The Logical Form of Interventionism

Michael Baumgartner

This paper argues that, notwithstanding the remarkable popularity of Woodward's (2003) interventionist analysis of causation, the exact definitional details of that theory are surprisingly little understood. There exists a discrepancy in the literature between the insufficient appreciation of the logical details of interventionism, on the one hand, and the amount of theoretical work interventionism is expected to do, on the other. The first part of the paper distinguishes four significantly different readings of the logical form of Woodward's analysis and identifies the two readings that best capture Woodward's intentions in (2003) and (2008a), respectively. In the second part, I show that these different readings are not clearly kept apart in the literature, and, moreover, that neither of them can do all the work that interventionists would like the theory to do.

I

Interventionist theories of causation and methodologies of causal reasoning embedded in that theoretical framework have become increasingly popular in recent years. The by far most thorough and elaborate presentation of a modern variant of interventionism has been given in Woodward (2003). Especially philosophers of the special sciences have received Woodward's theory with great enthusiasm.[1] This paper argues that, notwithstanding the remarkable popularity of Woodward's interventionism, the exact definitional details of that theory are surprisingly little understood. Most friends of interventionism apply that theory rather loosely and without a precise understanding of the logical form of its definitional core. To some extent, this lack of clarity about the logical form of interventionism stems from ambiguities in Woodward's own formulations. As will be shown in the first part of

[1] Cf. e.g. Reisman and Forber (2005), Shapiro and Sober (2007), or Campbell (2007).

the paper, the definition of (direct) causation Woodward proposes in (2003) allows for three deviating readings that are all (reasonably) faithful to the grammatical surface of Woodward's wording, yet differ significantly with respect to the truth conditions they attribute to causal statements. Furthermore, in a recent paper (Woodward 2008a), Woodward suggests yet another reading of his definition of causation, one that is incompatible with all three literal readings of his original account. Woodward never explicitly distinguishes between these different variants of his theory. Accordingly, many authors have overlooked the ambiguities contained in the definitional core of interventionism, to the effect that, depending on the context in which the theory is applied, different readings of it are implicitly presupposed. Section II, hence, demarcates the various readings of Woodward's interventionism by explicitly formalizing them. Moreover, it identifies the two readings that best capture Woodward's intentions in (2003) and (2008a), respectively. Finally, in section III, I show that neither of the possible readings of Woodward's analysis does all the theoretical work that friends of interventionism would like the latter to do.

II

The following is Woodward's original and most frequently cited definition of direct causation (Woodward 2003, 59):[2]

(M) A *necessary and sufficient condition* for X to be a (type-level) direct cause of Y with respect to a variable set $\mathbf{V}$ is that *there be* a *possible* intervention on X that will change Y or the probability distribution of Y *when* one holds fixed at some value *all* other variables Z_i in $\mathbf{V}$.

Before we take a closer look at the possible readings of (M), three things need to be noted about an account of causation that turns on (M). First, as Woodward points out repeatedly, it is non-reductive insofar as it does not spell out causation in non-causal terms. Rather, it defines causation in terms of the notion of an *intervention* on X with respect to Y, which Woodward (2003, 98) defines as a surgical (causal) manipulation of X that is not connected to Y on a path that does not go through X and that is (statistically) independent of all causes of Y that are not located on a path through X. Second, (M)

[2] The italics are mine. They are intended to emphasize the logical constants in (M). Note that Woodward is very explicit about the fact that he sees (M) to be a *definition* of direct causation (cf. e.g. Woodward 2003, 55, 60-61). We can confine our discussion to Woodward's notion of direct causation, because it is the core notion of his theory. Indirect (or contributing) causation is then defined based on direct causation.

provides a notion of causation that is relativized to a set of investigated variables. For simplicity, I shall subsequently only make this relativization explicit where necessary to avoid misunderstandings. And third, it is a variant of a counterfactual analysis of causation, because the notion of a *possible* intervention contained in (M), according to Woodward, should be interpreted to mean that if some intervention on X with respect to Y were to occur, (the probability distribution of) Y would change in some reproducible way (cf. Woodward 2003, 70-71, 88-90, Woodward and Hitchcock 2003).

The very beginning of (M) claims to supply a necessary and sufficient condition for any two variables to stand in the relation of direct causation. Accordingly, the main formal feature of (M) is a universally quantified biconditional. The left-hand side of this biconditional is constituted by the analysandum "X is a direct cause of Y with respect to a variable set $\mathbf{V}$" which—if variables of causal structures are chosen as domain of quantification—can be symbolized by $C_{\mathbf{V}}xy$ with $C_{\mathbf{V}}$ representing the relation "…is a direct cause of…relative to variable set $\mathbf{V}$". The right-hand side of the biconditional, in turn, provides the analysans of $C_{\mathbf{V}}xy$. It features three unmistakable logical constants—"there be", "possible", "all"—and one constant—"when"—which could be interpreted in terms of a conditional, or, as we shall see below, might also be taken to stand for mere conjunction.[3] The analysans first states (*de re*) the existence of a possible intervention on X with respect to Y which, relative to fixed domain semantics of modal logic, amounts to (*de dicto*) stating the possible existence of an intervention on X with respect to Y: $\Diamond\exists iIixy$, with I standing for the ternary relation "…is an intervention on…with respect to…".[4] Informally, the remainder of the analysans of $C_{\mathbf{V}}xy$ asserts that, while the possible intervention is performed on X and all other variables in $\mathbf{V}$ are held fixed, the value or the probability distribution of Y changes. Taken in combination, the two constituents of the right-hand side of the biconditional in (M) allow for three significantly different readings—depending on how "when" is interpreted:

(i) There possibly exists an i such that *if* i is an intervention on X with respect to Y and all

[3] Plainly, there also is an "or" in (M). This disjunction is intended to ensure that the theory is applicable to both deterministic and probabilistic structures. As this is of no relevance to the present context, I am not going to explicitly formalize that disjunction.

[4] In modal systems containing the Barcan Formula (BF) $\Diamond\exists xFx$ is even equivalent to $\exists x\Diamond Fx$. For details see e.g. Hughes and Cresswell (1996, 246). Irrespective of whether BF is presupposed, I take a *de dicto* reading of (M) that sidesteps metaphysical questions as to the manner of existence of possibilia to be more in line with the basic non-metaphysical approach followed in Woodward (2003). The subsequent discussion, however, in no way hinges on this preference of a *de dicto* reading of (M). That is, whoever prefers a *de re* reading may simply substitute $\exists i\Diamond Iixy$ for $\Diamond\exists iIixy$ in what follows.

other variables are held fixed, *then* i is accompanied by changes in Y (or of its probability distribution).

(ii) There possibly exists an i such that i is an intervention on X with respect to Y *and* if all other variables are held fixed, then i is accompanied by changes in Y (or of its probability distribution).

(iii) There possibly exists an i such that i is an intervention on X with respect to Y and all other variables are held fixed *and* i is accompanied by changes in Y (or of its probability distribution).

While in readings (i) and (ii) "when" is interpreted as a conditional (with varying scope), in (iii) it is taken to stand for conjunction. In (i) "if. . . then" is the main operator within the scope of the existential quantifier, in readings (ii) and (iii) the latter's scope is governed by "and". While (ii) features two conjuncts—the second conjunct being a conditional—, (iii) contains three conjuncts. How these readings affect the truth conditions of (M) and, thus, the analysis of causation provided by (M) is most clearly seen if the three complete logical forms of (M) induced by (i), (ii), and (iii) are contrasted. By introducing the predicates F representing ". . . is held fixed" and G standing for ". . . is accompanied by changes in . . . " and by restricting the domain of universal quantifiers to the variables in the set $\mathbf{V}$,[5] reading (i) yields (1), reading (ii) yields (2), and reading (iii) yields (3) as logical form of (M):

$$\forall x, y(C_{\mathbf{V}}xy \leftrightarrow \Diamond\exists i(Iixy \wedge \forall z(z \neq i \wedge z \neq x \wedge z \neq y \rightarrow Fz) \rightarrow Giy)) \tag{1}$$

$$\forall x, y(C_{\mathbf{V}}xy \leftrightarrow \Diamond\exists i(Iixy \wedge (\forall z(z \neq i \wedge z \neq x \wedge z \neq y \rightarrow Fz) \rightarrow Giy))) \tag{2}$$

$$\forall x, y(C_{\mathbf{V}}xy \leftrightarrow \Diamond\exists i(Iixy \wedge \forall z(z \neq i \wedge z \neq x \wedge z \neq y \rightarrow Fz) \wedge Giy)) \tag{3}$$

Not all of these readings of (M) capture the basic intuitions behind an interventionist analysis of causation equally well. Reading (1), for instance, certainly misses Woodward's intentions. The right-hand side of the biconditional in (1) turns out true if it is impossible to satisfy $Iixy$ or $z \neq i \wedge z \neq x \wedge z \neq y \rightarrow Fz$, for, in that case, the antecedent of the conditional within the scope of the existential quantifier is false which renders the conditional as a whole true. That is, if (M) is read in terms of

[5] This restriction could, of course, easily be formally expressed. To keep the formalizations as simple as possible I abstain from doing so here.

(1), it determines X to directly cause Y if either it is impossible that there exists an intervention on X with respect to Y or the other variables in the structure cannot be held fixed. Obviously, such a reading would have highly unwelcome consequences. For example, if the same variable is substituted for X and Y, i.e. $X = Y$, it is impossible to intervene (in Woodward's sense of the term) on X with respect to Y, because there cannot possibly exist a surgical manipulation of X that is not at the same time a manipulation of (the identical) Y.[6] More generally, it is impossible to intervene on any variable with respect to itself. Subject to (1), hence, every variable would turn out to trivially cause itself.[7] Or if there is only one variable among the variables other than I, X, and Y which cannot be held fixed, X would automatically be identified as direct cause of Y by (1). Both of these implications of (1) run counter to common causal intuitions. Woodward clearly does not have reading (1) in mind.

By contrast, if reading (2) is assumed, the possible existence of an intervention on X with respect to Y is rendered necessary for a direct causal dependency between these two variables. If there is no such possible intervention, the right-hand side of the biconditional in (2) is false, which amounts to the causal irrelevance of X to Y. While, according to (2), X only directly causes Y if there possibly exists an intervention on X with respect to Y, the fixability of the other variables in $\mathbf{V}$ is not necessary for a direct causal dependency between X and Y. For if not all remaining variables in $\mathbf{V}$ can be held fixed, the antecedent of the conditional in the second conjunct within the scope of the existential quantifier in (2) is false, which makes the whole conditional true. Hence, if there possibly exists an intervention on X with respect to Y and at least one of the remaining variables in the structure cannot be held fixed, both conjuncts of (2) are satisfied which, subject to (2), implies that X directly causes Y. Finally, according to reading (3), both the possible existence of an intervention on X with respect to Y and the fixability of all other variables turn out to be necessary conditions of a direct causal dependency between X and Y. (3) requires that in order for Y to be directly causally dependent on X there possibly exists an intervention on X with respect Y such that all other variables in $\mathbf{V}$ are fixed and Y changes its value or its probability distribution.

(3) is the strongest of all possible readings of (M)—it implies both (1) and (2). In (2003), Woodward provides a number of indications that (3) in fact is his intended reading. For instance, he sub-

[6] More specifically, condition (IV.3) of Woodward's (2003, 98) notion of an intervention cannot be met.

[7] This problem could be averted by adding a constraint to (M) that requires X and Y to be *different* variables. Yet, under reading (1), a version of (M) that is supplemented in that vein would still entail that non-manipulability is sufficient for causation, which is a very counterintuitive consequence.

scribes to the slogan "*No causal difference without a difference in manipulability relations, and no difference in manipulability relations without a causal difference*" (Woodward 2003, 61), or he explicitly characterizes the manipulability of X with respect to Y as a necessary condition of X causing Y (Woodward 2003, 112-113, 128). Even though he never positively identifies the fixability of the other variables in $\mathbf{V}$ as necessary for X to directly cause Y, he clearly wants to establish a tight conceptual connection between manipulability, difference-making in context, and causality. These basic intuitions behind interventionism are best captured by reading (3), which I hence take to be the reading of (M) Woodward has in mind in (2003).

In (2008a), however, he suggests a radically different reading of his definition of direct causation—without explicitly indicating to be modifying his original theory. In a context that discusses the problem of mental-to-physical causation, he says the following about his analysis of causation (Woodward 2008a, 224-225):

> I also assume that if a candidate causal claim is associated with interventions that are impossible for (or lack any clear sense because of) logical, conceptual or perhaps metaphysical reasons, then that causal claim is itself illegitimate or ill-defined. In other words, I take it to be an implication of (M) that a legitimate causal claim should have an intelligible interpretation in terms of counterfactuals [or, interchangeably, claims about the possible existence of interventions] the antecedents of which are coherent or make sense. (...) Thus if we have two apparently competing claims, the first contending some mental state is causally inert and the other contending that it causes some outcome, it must be possible to specify some set of (coherent, well-defined) interventions such that the two claims make competing predictions about what would happen under those interventions. If we cannot associate such an interventionist interpretation with one or both of the claims, the claim(s) in question lack a clear sense (...).

Plainly, if interventions on X with respect to Y are impossible, none of the readings of (M) discussed above yields that "X causes Y" is ill-defined or meaningless—contrary to what Woodward claims in this passage. Rather, (1), (2), and (3) assign definite truth-values to causal claims based on, among other things, whether interventions on X are possible or not. If—for whatever reason—it is impossible to intervene on X with respect to Y, (1) renders "X causes Y" true, whereas (2) and (3) render "X causes Y" false.

Hence, the quoted passage clearly shows that in (2008a) Woodward has an account of causation in mind that significantly differs from the literal readings of the theory discussed above. I am even inclined to say that he significantly modifies his theory in Woodward (2008a). As he does not make his modifications explicit, we are left to reconstruct his most recent understanding of interventionism on our own. Apparently, Woodward no longer takes the manipulability of X to be a necessary condition for X to cause Y. If X is not manipulable with respect to Y, the claim "X causes Y" shall newly be ill-defined and no longer false. The same holds for the fixability of the other variables in the set of investigated variables $\mathbf{V}$. Woodward (2008a, 256) indicates that if it is impossible to intervene on a variable X with respect to Y while holding the other variables in the structure fixed, it does not follow that X is causally irrelevant to Y. Rather, the impossibility to fix the other variables prohibits a coherent or meaningful interpretation of the claim "X causes Y". That is, according to Woodward (2008a), the possible existence of an intervention on X with respect to Y and the fixability of the other variables in the structure are *preconditions* of the meaningfulness of claims about causal dependencies among X and Y or about the absence of such dependencies. Rather than being necessary for causation, manipulability and fixability now turn out to be criteria for the well-definedness of causal claims. These considerations suggest that Woodward (2008a) reads (M) somewhat along the following lines:

(M') For all X, Y: *If* there possibly exists an intervention i on X with respect to Y such that all other variables Z in the pertaining variable set $\mathbf{V}$ are held fixed at some value, *then* X is a (type-level) direct cause of Y with respect to $\mathbf{V}$ *iff* i is accompanied by changes in Y (or of its probability distribution).

Formally, this amounts to (4), which is to be understood on the basis of the same interpretation as given for (1), (2), (3) above:

$$\forall x, y \Diamond \exists i (Iixy \wedge \forall z (z \neq i \wedge z \neq x \wedge z \neq y \rightarrow Fz) \rightarrow (C_{\mathbf{V}}xy \leftrightarrow Giy)) \tag{4}$$

(4) is weaker than (3), i.e. (3) implies (4) but not vice versa. Contrary to (3), (4) indeed does not assign a truth-value to "X causes Y" when it is impossible to intervene on X with respect to Y or to fix the other variables in the structure. Against the background of (4), the possibility to intervene on X while the other variables are held fixed can—in the vein of the passage quoted above—be argued

to be a precondition of "X causes Y" being a well-defined causal claim. This consequence of (4) can be given a *strong* and a *weak* reading. Subject to the strong reading, whenever manipulability or fixability are violated there is no objective fact of the matter whether "X causes Y" is true or not. In contrast, according to the weak reading, the existence of a causal dependency between X and Y can simply not be assessed within the interventionist framework in cases of violations of manipulability or fixability. That, however, does not exclude that X might be identified as cause of Y within some other theoretical framework in such cases. Or differently, if manipulability and fixability are violated the strong reading of (4) implies that "X causes Y" is meaningless, while the weak reading only implies that the interventionist framework is inapplicable to determining the truth-value of that causal claim. The passage quoted above indicates that Woodward favors the strong reading.

Overall, thus, our discussion of the logical form of Woodward's interventionist analysis of causation yields a double result. In (2003), he endorses an analysis as specified in (3), according to which manipulability and fixability are necessary conditions for direct causation, whereas in (2008a), he turns to an account along the lines of the strong reading of (4), according to which manipulability and fixability are preconditions of the well-definedness of causal claims. Independently of which version of interventionism one favors, though, the four versions discussed in this section must be expressly distinguished, for they have radically different implications.

III

As indicated in the introductory section, interventionism has been claimed to do all kinds of theoretical work in recent years—especially by philosophers of the special sciences. The remainder of this paper is going to show that neither of the four versions of interventionism can do all the work that interventionists would like the theory to do.

To show this, it suffices to consider what the four different versions of interventionism say about three simple causal statements—one on self-causation, one on macro-to-micro causation, and one on macro-to-macro causation. First, take statement (a):

(a) Every variable X is a (type level) cause of itself.

It is fair to say that the non-existence of ubiquitous self-causation is virtually a truism of the philosophy

$$\begin{tikzcd}
M \arrow[r] \arrow[dr] & M^* \\
P \arrow[u, Rightarrow] \arrow[r] & P^* \arrow[u, Rightarrow]
\end{tikzcd}$$

Figure 1: A causal structure as claimed to exist by non-reductive physicalists. M and M^* represent two types of macro phenomena that non-reductively supervene (symbolized by "$\Longrightarrow$") on the micro phenomena represented by P and P^*. Arrows of type "$\longrightarrow$" stand for the direct causal dependencies that are assumed to hold among these variables.

of causation. If at all, self-causation only occurs in exceptional feedback structures. Hence, the standard opinion among philosophers of causation is that (a) is a *false* statement. To theoretically reproduce the falsity of (a), interventionism must be understood in terms of either (2) or (3). For, as we have seen in the previous section, interventionism of type (1) yields that (a) is a true statement, because it is impossible to intervene on any variable with respect to itself. By contrast, subject to interventionism of type (4), the impossibility to intervene on any variable with respect to itself entails that (a) is neither true nor false but ill-defined or meaningless. Only versions (2) and (3), according to which manipulability is necessary for causation, determine (a) to be false (and meaningful).

Second, consider the graph in figure 1, which depicts a causal structure over the variable set $\mathbf{V} = \{M, M^*, P, P^*\}$, where M and M^* represent two types of macro phenomena and P and P^* represent the two types of micro phenomena that realize particular values of M and M^*, respectively. Suppose, we analyze this structure from the perspective of a non-reductive physicalist who takes macro properties to supervene on micro properties in a non-reductive way and who subscribes to the causal closure of the physical as well as to the existence of both macro-to-micro causation and macro-to-macro causation. That is, M and M^* are taken to supervene on P and P^* such that $M \neq P$ and $M^* \neq P^*$; moreover, M shall be assumed to be a direct cause of M^* and of P^* (relative to $\mathbf{V}$), while P is a direct cause of P^*. Against this background, we now inquire what interventionism says about the truth-values and meaningfulness of the following two statements:

(b) M is a (type level) direct cause of P^* relative to $\mathbf{V}$.

(c) M is a (type level) direct cause of M^* relative to $\mathbf{V}$.

Let us begin with (b). Recently, some non-reductive physicalists with sympathies for interventionism (e.g. Shapiro and Sober 2007) have argued that Woodward's interventionism successfully accounts for macro phenomena causing effects of their supervenience bases and, hence, renders statements like (b) true. To determine which reading of (M) indeed yields that result we, among other things, have to assess whether it is possible to intervene on M with respect to P^*. According to Woodward (2003, 98), interventions on M with respect to P^* must be (statistically) independent of all causes of P^* that are not located on a path from M to P^*. In virtue of the structure in figure 1, P is located on a causal path to P^* that does not include M. Hence, any possible intervention on M with respect to P^* must be independent of changes in P. Yet, M supervenes on P, which implies that every change induced on M is *necessarily* correlated with changes in P. It is therefore *impossible* to intervene on M with respect to P^*. Moreover, this impossibility does not hinge on the choice of the investigated variable set $\mathbf{V}$. Contrary to Woodward's notion of direct causation, his notion of an intervention is not relativized to a particular variable set.[8] An intervention on M with respect to P^* would have to be independent of *all* other causes of P^*, irrespective of whether they are contained in $\mathbf{V}$ or not. Such a variable cannot possibly exist.[9]

This finding answers the question as to how the different variants of interventionism evaluate statement (b). According to (2) and (3), a violation of manipulability entails that (b) is false. By contrast, (4) determines (b) to be meaningless. The only variant of interventionism that renders (b) true is (1). Hence, non-reductive physicalists as Shapiro and Sober (2007) who claim that Woodward's interventionism immunizes non-reductive physicalism against the threat posed by the problem of causal exclusion, presumably, read Woodward's theory along the lines of (1).[10] At the same time, however, these authors, most likely, would not want to subscribe to the truth of (a), which is entailed by (1). That is, when it comes to cases of self-causation they draw on a different understanding of interventionism.

Finally, let us turn to statement (c). While interventions on M with respect to P^* would, impossibly, have to be independent of P, that is not required for interventions on M with respect to M^*.

[8] Strevens (2007, 243) contends that Woodward's (2003, 98) definition (IV) of an intervention variable is implicitly relativized to a variable set. To this, Woodward (2008b) replies, correctly in my view, by insisting that (IV) defines the notion of an intervention variable not by drawing on the relativized notion of causation provided by (M), but by drawing on a de-relativized notion of causation *simpliciter* which is defined via existential generalization of (M). In response, Strevens (2008) argues that this de-relativization of (M) gets the interventionist account involved in a vicious circle. I cannot enter this discussion here. For the purposes of this paper, I simply interpret Woodward's definition (IV) in the way that is most faithful to its wording, that is, in the non-relativized way.

[9] For a detailed presentation of this line of reasoning cf. Baumgartner (2009), Baumgartner (2010).

[10] Also Menzies (2008, 206) explicitly subscribes to a conditional reading of interventionist causation in the vein of (1).

Even though P is a cause of P^* which constitutes the supervenience base of M^*, P is no cause of M^* in the structure of figure 1, because the supervenience relation is non-causal. In consequence, P may vary while M is manipulated with respect to M^*. That is, it is very well possible that there exist interventions on M with respect to M^*—and let us moreover assume that these interventions are indeed accompanied by changes in M^*. By contrast, it is still not possible to hold P fixed while M is manipulated. Every intervention on M is necessarily correlated with changes in its supervenience base P. Thus, there is a variable in $\mathbf{V}$ that differs from M and M^* and that cannot be held fixed while M is intervened upon, i.e. it is impossible to satisfy the fixability requirement.

Based on these findings the different variants of interventionism assign the following truth-values to (c). According to (1), violations of either manipulability or fixability are sufficient for the truth of corresponding causal claims. That is, subject to (1), (c) is true. In virtue of (2), fixability of the other variables in $\mathbf{V}$ is not necessary for direct causation. Hence, given that interventions on M with respect to M^* are accompanied by changes in M^*, (2) renders (c) true. (3), in turn, requires both manipulability and fixability for direct causation and, correspondingly, determines (c) to be false. Finally, according to (4), the non-fixability of all other variables in $\mathbf{V}$ violates one precondition of the well-definedness of (c), which is, therefore, meaningless.

In sum, there is not one single variant of interventionism that yields the—intuitively desirable—falsity of (a) and that adequately reproduces the causal structure in figure 1, i.e. that mirrors the truth of (b) and (c). Of the two of Woodward's intended readings, (3) implies the falsity of (a), (b), and (c), whereas (4) implies the ill-definedness of all of these causal statements. Both the falsity of (a) and the truth of (b) and (c) can only be theoretically accounted for on the basis of Woodward's interventionism if that theory is understood differently depending on the context of application.

IV

To conclude, there exists a discrepancy in the literature between the lack of clarity about the logical details of interventionism, on the one hand, and the work interventionism is expected to do, on the other. Ambiguities and a certain degree of implicitness in Woodward's own formulations are partly responsible for the insufficient appreciation of the logical details of interventionism in the literature. I suspect, however, that interventionism also tends to be applied rather loosely because the theory has so much (pre-theoretic) intuitive appeal that it is easily endorsed without a thorough inspection of its logical form. This paper has shown that intuitive appeal is not enough. Before interventionism can be put to work, its logical details have to be carefully worked out and clearly understood. I have distinguished four conceivable readings of the definitional core of Woodward's interventionism, none of which serves all desirable purposes. That is, our clarifications of the logical form of interventionism suggest that certain friends of interventionism might have subscribed to that framework just somewhat prematurely.

References

Baumgartner, M. (2009). Interventionist causal exclusion and non-reductive physicalism. *International Studies in the Philosophy of Science 23*, 161–178.

Baumgartner, M. (2010). Interventionism and epiphenomenalism. *Canadian Journal of Philosophy 40*, 359–384.

Campbell, J. (2007). An interventionist approach to causation in psychology. In A. Gopnik and L. Schulz (Eds.), *Causal Learning. Psychology, Philosophy, and Computation*, pp. 58–66. Oxford: Oxford University Press.

Hughes, G. E. and M. J. Cresswell (1996). *A New Introduction to Modal Logic*. London: Routledge.

Menzies, P. (2008). The exclusion problem, the determination relation, and contrastive causation. In J. Hohwy and J. Kallestrup (Eds.), *Being Reduced. New Essays on Reduction, Explanation, and Causation*, pp. 196–217. Oxford: Oxford University Press.

Reisman, K. and P. Forber (2005). Manipulation and the causes of evolution. *Philosophy of Science 72*, 1113–1123.

Shapiro, L. and E. Sober (2007). Epiphenomenalism. The dos and don'ts. In G. Wolters and P. Machamer (Eds.), *Thinking about Causes: From Greek Philosophy to Modern Physics*, pp. 235–264. Pittsburgh: University of Pittsburgh Press.

Strevens, M. (2007). Review of Woodward, Making Things Happen. *Philosophy and Phenomenological Research LXXIV*, 233–249.

Strevens, M. (2008). Comments on Woodward, Making Things Happen. *Philosophy and Phenomenological Research LXXVII*, 171–192.

Woodward, J. (2003). *Making Things Happen*. Oxford: Oxford University Press.

Woodward, J. (2008a). Mental causation and neural mechanisms. In J. Hohwy and J. Kallestrup (Eds.), *Being Reduced: New Essays on Reductive Explanation and Special Science Causation*, pp. 218–262. Oxford: Oxford University Press.

Woodward, J. (2008b). Response to Strevens. *Philosophy and Phenomenological Research LXXVII*, 193–212.

Woodward, J. and C. Hitchcock (2003). Explanatory generalizations, part I: A counterfactual account. *Noûs 37*, 1–24.