipulability*, which I have argued entails *Non-Supervenience*. *Non-Supervenience* in turn prevents the variables from \mathcal{M}_P and \mathcal{M}_{M_I} from being incorporated into a single model, and so the non-embedding constraint is met. Given that the non-embedding constraint is met, we are justified in rejecting *Exclusion_j*³⁷.

Before concluding, I will make explicit an assumption on which my argument has depended, and indicate an open question that arises when the assumption is dropped. Call the assumption the *subvenience-sufficiency* (SS) assumption. The (SS) assumption is that at least one variable in \mathcal{M}_{M_I} has a possible value that is metaphysically necessitated by the possible values of one or more variables in \mathcal{M}_P . This will be the case

³⁷Baumgartner (2009; forthcoming) has alleged that Woodward faces a dilemma concerning *Independent Manipulability*. Either he accepts the condition or he does not. If he does not, it can be shown that by his own lights M_x does not cause A . If he does, then since there is no causal model \mathcal{M}_{P_M} , M_x does not cause A . Therefore M_x does not cause A . The error is with the inference of the second horn. For M_x to cause A it is not required that M_x and A be capable of appearing in any candidate causal model we dream up—it is simply required that there exist some causal model in which an intervention on M_x would change A . The non-existence of \mathcal{M}_{P_M} has no implication regarding the existence of \mathcal{M}_{M_I} ; Baumgartner has overlooked the model relativity of interventionist causal claims. Baumgartner (2009) further alleges that Woodward changed his mind concerning *Independent Manipulability*, rejecting it in Woodward (2003) and endorsing it in Woodward (2008). In my view a reasonable reading of Woodward has *Independent Manipulability* implicit in the earlier work.

whenever the M-properties supervene on the P-properties that are causally sufficient for A in \mathcal{M}_P , or when \mathcal{M}_P explicitly includes all of the P-properties on which the M-properties supervene. I have argued that when this assumption is granted, the non-embedding constraint is satisfied, and that we may therefore reject *Exclusion₃*. However if we drop this assumption there remains a final challenge. Let us retain our assumption that it is only P_x and P_y together that are metaphysically sufficient for M_x , and further suppose that P_x alone is causally sufficient for A in a model \mathcal{M}_{P^*} that does not contain variable P_y . Surely, the challenge alleges, there exists a causal model \mathcal{M}_{PM^*} containing M_x and P_x but *not* P_y . Indeed, *Independent Manipulability* is satisfied here. Because P_y is omitted from the model, we can hold P_x fixed and change M_x by changing P_y . And in the other direction, we can hold M_x fixed and change P_x just in case the possible values of P_x include other subsets of subvenience bases of M_x . Moreover, given that \mathcal{M}_{P^*} is effectively closed with respect to M, M_x is at risk of being rendered epiphenomenal in \mathcal{M}_{PM^*} . If so, the non-embedding constraint fails for models \mathcal{M}_{P^*} and \mathcal{M}_{M_I} . Moreover, the challenge goes, this will be generally true for sufficient physical causes that are subsets of the subvenience bases of mental causes. The exclusion problem would still have bite, if only against—for example—externally individuated mental states for which a subset of their subvenience base is sufficient for some of their effects. Still, perhaps this situation is pervasive, because all mental states are externally individuated (Fisher 2007), or for other reasons (Bennett 2003, pp. 484–485), so the challenge is a serious one. The issues the challenge raises, however, require a degree of detail that I do not here have the space to provide. In particular, in my view answering the challenge requires making use of the contrastive nature of causation as understood within the interventionist framework. For the present paper, I am content to have demonstrated that when the (SS) assumption is made, the interventionist theory of causation licenses the rejection of *Exclusion₃*.

5 Conclusion

To solve the exclusion problem is to find independent reason to reject at least one premise on which it is founded. Interventionism is an independently attractive theory of causation that, I have argued, provides reasons of exactly this character. For interventionism entails that for all mental states that are metaphysically necessitated by physical states sufficient for their effects, the exclusion premise is false. Causal compatibilism can be adopted not merely because it is demanded by our preferred metaphysics of mind, but because it is entailed by interventionism.

References

- Baker, Lynne Rudder. 1993. "Metaphysics and Mental Causation", in *Mental Causation*, edited by John Heil and Alfred Mele, Oxford University Press, Oxford, pp. 75–96. URL: <http://www.people.umass.edu/lrb/files/bak93metS.pdf>.
- Baumgartner, Michael. 2009. "Interventionist Causal Exclusion and Non-reductive Physicalism", in *International Studies in the Philosophy of Science*, Vol. 23, No. 2, July 2009, pp. 161–178. URL: <http://dx.doi.org/10.1080/02698590903006909>.
- . forthcoming. "Interventionism and Epiphenomenalism", in *Canadian Journal of Philosophy*, Vol. URL: http://www.staff.unibe.ch/baumgart/docs/inter_epi.pdf.
- Bennett, Karen. 2003. "Why the Exclusion Problem Seems Intractable, and How, Just Maybe, To Tract It", in *Noûs*, Vol. 37, No. 3, September 2003, pp. 471–497. URL: <http://dx.doi.org/10.1111/1468-0068.00447>.
- . 2008. "Exclusion Again", in *Being Reduced: New Essays on Reduction, Explanation and Causation*, edited by Jesper Kallestrup and Jakob Hohwy, Oxford University Press, Oxford, pp. 280–305.
- Blackburn, Simon. 1991. "Losing Your Mind: Physics, Identity, and Folk Burglar Prevention", in *The Future of Folk Psychology: Intentionality and Cognitive Science*, edited by John D. Greenwood, Cambridge University Press, Cambridge, pp. 196–225. Reprinted in Blackburn (1993, pp. 229–254).
- . 1993. *Essays in Quasi-Realism*, Oxford University Press, Oxford.
- Block, Ned. 1980. *Readings in Philosophy of Psychology*, edited by Ned Block. Vol. 1. Harvard University Press, Cambridge MA.
- . 2003. "Do Causal Powers Drain Away?", in *Philosophy and Phenomenological Research*, Vol. 67, No. 1, July 2003, pp. 133–150. URL: <http://dx.doi.org/10.1111/j.1933-1592.2003.tb00029.x>.
- Bontly, Thomas D. 2002. "The Supervenience Argument Generalizes", in *Philosophical Studies*, Vol. 109, No. 1, May 2002, pp. 75–96. URL: <http://dx.doi.org/10.1023/A:1015786809364>.
- Burge, Tyler. 1993. "Mind-Body Causation and Explanatory Practice", in *Mental Causation*, edited by John Heil and Alfred Mele, Oxford University Press, Oxford, pp. 97–120. Reprinted with postscript in Burge (2007a, pp. 344–362).
- . 2007a. *Foundations of Mind*, Vol. 2. Philosophical Essays. Oxford University Press, Oxford.

- Burge, Tyler. 2007*b*. “Postscript: Mind-Body Causation and Explanatory Practice”, in *Foundations of Mind*. Vol. 2, Philosophical Essays, Oxford University Press, Oxford, pp. 363–382.
- Chalmers, David J. 2002. *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers. Oxford University Press, Oxford.
- Clark, Andy and Peter Millican. 1999. *Connectionism, Concepts, and Folk Psychology: The Legacy of Alan Turing*, edited by Andy Clark and Peter Millican. Vol. 2. Oxford University Press, Oxford.
- Davidson, Donald. 1963. “Actions, Reasons, and Causes”, in *The Journal of Philosophy*, Vol. 60, No. 23, November 1963, pp. 685–700. Reprinted in Davidson (2001, pp. 3-19). URL: <http://dx.doi.org/10.2307/2023177>.
- . 1967. “Causal Relations”, in *The Journal of Philosophy*, Vol. 64, No. 21, November 1967, pp. 691–703. Reprinted in Davidson (2001, pp. 149-162).
- . 1970. “Mental Events”, in *Experience and Theory*, edited by Lawrence Foster and Joe William Swanson, University of Massachusetts Press, Amherst MA, pp. 79–101. Reprinted in Block (1980, pp. 107–119) and Moser and Trout (1995, pp. 11–126) Davidson (2001, pp. 207-227). URL: <http://dx.doi.org/10.1093/0199246270.003.0011>.
- . 1995. “Laws and Cause”, in *Dialectica*, Vol. 49, No. 2-4, June 1995, pp. 263–279.
- . 2001. *Essays on Actions and Events*, 2nd edition. Oxford University Press, Oxford. URL: <http://dx.doi.org/10.1093/0199246270.001.0001>.
- Feigl, Herbert. 1958. “The “Mental” and the “Physical””, in *Concepts, Theories, and the Mind-Body Problem*, edited by Herbert Feigl, Michael Scriven, and Grover Maxwell. Vol. 2, Minnesota Studies in the Philosophy of Science, University of Minnesota Press, Minneapolis, pp. 370–497. Reprinted in Feigl (1968).
- . 1968. *The “Mental” and the “Physical”: The Essay and a Postscript*, University of Minnesota Press, Minneapolis.
- Feigl, Herbert and May Brodbeck. 1953. *Readings in the Philosophy of Science*, edited by Herbert Feigl and May Brodbeck. Appleton-Century-Crofts, New York.
- Field, Harry. 2003. “Causation in a Physical World”, in *The Oxford Handbook of Metaphysics*, edited by Michael J. Loux and Dean W. Zimmerman, Oxford University Press, Oxford, pp. 435–460. URL: <http://philosophy.fas.nyu.edu/docs/IO/1158/Cause.pdf>.
- Fisher, Justin C. 2007. “Why Nothing Mental is Just in the Head”, in *Noûs*, Vol. 41, No. 2, June 2007, pp. 318–334. URL: <http://dx.doi.org/10.1111/j.1468-0068.2007.00649.x>.

- Gibbons, John. 2006. "Mental Causation without Downward Causation", in *The Philosophical Review*, Vol. 115, No. 1, January 2006, pp. 79–103. URL: <http://dx.doi.org/10.1215/00318108-115-1-79>.
- Godfrey-Smith, Peter. 2005. "Folk Psychology as a Model", in *Philosophers' Imprint*, Vol. 5, No. 6, November 2005, pp. 1–16. URL: <http://www.philosophersimprint.org/005006/>.
- Goldman, Alvin I. 1969. "The Compatibility of Mechanism and Purpose", in *The Philosophical Review*, Vol. 78, No. 4, October 1969, pp. 468–482.
- Halpern, Joseph Y. and Judea Pearl. 2005a. "Causes and Explanations: A Structural-Model Approach. Part I: Causes", in *British Journal for the Philosophy of Science*, Vol. 56, No. 4, December 2005, pp. 843–887. URL: <http://dx.doi.org/10.1093/bjps/axi147>.
- . 2005b. "Causes and Explanations: A Structural-Model Approach. Part II: Explanations", in *British Journal for the Philosophy of Science*, Vol. 56, No. 4, December 2005, pp. 889–911. URL: <http://dx.doi.org/10.1093/bjps/axi148>.
- Heckman, James J. 2005. "The Scientific Model of Causality", in *Sociological Methodology*, Vol. 35, No. 1, August 2005, pp. 1–98. URL: <http://dx.doi.org/10.1111/j.0081-1750.2006.00163.x>.
- Heil, John and Alfred Mele. 1993. *Mental Causation*, edited by John Heil and Alfred Mele. Oxford University Press, Oxford.
- Hill, Christopher S. 1991. *Sensations: A Defense of Type Materialism*, Cambridge University Press, Cambridge.
- Hitchcock, Christopher. 2001. "The Intransitivity of Causation Revealed in Equations and Graphs", in *The Journal of Philosophy*, Vol. 98, No. 6, June 2001, pp. 273–299.
- Horgan, Terence. 1997. "Kim on Mental Causation and Causal Exclusion", in *Noûs*, Vol. 31, Supplement: Philosophical Perspectives, 11, Mind, Causation, and World 1997, pp. 165–184.
- Jackson, Frank and Philip Pettit. 1988. "Functionalism and Broad Content", in *Mind*, Vol. 97, No. 387, July 1988, pp. 381–400. Reprinted in Jackson, Pettit, and Smith (2004, pp. 95–118). URL: <http://dx.doi.org/10.1093/mind/XCVII.387.381>.
- . 1990a. "Causation in the Philosophy of Mind", in *Philosophy and Phenomenological Research*, Vol. 50, Supplement, Autumn 1990, pp. 195–214. Reprinted with postscript in Clark and Millican (1999, pp. 75–100) and Jackson, Pettit, and Smith (2004, pp. 45–68). URL: <http://dx.doi.org/10.2307/2108039>.

- Jackson, Frank and Philip Pettit. 1990b. "Program Explanation: A General Perspective", in *Analysis*, Vol. 50, No. 2, March 1990, pp. 107–117. Reprinted in Jackson, Pettit, and Smith (2004, pp. 119–130). URL: <http://dx.doi.org/10.2307/3328853>.
- Jackson, Frank, Philip Pettit, and Michael Smith. 2004. *Mind, Morality, and Explanation: Selected Collaborations*, Oxford University Press, Oxford.
- Kallestrup, Jesper and Jakob Hohwy. 2008. *Being Reduced: New Essays on Reduction, Explanation and Causation*, edited by Jesper Kallestrup and Jakob Hohwy. Oxford University Press, Oxford.
- Kim, Jaegwon. 1997. "Does the Problem of Mental Causation Generalize?", in *Proceedings of the Aristotelian Society*, Vol. 97, No. 3, pp. 281–297. URL: <http://dx.doi.org/10.1111/1467-9264.00017>.
- . 1998. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*, MIT Press, Cambridge MA. URL: <https://cognet.mit.edu/library/books/view?isbn=0262611538>.
- . 2003. "Blocking Causal Drainage and Other Maintenance Chores with Mental Causation", in *Philosophy and Phenomenological Research*, Vol. 67, No. 1, July 2003, pp. 151–176. URL: <http://dx.doi.org/10.1111/j.1933-1592.2003.tb00030.x>.
- . 2005. *Physicalism, Or Something Near Enough*, Princeton University Press, Princeton.
- Loar, Brian. 1990. "Phenomenal States", in *Philosophical Perspectives*, Vol. 4, Action Theory and Philosophy of Mind 1990, pp. 81–108. Revised version published as Loar (1997). URL: <http://dx.doi.org/10.2307/2214188>.
- . 1997. "Phenomenal States (Second Version)", in *The Nature of Consciousness: Philosophical Debates*, edited by Ned Block, Owen J. Flanagan, and Güven Güzeldere, MIT Press, Cambridge MA, pp. 597–616. Revised version of Loar (1990). Reprinted in Chalmers (2002, pp. 295–310) and Ludlow, Nagasawa, and Stoljar (2004, pp. 219–240). URL: <http://www.nyu.edu/gsas/dept/philo/courses/consciousness97/papers/loar.html>.
- Ludlow, Peter, Yujin Nagasawa, and Daniel Stoljar. 2004. *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, edited by Peter Ludlow, Yujin Nagasawa, and Daniel Stoljar. MIT Press, Cambridge MA.
- Mellor, D. H. 1995. *The Facts of Causation*, Routledge, London.
- Menzies, Peter. 2003. "The Causal Efficacy of Mental States", in *Physicalism and Mental Causation: The Metaphysics of Mind and Action*, edited by Sven Walter and Heinz-Dieter Heckmann, Imprint Academic, Exeter. Reprinted

- in Monnoyer (2004). URL: <http://www.phil.mq.edu.au/staff/pmenzies/Mental%20Causation.pdf>.
- Monnoyer, Jean-Maurice. 2004. *La Structure du Monde: Objets, Propriétés, États de Choses*, edited by Jean-Maurice Monnoyer. Vrin, Paris.
- Moser, Paul K. and J. D. Trout. 1995. *Contemporary Materialism: A Reader*, edited by Paul K. Moser and J. D. Trout. Routledge, London.
- Noordhof, Paul. 1997. “Making the Change: the Functionalist’s Way”, in *British Journal for the Philosophy of Science*, Vol. 48, No. 2, pp. 233–250. URL: <http://dx.doi.org/10.1093/bjps/48.2.233>.
- Papineau, David. 2002. *Thinking about Consciousness*, Oxford University Press, Oxford. URL: <http://dx.doi.org/10.1093/0199243824.001.0001>.
- . 2007. “Phenomenal and Perceptual Concepts”, in *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, edited by Torin Alter and Sven Walter, Oxford University Press, Oxford, pp. 111–144. URL: <http://dx.doi.org/10.1093/acprof:oso/9780195171655.003.0007>.
- Pereboom, Derk and Hilary Kornblith. 1991. “The Metaphysics of Irreducibility”, in *Philosophical Studies*, Vol. 63, No. 2, August 1991, pp. 125–145. URL: <http://dx.doi.org/10.1007/BF00381684>.
- Price, Huw and Richard Corry. 2007. *Causation, Physics and the Constitution of Reality: Russell’s Republic Revisited*, edited by Huw Price and Richard Corry. Oxford University Press, Oxford.
- Price, Huw and Brad Weslake. forthcoming. “The Time-Asymmetry of Causation”, in *The Oxford Handbook of Causation*, edited by Helen Beebe, Christopher Hitchcock, and Peter Menzies, Oxford University Press, Oxford. URL: <http://philsci-archive.pitt.edu/archive/00004475/>.
- Robb, David and John Heil. 2003. “Mental Causation”, in *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Stanford University, Stanford. URL: <http://plato.stanford.edu/entries/mental-causation/>.
- Ross, Don and David Spurrett. 2005. “What to Say to a Sceptical Metaphysician: A Defense Manual for Cognitive and Behavioral Scientists”, in *Behavioral and Brain Sciences*, Vol. 27, No. 5, October 2005, pp. 603–627. URL: <http://dx.doi.org/10.1017/S0140525X04000147>.
- Russell, Bertrand. 1912–1913. “On the Notion of Cause”, in *Proceedings of the Aristotelian Society*, Vol. 13, pp. 1–26. Reprinted in Russell (1918, pp. 142–164) and Russell (2003, pp. 163–182). Simplified version published as Russell (1914).

- Russell, Bertrand. 1914. "On the Notion of Cause, with Applications to the Free Will Problem", in *Our Knowledge of the External World: As a Field for Scientific Method in Philosophy*, Open Court, Chicago, pp. 214–246. Reprinted in Feigl and Brodbeck (1953, pp. 387–407). URL: <http://www.hist-analytic.org/Russellcause.pdf>.
- . 1918. *Mysticism and Logic*, Longmans Green, London.
- . 2003. *Russell on Metaphysics: Selections from the Writings of Bertrand Russell*, edited by Stephen Mumford. Routledge, London.
- Shapiro, Lawrence A. forthcoming. "Lessons from Causal Exclusion", in *Philosophy and Phenomenological Research*, Vol. URL: <http://philosophy.wisc.edu/shapiro/HomePage/Lessons.pdf>.
- Shapiro, Lawrence A. and Elliott Sober. 2007. "Epiphenomenalism—The Do's and the Don'ts", in *Thinking about Causes: From Greek Philosophy to Modern Physics*, edited by Peter Machamer and Gereon Wolters, Pittsburgh-Konstanz Series in the Philosophy and History of Science, University of Pittsburgh Press, Pittsburgh, pp. 235–264. URL: <http://philosophy.wisc.edu/sober/shapiro%20and%20sober%20epi%201%202%2006.pdf>.
- Stoljar, Daniel. 2008. "Distinctions in Distinction", in *Being Reduced: New Essays on Reduction, Explanation and Causation*, edited by Jesper Kallestrup and Jakob Hohwy, Oxford University Press, Oxford, pp. 263–279.
- Strevens, Michael. 2007. "Review of Woodward, *Making Things Happen*", in *Philosophy and Phenomenological Research*, Vol. 74, No. 1, January 2007, pp. 233–249. URL: <http://dx.doi.org/10.1111/j.1933-1592.2007.00012.x>.
- Thomson-Jones, Martin. 2007. "Overdetermination, Mental Causation, and the Staggered Firing Squad: Problems with the Case for Compatibilism". Unpublished manuscript.
- Van Gulick, Robert. 1992. "Three Bad Arguments for Intentional Property Epiphenomenalism", in *Erkenntnis*, Vol. 36, No. 3, May 1992, pp. 311–332. URL: <http://dx.doi.org/10.1007/BF00204132>.
- Weslake, Brad. 2006. "Review of *Making Things Happen*", in *Australasian Journal of Philosophy*, Vol. 84, No. 1, March 2006, pp. 136–140. URL: <http://dx.doi.org/10.1080/00048400600571935>.
- Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*, Oxford University Press, New York. URL: <http://dx.doi.org/10.1093/0195155270.001.0001>.
- . 2008. "Mental Causation and Neural Mechanisms", in *Being Reduced: New Essays on Reduction, Explanation and Causation*, edited by

Jesper Kallestrup and Jakob Hohwy, Oxford University Press, Oxford, pp. 218–262.

Yablo, Stephen. 1992. “Mental Causation”, in *The Philosophical Review*, Vol. 101, No. 2, April 1992, pp. 245–280. Reprinted in Yablo (2009, pp. 222–248).

———. 1997. “Wide Causation”, in *Noûs*, Vol. 31, No. Supplement: Philosophical Perspectives, 11, Mind, Causation, and World, pp. 251–281. Reprinted in Yablo (2009, pp. 275–306).

———. 2009. *Thoughts: Papers on Mind, Meaning, and Modality*, Oxford University Press, Oxford. URL: <http://dx.doi.org/10.1093/acprof:oso/9780199266463.001.0001>.