Getting Rid of Interventions

ALEXANDER REUTLINGER

University of Cologne

1. Introduction

In his highly influential book *Making Things Happen*, James Woodward argues for an interventionist theory of causation. In the recent debate, the interventionist theory has been enthusiastically received as an adequate theory of causation in the special sciences. Especially, the interventionist theory has been taken to apply to the social and the biomedical sciences (for instance and apart from Woodward's own work, by Craver 2007 in philosophy of neuroscience, Waters 2007 in philosophy of biology, Campbell 2007 and Shapiro & Sober 2007 in philosophy of psychology, and Hausman 2009 in philosophy of economics).

The key idea underlying interventionist account – which also defines other causal notions besides direct causation – is that, roughly, X is a cause of Y iff there is a possible intervention on X that changes Y (cf. Woodward 2003: 59 for Woodward's central definition **M**). For instance, Woodward defines the notion of a direct (type-level) cause as follows:

> 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) when all other variables are held fixed at some value by interventions. (Woodward 2003: 55)

The central character of interventions in Woodward's framework is obvious: Woodward's definitions of various causal notions imply that the truth of a causal statement "X directly causes Y" requires the existence of a possible intervention on X. Another way to make the same point is that if "X directly causes Y" is true, then the following interventionist counterfactuals have to be true: "if there were an intervention $I = i$ on X such that $X = x$, then $Y = y$ would be the case", and "if

there were an intervention $I = i^*$ on X such that $X = x^*$, then $Y = y^*$ would be the case" (with i≠i*, x≠x*, y≠y*). Woodward requires that interventions be merely *in principle possible*. Woodward interprets in principle possibility as *logical possibility*. Call this the *modal character* of interventions.

Let me add two clarifications at this point. *First*, what kind of a project is Woodward's interventionist theory of causation? It is crucial to emphasize that Woodward is *not* concerned with the methodology of causation, i.e. the construction of algorithms that allow to infer causal models from statistical data (cf. Woodward 2003: 38).[1] The aim of interventionists is to provide a semantic account of causal statements. Woodward is very explicit about the semantic goal of his approach: "my aim is to give an account of the *content* or *meaning* of various locutions, such as X causes Y […]" (Woodward 2003: 38, my emphasis, cf. also 2003: 7-9, 2008: 194-196).[2]

Second, how does the claim that *there is a possible intervention* relate to interventionist counterfactuals of the form "*if there were an intervention $I = i$ on X such that $X = x$, then $Y = y$ would be the case*"? According to the standard possible worlds semantics, sentences such as "it is possible to intervene on X" or, alternatively, "there is a possible intervention $I = i$ on X" are true iff $I = i$ is an intervention on X in at least one possible world. When we describe – following interventionists – what would happen under an intervention $I = i$ by using an interventionist counterfactual, we evaluate the conditional in a possible – actual or counterfactual – world in which (among other constraints) $I = i$ is an intervention

---

[1] "By contrast [to Pearl's and Spirtes, Glymour, and Scheines's work], I have nothing to say about issues having to with calculating quantitative magnitudes, estimation, identifiability, or causal inference. Instead, my enterprise is, roughly, to provide an account of the meaning or content of just those qualitative causal notions that Pearl (and perhaps Spirtes et al.) take as primitive. *Because my project is semantic or interpretative, and is not intended as a contribution to practical problems of causal inference* […]." (Woodward 2003: 38, my emphasis)

[2] To be fair, I have to concede that Woodward (2008) initially describes his project as methodological in a reply to Michael Strevens' worry that interventionism suggests that causal facts are fundamental. However, although Woodward initially characterizes his project as being methodological in the opening pages of his reply to Strevens, this label does not match the questions he takes his project to answer: "given that a directed graph or a system of equations can be used, qua representational device, to represent both patterns of correlations and systems of causal relationships, what conditions have to be met for these devices to accurately represent the latter rather than the former? […] If we wanted to try to capture the notion of one events being an actual cause of another within a structural equations or directed graph framework, how would we do so?" (cf. Woodward 2008: 194f). I think these questions are solely concerned with the semantic interpretation of causal statements (relative to a causal model) and the explication of causal notions.

on *X*. One might worry that referring to possible world semantics in order to understand Woodward's account of causation is wrong-headed, because Woodward explicitly rejects Lewis's possible world semantics for counterfactuals (cf. Woodward 2003: 133-145). However, taking a closer look reveals that Woodward merely objects to Lewis's similarity measure for the closeness of world (cf. Woodward 2003: 139, 142 for two counterexamples against Lewis's similarity metric). Woodward's position is entirely coherent with the general idea of Lewisian semantics for counterfactuals: that is, a counterfactual is true at a world w iff the consequent is true in the closest antecedent-worlds.[3] This general idea of Lewis's semantics can be distinguished from Lewis's specific proposal for selecting the closest worlds (i.e. his similarity metric). One appealing way to understand Woodward is to say that he deviates from Lewis by using a different measure for closeness of worlds. From Woodward's point of view, the most obvious candidate for such a measure is this one: the closest antecedent-worlds are those in which the antecedent is the outcome of an intervention (cf. Woodward 2003: 135f., Woodward & Hitchcock 2003: 13f. for the claim that interventions play a similar role as Lewisian small miracles by selecting the closest possible antecedent-worlds). I will return to interventionist semantics in Section 4. The main point I would like to stress here is that using standard possible worlds semantics in order to understand (a) existential claims about possible interventions and (b) interventionist counterfactuals is compatible with Woodward's theory of causation (including his objections to Lewis's similarity metric).

However, despite the *prima facie* virtues[4] of interventionist theories of causation, I will argue that the key notion of the interventionist approach – the notion of a possible intervention – turns out to be deeply problematic. In particular, I will argue that Woodward's notion of an intervention is problematic

---

[3] However, Woodward's worlds should be understood as model worlds or small worlds, i.e. assignments of values to variables in a causal model (cf. Pearl 2000: 207). In this respect, Woodwardian worlds differ from Lewisian worlds because the latter are as typically as detailed as the real spatio-temporal entity we inhabit (cf. Hüttemann 2004: 113).

[4] Interventionist theories are considered to be successful because (a) they explicate several kinds of causation that are referred to in the sciences (e.g. actual causes, type-level causes, indeterministic causes etc.), (b) they account for several typical features of causation (e.g. the time-asymmetry of causation), and (c) they provide a successful description of intuitively possible causal scenarios (such as preemption scenarios),

because of the modal character of possible interventions. I will argue for two claims against Woodward:

1. *Either* merely logically possible interventions are *dispensable* for the semantic project of providing an account of the meaning of causal statements. If interventions are indeed dispensable, the interventionist theory collapses into (some sort of) a counterfactual theory of causation.[5] Thus, the interventionist theory is not tenable as a theory of causation in its own right.

2. *Or*, if one maintains that merely logically possible interventions are indispensable, then interventions with this modal character lead to the fatal result that interventionist counterfactuals are evaluated *inadequately*. Consequently, interventionists offer an inadequate theory of causation.

What I argue in this paper is compatible with the view that interventions are important (and maybe even indispensable) for the methodological task of *discovering* causal relations or inferring causal models from statistical data (i.e. the methodological project which Woodward explicitly distinguishes from his own). My arguments are solely directed against the interventionist theory as a semantic project. In SECTION 2, I will briefly describe Woodward's notion of an intervention. In SECTION 3, I will discuss which kind of possibility is presupposed in Woodward's notion of possible interventions. I will present a case that is supposed to show that it would be unwise to require that interventions must be *physically* possible. Woodward himself discusses this kind of counterexample against the requirement that interventions need to be physically possible. Reacting to the counterexample, Woodward rejects the claim that interventions have to be physically possible and requires interventions to be merely logically possible. In the central SECTION 4, I present two arguments against the adequacy of an account of causation framed in terms of logically possible interventions. In the first argument, I will show that merely logically possible interventions are superfluous

---

[5] Alternatively, Woodward's interventionist theory collapses into invariance theories of causation – i.e. to the extent to which invariance theories can be formulated without the Woodwardian notion of an intervention.

and can be dispensed with for pursuing the semantic project of providing truth-conditions for causal claims (first claim above). In the second argument, I will take a different stance and leave aside the indispensability arguments. I will show that relying on merely logically possible interventions conflicts with the standard approaches to the meaning of counterfactuals. If an existential claim about a merely logically possible intervention figures in the antecedent of an interventionist counterfactual, then this leads to the fatal result that this interventionist counterfactual is evaluated as false, although we would take these counterfactuals to be true. I will argue that also the obvious strategy to choose from an interventionist point of view – i.e. to adopt an interventionist semantics (cf. Reutlinger 2011) – does not solve the problem. In SECTION 5, I conclude that when analyzing causal concepts and stating the truth conditions of causal claims we best get rid of Woodwardian interventions.

2. Woodward's Concept of an Intervention

Woodward defines interventions in two steps: *first* he defines an intervention variable, *then* he uses the notion of an intervention variable in order to define the notion of an intervention. A variable I is an intervention variable for X relative to Y iff the conjunction of the following conditions is satisfied:

(a) I is a cause of X.

(b) There is at least some value of I such that if I takes this value, then X depends only on I and X depends on no other variables, i.e. I is the only cause of X.

(c) I is not a direct cause of Y, and if I is a cause of Y then I is an indirect cause of Y via a causal path leading through X and a – possibly empty – set of intermediate variable $Z_1, \ldots, Z_n$.

(d) I is probabilistically independent of other causes $W_1, \ldots, W_n$ of Y, which are not on a causal path leading from X to Y.

(e) I does not alter the relationship between Y and its causes $W_1, \ldots, W_n$, which are not on a causal path leading from X to Y.

The notion of an intervention variable is, in turn, used to define the notion of an intervention: any value i_i of an intervention variable I (for X relative to Y) is an *intervention on X* iff it is the case that the value of X counterfactually depends on the fact that I has the value i_i (cf. Woodward 2003: 98). Among other features of interventions, Woodward stresses the modal character of interventions. Interventions are not required to be possible in the sense that they are feasible actions for human agents. Woodward's key idea consists in the assumption that it is in principle possible, or logically possible, to intervene. As I stated in SECTION 1, the main goal of this paper is to investigate the modal character of interventions and the problems arising from it.

3. A Counterexample: Why Some Interventions Are Not Even Physically Possible

Recall that according to Woodward, if X is a direct cause of Y, then there is a *possible* intervention on X that changes Y. One might wonder which kind of possibility is in play in the interventionist theory of causation. Could the required kind of possibility simply refer to the abilities of human agents to intervene? Woodward and other interventionists explicitly (and rightly, I think) deny that interventions have to be practically possible, i.e. interventions are not required to be feasible actions for human beings. Woodward (2003: 103f., 123-127) argues against this anthropocentric concept of an intervention as defended by advocates of agency theories of causation (such as Menzies & Price 1993). Woodward objects that agency theories are committed to two problematic metaphysical claims: (a) the capacity of agents to intervene is "a fundamental and irreducible feature of the world and not just a variety of causal interaction among others" (p. 123), and (b) an agency theory "leads us toward an undesirable kind of anthropomorphism or subjectivism regarding causation" (p. 123). As Woodward believes that causal claims have mind-independent truth-makers, he rejects practical possibility (or abilities of an agent to intervene) as the right kind of modality required for his concept of an intervention.

So, is it adequate to think that interventions are *physically* possible? Woodward addresses this question by presenting and discussing a counterexample

against the assumption that physically possible interventions are necessary for determining the truth conditions of causal claims:

> Suppose that […] [*X*]s *occur only spontaneously in the sense that they themselves have no causes*. There are no further factors [I_1, …, I_n] that affect whether or not [*X*] happens, and this is a matter of physical law. (I take this at least to be a logically coherent possibility.) Thus, it is *physically impossible* to carry out an intervention that changes whether [*X*] occurs. Nonetheless, it seems quite possible that the [*X*]s themselves might well have further effects [*Y*]. (Woodward 2003: 130, my emphasis, my alteration of notation)

The general form of Woodward's counterexample is:

> *X* is a cause of *Y* but there is no physically possible intervention $I = i$ on *X*.

I will distinguish two senses of "spontaneously" in the quote by Woodward.

(i)   There are cases in which there is *no* physically possible *deterministic* intervention *but* there *is* a physically possible *indeterministic* intervention on *X*, and

(ii)   There are cases in which there is *neither* a physically possible *deterministic* intervention *nor* a physically possible *indeterministic* intervention on *X*.

Two concrete examples from physics illustrate these cases:

(i) *The No-Deterministic-But-Indeterministic Intervention Case*. Uranium decays "only spontaneously" in the sense that there is no deterministic, physically possible way to manipulate Uranium such that it surely decays at a time *t*. Nonetheless, the decay of Uranium causes a flash on a screen (as a measurement in an experimental setup). Yet, one *can* raise the probability that the Uranium atom decays (e.g. by increasing the energy of the nucleus of the Uranium atom). In other words, there is a physically possible indeterministic intervention on Uranium decay.

(ii) *The Neither-Deterministic-Nor-Indeterministic Intervention Case.* Tim Maudlin (2002: 149f) presents the Big Bang as a case where there is neither a deterministic physically possible way to intervene, nor an indeterministic way to intervene. In other words, according to Maudlin's example, there is no physically possible indeterministic cause of the Big Bang which raises or lowers the probability of the occurrence of the Big Bang. Nonetheless, one would like to maintain that the Big Bang – if anything – has a plentitude of direct and indirect effects.

I will describe the Big Bang case in detail in SECTION 4. However, my arguments do not depend on the Big Bang case. There are other "less cosmological" examples of the *Neither-Deterministic-Nor-Indeterministic Intervention Case.* Although I will mainly focus on the Big Bang case in the rest of the paper, let me briefly present two other examples.

Physical Constants. Many people intuitively believe that physical constants have a causal influence on the behavior of physical objects. For instance, the gravitational constant, the Planck constant, and the constant representing the speed of light can be taken to be causes. Still, it is physically impossible to intervene on a physical constant because changing the constants amounts to changing the physical laws in a non-local way. Changing the laws creates physically impossible worlds.[6] Hence, we have another example of the *Neither-Deterministic-Nor-Indeterministic Intervention Case*: *if* we assume that physical constants count as causes,[7] then there is no physically possible intervention on a physical constant such that the value of the constant changes.[8]

Norton's Dome. John Norton (2007: 22-28) argues for a case of uncaused events that is compatible with Newtonian mechanics. Norton imagines a symmetrically shaped dome that is located in downward directed gravitational

---

[6] Woodward (2003: 208f) discusses the constant speed of light, and observes that the generalization "All physical processes propagate at a speed less than or equal to that of light" is not invariant under physically possible interventions.
[7] Of course, one can disagree with the assumption in the antecedent that physical constants are indeed causes. However, let us assume that they are causes at least for the sake of the argument.
[8] Thanks to Kristina Engelhard for suggesting this example.

field. A mass is at rest on top of the dome, on the apex. Let us represent the dome-scenario by the following figure.
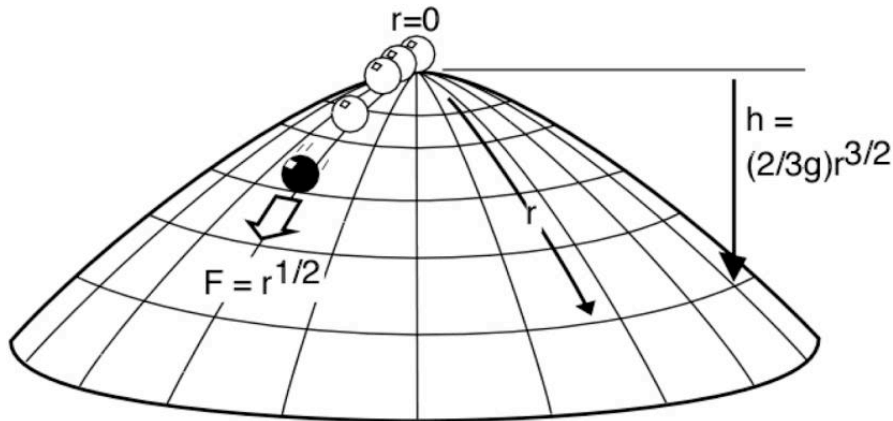
Source: Norton 2007: 23.

Norton claims that the following scenario is consistent with Newtonian mechanics:

> It is a mass at rest in a physical environment that is completely unchanging for an arbitrary amount of time – a day, a month, an eon. Then without any external intervention or any change in the physical environment, the mass spontaneously moves off in an arbitrary direction with the theory supplying no probabilities for the time or direction of the motion. (Norton 2007: 22f.)

Norton interprets the acceleration of the mass as an uncaused event (cf. Norton 2007: 24). I will use Norton's dome scenario in the following way: suppose that the acceleration of the mass has an effect e (for instance, the mass hits a little bell that is located at the base of the dome).[9] Further suppose that we posit very restricted worlds such that the worlds only contain the mass, the dome, the gravitational field, and the effect e. In these restricted dome-worlds, there exists no physically possible event that may play the role of an intervention on the mass, because *all* that exists in the restricted dome-worlds is the mass, the dome, the gravitational field, and the effect e. In other words, restricted dome-worlds conform to Norton's initial description of the dome-scenario – the idea that the

---

[9] This is an assumption that Norton is certainly not willing to go along with, because he denies that Newtonian mechanics (and other physical theories after Newton) describe causal relations.

"physical environment that is completely unchanging for an arbitrary amount of time" – in an extreme way by isolating the dome. Hence, in these restricted dome-worlds it is the case that (a) the uncaused acceleration of the mass itself is a cause of event *e*, and (b) a physically possible intervention does not exist. In other words, we have a case of causation without the physical possibility to intervene. I conclude that, if the restricted version Norton's dome-scenario is correct, then we have another example for the *Neither-Deterministic-Nor-Indeterministic Intervention Case*.[10]

The *Neither-Deterministic-Nor-Indeterministic Intervention Case* is *stronger* than the *No-Deterministic-But-Indeterministic Intervention Case*, because Woodward can easily admit that interventions, and causation in general, need not be deterministic as assumed in the *No-Deterministic-But-Indeterministic Intervention Case*. However, one could insist that the definition of an intervention variable at least appears to presuppose that an intervention variable *I* for *X* is a *deterministic* cause of *X*, because the formulation that "*I* acts *as a switch* for all the other variables that cause *X*" (cf. Woodward 2003: 98, my emphasis), as it stands, seems to refer to deterministic causation only.[11] If that were a correct way of understanding Woodward's definition, then the Uranium case would be more problematic for interventionists. But be that as it may. A charitable reading of Woodward's definition of an intervention should allow for indeterministic interventions.

At any rate, the *No-Deterministic-But-Indeterministic Intervention Case* is only problematic conditional on the assumption that only deterministic physically possible interventions are allowed. In contrast, the *Neither-Deterministic-Nor-Indeterministic Intervention Case* is unconditionally trouble for Woodward (as Woodward knows well), because there is *absolutely no* (neither deterministic nor indeterministic) physically possible way to intervene. For this reason, I will concentrate on the *Neither-Deterministic-Nor-Indeterministic Intervention Case*.

What do these counterexamples show? One might be inclined to object: "isn't the Big Bang case a rather far-fetched counterexample from physics?". However, even if these counterexamples are far-fetched ones from physics, they

---

[10] I'd like to thank Carl Craver and John Norton for very helpful discussions of the dome-scenario.
[11] Thanks to Michael Baumgartner for pointing this out.

should raise the following suspicion: in general and apart from these specific cases of the counterexamples, the possibility to intervene on the cause X seems to be completely irrelevant for the truth conditions of "*X* causes *Y*". Apart from this worry concerning the specific choice of counterexamples from physics, I think that the force of my arguments (that I will present in SECTION 4) does not necessarily depend on the existence of actual cases of the general form of the counterexample. All of my worries should still be justified even if it turned out that, as Woodward acknowledges, it is merely a "logically coherent possibility" (Woodward 2003: 130) that *X* is a cause of *Y* but there is no physically possible intervention $I = i$ on X – even if there are no actual (physical) cases of this kind.

Woodward rejects these kinds of counterexamples by pointing out that interventions need not even be physically possible. Woodward (2003: 128f) understands the relevant sense physical possibility as follows:

> An intervention $I = i$ is physically possible iff $I = i$ is consistent with some set of possible initial conditions and the actual laws.[12]

Woodward (2003: 128f) claims that the counterexamples show that an intervention on X is not be required to be physically possible. Instead of being physically possible, interventions are required to be possible in the sense that they are merely "logically possible" or not "ill-defined for conceptual or metaphysical reasons" (Woodward 2003: 128, 132). Note that, although Woodward introduces a distinction of logical, conceptual, and metaphysical possibility, he refers only to logical possibility in the large majority of cases and, further, he seems to use these kinds of modality interchangeably. For this reason, I will not explore the potential differences between these kinds of modality in this paper.

Is Woodward's strategy to deal with the counterexamples by weakening the required notion of possibility really convincing? I think it is *not*. In the next

---

[12] Woodward also considers another stronger reading of physical possibility: an intervention $I = i$ is *strongly* physically possible iff $I = i$ is consistent with *actual* initial conditions and the actual laws. However, Woodward – correctly, I think – dismisses strong physical possibility as a too demanding requirement for the interventionist theory of causation (cf. Woodward 2003: 128).

section, I will present two arguments against Woodward's strategy to deal with the counterexamples.

4. Two Arguments against the Need for Logically Possible Interventions

In this section, I will present two arguments against Woodward's above-mentioned strategy. The first argument aims to establish that interventions are dispensable for Woodward's semantic project as long as one allows for them to be merely logically possible. The second argument supports the claim that interventions with this modal character lead to the fatal result that interventionist counterfactuals are evaluated *inadequately*.

First Argument

The first argument addresses the claim that merely logically possible interventions can be dispensed with – which would be lead to a collapse of the interventionist theory as it stands. When I claim that merely logically possible interventions are dispensable I mean by this that interventions with this modal character fail to contribute non-trivially to the truth conditions of causal claims, i.e. interventions can be eliminated without loss. The conclusion of the first arguments is that interventions are dispensable for stating the truth conditions of causal claims such as "X is a direct type-level cause of Y". In other words, the first argument is a dispensability argument. I use the notion of dispensability in analogy to its use in indispensability arguments in philosophy of mathematics (cf. Putnam 1975, Field 1980): a term in a theory (in our case, the notion of an intervention) is dispensable if this term can be eliminated without compromising the strength of the theory (that is, the adequacy of an explication of causation).

Suppose that the causal claim "the Big Bang is a cause of Y" is true (for some Y). According to the interventionist theory, "the Big Bang is a cause of Y" is true iff the following two interventionist counterfactuals are true: (a) "if there were an intervention $I = i$ on the Big Bang such that the Big Bang occurred as it actually did, then $Y = y$ would be the case as it actually is", and (b) "if there were an intervention $I = i^*$ on the Big Bang such that the Big Bang were to occur

differently than it actually did, then *Y* would take a counterfactual value *y**". However, the existence of a logically possible intervention seems to be completely dispensable for stating the truth conditions of "the Big Bang is a cause of Y". I will now begin to argue for this claim by way of an example of an intervention on the Big Bang suggested by Tim Maudlin. Having pointed out that it is neither practically nor physically possible to intervene on the Big Bang, Maudlin presents an instructive example of a physically impossible[13] intervention on the Big Bang:

> It is surely *physically* impossible that the Big Bang have [sic!] been controlled […] but it nonetheless has *effects*. Physical considerations […] may leave the philosophical mind unimpressed. It is at least *metaphysically* possible to control the Big Bang. God might decide, for example, to employ the following scheme: if He is forgiving the universe will be open, ever-expanding; if He is jealous the universe will be closed […]. (Maudlin 2002: 149f)

Maudlin's key idea can be put as follows: although there is no physically possible intervention on the Big Bang, *God* (I use italics for variables) is a logically possible (deterministic or indeterministic) intervention variable on the variable *Big Bang* relative to the variable *Evolution of the Universe* (after the Big Bang).[14] Maudlin's scenario can be depicted by the following graph:

God

                    Big Bang ⟶ *Evolution of the Universe*

In this graph, the variables *God*, *Big Bang* and *Evolution of the Universe* have the following possible values:

- The variable *God* ranges over the set of possible values {forgiving; jealous}.

---

[13] Maudlin (2002: 150) describes the following scenario as "at least metaphysically possible". I think that for this reason Maudlin's example qualifies for my current purpose.

[14] One might object that "there is a cause of the Big Bang" is conceptually impossible, because, by definition, the Big Bang cannot have a physical cause. Since God need not be a physical being Maudlin's example nicely avoids the objection that an intervention on the Big Bang is conceptually ill-defined.

- The binary variable *Big Bang* ranges over the set of possible values $\{1, 2\}$. These values represent two different ways in which the Big Bang might have taken place, as Maudlin's scenario suggests that the two possible moods of god (i.e. to be jealous and to be forgiving) lead to different initial conditions of the universe.
- The variable *Evolution of the Universe* has the values $\{$open, closed$\}$.

Let me re-describe Maudlin's divine intervention scenario in terms of interventions on *Big Bang*.

(a) *God* is a cause of the *Big Bang*.

(b) There is at least some value of *God* such that if *God* takes this value, then *Big Bang* depends *only* on *God* and *Big Bang* depends on no other variables, i.e. *God* is the only cause of *Big Bang*.

(c) *God* is not a direct cause of *Evolution of the Universe*, and if *God* is a cause of *Evolution of the Universe* then *God* is an indirect cause of *Evolution of the Universe* via a causal path leading through *Big Bang* and a – possibly empty – set of intermediate variable $Z_1, \ldots, Z_n$.

(d) *God* is probabilistically independent of other causes $W_1, \ldots, W_n$ of the *Evolution of the Universe*, which are not on a causal path leading from the *Big Bang* to *the Evolution of the Universe*.

(e) *God* does not alter the relationship between *Evolution of the Universe* and its causes $W_1, \ldots, W_n$, which are not on a causal path leading from *Big Bang* to *Evolution of the Universe*.

If *God* takes either of the values in its range, then the probability increases that the Big Bang takes place in a different manner (as indicated by *Big Bang*'s range $\{1, 2\}$) and different evolutions of the universe result (as indicated by *Evolution of the Universe*'s range $\{$open, closed$\}$). All of this is in perfect accordance with Maudlin's story about the possible scenarios. Thus, each of the statements *God = forgiving* and *God = jealous* represents a merely logically possible intervention on

Big Bang relative to *Evolution of the Universe*. Maudlin's scenario illustrates a merely logically possible intervention on the Big Bang.

Clearly, interventionists are committed to accept the presented intervention – because it is logically possible that there is something about the moods of God that has the desired intervening effect on the Big Bang. Obviously, it is easy to come up with many other examples of merely logically possible interventions (e.g., the physical constants case and the restricted dome case, see SECTION 3). If this is correct, then a problem arises for the interventionist: once one admits merely logically possible interventions, interventions become semantically dispensable. Let me argue for this claim by focusing on the Big Bang scenario: suppose that the causal claim "*Big Bang* is a cause of *Evolution of the Universe*" is true. Interventionists claim that "*Big Bang* is a cause of *Evolution of the Universe*" is true, roughly, iff the following two interventionist counterfactuals are true: (a) "if there were an intervention *God = forgiving* on *Big Bang* such that the *Big Bang = 1*, then it would be the case that *Evolution of the Universe = open*", and (b) "if there were an intervention *God = jealous* on *Big Bang* such that the *Big Bang = 2*, then it would be the case that *Evolution of the Universe = closed*". In this scenario, the existence of a logically possible intervention apparently becomes *dispensable* for stating the truth conditions of "Big Bang is a cause of *Evolution of the Universe*", because the existential claim "there is a logically possible intervention on *Big Bang*" (say, in the second counterfactual) does not add anything substantial to the following counterfactual (in which the existential claim about the intervention is *eliminated*):

> "if it were the case that *Big Bang = 2*, and other causes of *Evolution of the Universe*, W_1, …, W_n, had been held fixed (if there are any), then it would be the case that *Evolution of the Universe = closed*, *and* it is not a logical contradiction to say that the *Big Bang = 2*".

To say the least, it is not obvious what one gains by adding an existential claim about a merely logically possible intervention to the conditional above. We seem to be able to provide a complete description of the causal information contained in

the Big Bang scenario by simply using counterfactuals that do not refer to interventions. However, the *argument* for the dispensability of Woodwardian interventions is ultimately justified in three ways. The details of the dispensability argument are provided in Reutlinger (2011). I merely intend to provide a sketch of the arguments here.

First, if we can find a successful semantics for counterfactuals that does not rely on Woodwardian interventions, then Woodwardian interventions are dispensable for stating the truth-conditons of counterfactuals. Several standard accounts of semantics for counterfactuals *are* in fact able to successfully evaluate the conditional above without employing the Woodwardian notion of an intervention (these standard accounts are presented in the *second argument* below; recent improvements of the standard accounts include Pearl 2000, Schaffer 2004, Leitgeb forthcoming).[15] Thus, Woodwardian interventions are dispensable for stating the truth-conditions of counterfactuals.

Second, if a counterfactual theory of causation – that makes use of these (improved) standard semantics for counterfactuals – is able to satisfy the *criteria of adequacy* for a theory of causation *at least as well as* the interventionist theory, then interventions seem to be dispensable for an adequate theory of causation. There are such theories of causation. Thus, interventions are dispensable.

Let me provide some details and examples here. For instance, an adequate theory of causation is expected to satisfy the following criteria of adequacy:

a. an adequate theory ought to explicate several *kinds of causation* that are referred to in the sciences (e.g. actual causes, type-level causes, deterministic and indeterministic causes, contributing and total causes).

b. An adequate theory of causation has to account for several typical *features of causation* (e.g. the time-asymmetry of causation, the sensitivity to changes in background conditions, a distinction between genuine causation and accidental correlation).

---

[15] Although Pearl (2000) relies on the notion of an intervention, a Pearlian intervention differs from a Woodwardian intervention because the former is not committed to the view that interventions have to be modeled as exogenous causes. A Pearlian intervention simply amounts to assigning a specific value to variable X in a causal model while other causal influence are disrupted. This assignment of a value to X is not required to be modeled as the result of an exogenous cause I (i.e. an intervention variable as in the Woodwardian case). I do not raise any objections to the Pearlian kind of interventions.

c. It is a criterion of adequacy to provide a successful description of intuitively possible *causal scenarios* (such as preemption scenarios, common cause scenarios, and common-cause scenarios).

In my (2011: chapter 8), I argue for a counterfactual theory of causation – the comparative variability theory – which preserves crucial features of the interventionist account (without relying on interventions, of course). Most importantly, my counterfactual approach adopts a key feature of Woodward's theory that guarantees its (limited) success: specific background conditions (i.e. variables that are not on a directed path from the cause variable to the effect variable) have to be *held fixed* in a possible situation in which the cause variable takes a counterfactual value.[16] The upshot of my approach is that causal dependence amounts to *conditional* counterfactual dependence – i.e. conditional on a certain fixed causal background, a causal field (cf. Field 2003: 452). However, this feature of *holding fixed the causal background* can be preserved *even if* the notion of an intervention is *eliminated*. I argue that the comparative variability theory is able to satisfy the above-listed criteria of adequacy at least as well as the interventionist (and, in some cases, even better). Thus, interventions are dispensable. However, this result does *not* exclusively depend on my particular theory of causation. Other counterfactual theories that do not employ the strong Woodwardian notion of an intervention (such as Field 2003, Schaffer 2004, Halpern & Pearl 2005, Menzies 2007) can also be used to establish the dispensability argument.

Third, I will argue in the second argument (below) that there is no easy way out for interventionists: it is unfortunately *not* the case that interventions are *dispensable but harmless*. Woodwardian interventions are not harmless because they lead to a severe problem regarding the evaluation of counterfactuals.[17] If interventions create their own problems (given that we can state the truth-conditions of counterfactual conditionals and causal claims without referring to

---

[16] Woodward uses the notion of the redundancy range in order to hold fixed the causal background (cf. Woodward 2003: 83).

[17] In Reutlinger (under review), I present another problem for interventionists: they fail to account for the time-asymmetry of causation.

interventions), then we have an additional reason to eliminate interventions without regret.

I conclude from these three considerations that intervention can be eliminated without loss. In other words, they are dispensable. Tim Maudlin suggests a similar conclusion with respect to his scenario of a physically impossible divine intervention on the Big Bang:

> The temptation to use these fantastic theological scenarios does not arise from any deep conceptual connection between causation and signals [or 'controllability' as Maudlin also says – these expressions are equivalent with possible interventions in our case]. *The scenarios rather illustrate a counterfactual connection: had the Big Bang been different so would the later course of the universe have been.* (Maudlin 2002: 150, my emphasis)

The upshot of discussing Maudlin's divine intervention scenario is that merely logically possible interventions turn out to be eliminable without loss. As Maudlin points out, the decisive issue is counterfactual dependence which can be stated without reference to (merely logically possible) interventions.

Second Argument

The second argument against logically possible interventions relies on three alternative approaches to the semantics of counterfactual conditionals. Let me briefly motivate why determining the truth conditions of counterfactuals matters for our present concerns. Let us recall the motivation for employing interventions. The main motivation for introducing interventions is closely tied to counterfactuals. Woodward's definitions of causation imply a specific kind of counterfactuals whose arguments are propositions that some variable takes a certain value in its range: "if there were an intervention $I = i$ on X such that $X = x$, then $Y = y$ would be the case" (cf. Woodward 2003: 15). In the first argument, I argued that logically possible interventions are dispensable for the truth conditions of causal statements. Now, in the second argument, I will change the perspective: for the sake of the argument, I will grant that merely logically possible interventions do play an indispensable role for the interventionist theory. I argue that Woodward has to pay price for this claim: merely logically possible interventions lead to the fatal result that interventionist counterfactuals involving

existential claims about merely logically possible interventions are inadequately evaluated as false. If this reasoning were sound, this would also be an unwelcome result for interventionists. Call this the *Problem of Inadequate Evaluation*. I will show that *Problem of Inadequate Evaluation* is fatal for the interventionist theory of causation by proceeding in this way: I argue that the *Problem of Inadequate Evaluation* remains a problem for these standard semantics. And, I will argue that even if one relies on one of a variant of interventionist semantics one fails to provide a satisfying solution to the *Problem of Inadequate Evaluation*, because relying on interventionist semantics is either (i) ad hoc, (ii) non-practical, or (iii) subject to the dispensability argument. For brevity's sake, I will restrict the second argument to the strong case of merely logically possible interventions on the Big Bang.

The modal character of logically possible interventions is not as innocent as Woodward suggests, because possible worlds at which interventions of this kind exist deviate strongly from the actual world with respect to laws of nature. This is problematic for interventionists because: (a) interventionists crucially rely on (interventionist) counterfactuals, and (b) worlds with strongly different laws are *ruled out* for the evaluation of counterfactuals by the three standard approaches to the semantics of counterfactuals. Consequently, if we evaluate counterfactuals in worlds where merely logically possible interventions on the antecedent variable exist, then the counterfactuals are false (*Problem of Inadequate Evaluation*).

Let me argue for this claim in more detail. Assume there is a world *w* such that a merely logically possible intervention on the Big Bang exists in *w*. By definition, worlds where merely logically possible interventions exist are physically impossible worlds, i.e. worlds differing in laws from the actual world. According to our example, a world where an intervention is carried out on the Big Bang instantiates laws of nature that *differ* from the actual laws (namely, it is in accordance with *these other-worldly* laws that there is a deterministic or an indeterministic intervention on the Big Bang). Yet, is it the case that *any* violation of actual laws in a world *w* rules out *w* for the evaluation of counterfactuals (which are true or false in the actual world)? The answer is "No". One might

wonder whether a logically possible intervention is just a *local* violation of a law. For this reason, an intervention is sometimes (cf. Woodward 2003: 135f) compared to a small miracle in the similarity heuristic of Lewis's semantics (cf. Lewis 1979: 47f.). According to Lewis, the closest possible worlds are those that differ minimally from the actual world in particular matters of fact such that a counter-nomological (i.e. violating the laws of the actual world) event *m* occurs just before the antecedent event *c* (the cause) occurs. The event *m* is a small miracle because after the occurrence of *m* (i.e. after setting *new* initial conditions) the world again conforms to the actual laws. But note that this comparison is not wholly justified: although interventions and small miracles are both changes of particular matters of fact in possible worlds, Lewis does not argue that small miracles are *causes* of the antecedent event *c* in question. And it is precisely the *causal* character of interventions – which interventionists explicitly highlight – that leads into trouble.

There is a further and even more important reason to believe that an intervention on the Big Bang has to be clearly distinguished from a small miracle. In the closest worlds at which the actual laws are preserved to a maximal degree, there is no small miracle (that is, no set of counterfactual initial conditions) such that a state of the world evolves in which the Big Bang occurs differently than it actually does. In other words, such an intervention is not even possible in the sense of physical possibility. As opposed to that, worlds in which merely logically possible interventions exist instantiate physically impossible (viewed from the actual world) *nomic* connections between the intervention variable *God* and the cause variable *Big Bang*. Woodward seems to be committed to these physically impossible nomic connections: Woodward argues that true (actual and type-level) causal statements are "backed up" by at least non-strict, invariant generalizations (cf. Woodward 2003: 146f, 244; also his definition M of causation, p. 59, implies invariant relation between causal relata; pp. 245-255 on the claim that causal relations are backed up by invariant generalizations). In other words, there is an invariant (or, in some cases, nomic), *repeatable* connection between a type of intervention (e.g. *God = forgiving*) and a type of cause (e.g. *Big Bang = 1*) in world *w*. Thus, if it is true in a world *w* that there is an intervention on *Big Bang*

(i.e., according to the definition of an intervention, at least "*God* is a cause of *Big Bang*" is true in *w*), then there is an invariant generalization (or law) in world *w* that "backs up" this the causal relation between the intervention variable *God* and the cause variable *Big Bang*. I conclude that interventions differ from small miracles. Merely-logically possible-intervention-worlds are worlds with *truly* different, non-actual laws in the sense that they are counter-legal to a higher degree than small-miracle-worlds are.

This result is quite *problematic*: according to the three standard semantics of counterfactuals, worlds truly differing in laws from the actual world are not adequate to evaluate counterfactuals.

(I)   According to Lewis's possible worlds semantics, a counterfactual is (non-vacuously) true iff the consequent is true in the closest antecedent-worlds (cf. Lewis 1973: 16). Worlds that differ in laws in higher degree than small miracle worlds are not among the closest antecedent-worlds that are relevant for fixing the truth conditions of counterfactuals (cf. Lewis 1979: 47f). Thus, (merely logically possible) intervention-worlds are *not among the closest* worlds, i.e. they are big-miracle worlds involving global violations of actual laws.

(II)   According to the meta-linguistic account (Goodman 1983), a counterfactual is true iff the consequent can be logically derived from a set of premises consisting of (1) laws of nature, and (2) the antecedent and other initial conditions. Using truly non-actual law statements as premises seems to be misguided, if one would like to evaluate a counterfactual at the actual world. According to the Goodmanian approach, it is the case that if a counterfactual is supposed to be true at the *actual* world, then the consequent has to be inferred from (maximally) *actual* laws (that do not include counterfactual invariant connections between a merely logically possible type of intervention and a type of cause) and singular statements.

(III)   According to the suppositional theory (Ramsey 1929: 247, Adams 1975: Chapter 4, Skyrms 1994, Edgington 2008), a conditional is acceptable (to the degree that) if one "hypothetically" adds the antecedent to one's

knowledge, then one also believes that the consequent is true.[18] This procedure is known as the Ramsey-Test. Beliefs about which laws are true in the actual world are important for carrying out the Ramsey-Test (cf. Leitgeb forthcoming Section 2). Supposing that p and adding p to a stock of beliefs including *truly non-actual* laws does not make sense. The *relevant* epistemically possible worlds (in a communicative situation) are at least partly constituted by maximally *actual* laws (that, again, do not include counterfactual invariant connections between a type of a merely logically possible intervention and a type of cause).

To sum up, merely logically possible worlds in which there are interventions on the Big Bang are truly counter-legal (i.e. they are counter-legal to a higher and more global degree than small-miracle-worlds). This result leads to a conflict with the three standard semantics of counterfactuals: if we evaluate a counterfactual in worlds with truly non-actual laws, then we have to evaluate this counterfactual as *false*. Therefore, worlds where interventions on the Big Bang exist are not appropriate for evaluating counterfactuals: they lead to inadequate results (*Problem of Inadequate Evaluation*). For instance, we have to say, from an interventionist point of view, that the counterfactual "if the Big Bang had occurred differently, then it universe would have evolved differently" is false, because interventionists reformulate this conditional by adding an existential claims about a merely logically possible intervention on the Big Bang to the antecedent of the conditional.

Now, the burden of proof is on the interventionists' side: interventionists have to show (a) why the standard approaches are wrong in rejecting merely logically possible interventions, and (b) how to evaluate counterfactuals with a

---

[18] The basic idea of the Ramsey-Test was historically first used to determine the assertability conditions of indicative conditionals by Adams (1975). Adams analyzes the degree of assertability of an indicative conditional in terms of subjective probability: the degree of assertability of "if p, then will q" equals the subjective probability of q given p. Although this fact is often ignored, Adams (1975: Chapter 4) also argues that the suppositional theory can be applied to *counterfactuals* in a slightly modified way. In particular, Skyrms has developed the most sophisticated account of Adams's original idea. According to Skyrms (1994: 13-15), one determines the pragmatic meaning of a counterfactual in terms of the degree of assertability: the degree of assertability of the counterfactual "if it were the case that p, then it would be that case that q" equals the subjectively expected objective conditional probability of q given p.

semantics that includes interventions. This is, of course, an unpleasant result for interventionists.

Yet, let us take a step back. Maybe the situation is not so unpleasant for interventionists after all. A natural response to my second argument could consist in adopting an *interventionist semantics* of counterfactuals. As already mentioned in Section 1, an interventionist might propose an interventionist version of possible worlds semantics along the following lines: a counterfactual is true iff the consequent is true in the closest antecedent-worlds. The closest worlds are those worlds in which the antecedent is true in virtue of an intervention on the antecedent-variable (cf. Reutlinger 2011: Chapter 3 for a more detailed discussion of this semantics as well as for Goodmanian and suppositionalist versions of interventionist semantics).

Although turning to interventionist semantics might be a tempting option for an advocate of an interventionist theory of causation, this strategy does not help to refute the second argument for three reasons.

First, adopting an interventionist semantics comes close to an *ad hoc move* in order to vindicate the interventionist theory of causation. It might be correct that the interventionist semantics can handle counterexamples to Lewis's original semantics of counterfactuals. Dealing with these counterexamples successfully is a prima facie reason in favor of interventionist semantics. However, the interventionist semantics does not explicitly presuppose a specific modal character of interventions. Building the *problematic* assumption of the existence of merely logically possible interventions into the semantics of counterfactuals seems to be motivated by one goal only: one excludes cases in which intervening is physically impossible (such as the Big Bang scenario) as counter-examples to the interventionist theory. Therefore, legitimizing merely logically possible interventions by incorporating them into the interventionist semantics is an ad hoc move to defend the interventionist theory, and therefore it is not convincing.

Second, Woodward argues that his interventionist theory is a "practical" theory of causation – by contrast Lewis's counterfactual theory and Dowe's conserved quantity theory are "impractical" theories (cf. Woodward 2003: 28-38). According to Woodward, a theory of causation is "practical" if it succeeds to

"connect causal knowledge with some goal that has practical utility" (cf. Woodward 2003: 30). Examples of a goal that has practical utility are successful manipulation (for instance, in medical, political, and scientific contexts) and reliable experimentation in the sciences. So, Woodward claims that if we come to believe that "X causes Y" and we understand the truth conditions of this claim in accord to the interventionist theory, then we can easily connect this causal knowledge to practical goals such as manipulation and experimentation. For instance, if we endorse an interventionist theory of causation, we know that we are in principle able to manipulate the effect Y by manipulating the cause X. We also know that if we want to design an experimental set-up which involves Y, then the belief that X causes Y is indeed relevant (for instance, because we might want to isolate Y from its cause X, or because we might want to exploit the causal relation between X and Y for manipulation within the experiment). However, even if one agrees with Woodward that theories of causation ought to be practical in the sense he suggests a problem arises for the interventionist: if one allows interventions to be merely logically possible, then the interventionist theory loses its practical character. For example, suppose that there it is merely logically possible to intervene on the Big Bang, and suppose that the statement "the Big Bang causes event e" is true. It is hard to see which practical goal can be achieved by this causal knowledge. For instance, if we had the goal to manipulate e then the interventionist theory does not explain how we can realize this goal because it is merely logically possible to change e by intervening on the Big Bang. I doubt that the knowledge that the occurrence of e counterfactually depends on the change of the Big Bang which is the outcome of a merely logically possible intervention satisfies Woodward's own requirement that a theory of causation ought to explain why we pursue certain practical goals. Quite the opposite is the case: accepting merely logically possible interventions undermines the appraised practical character of the interventionist theory of causation, because the link to practical goals (such as successful manipulation and reliable experimental practice in the sciences) is disrupted in cases involving interventions with this modal character. Therefore, relying on an interventionist semantics which allows merely logically possible interventions is not convincing.

Third, if an interventionist attempts to refute the second argument by adopting an interventionist semantics, then merely logically possible interventions still remain subject to the dispensability argument. As argued above, merely logically possible interventions are dispensable. If interventions are dispensable then relying on an interventionist semantics obviously fails to reject the second argument.

5. Conclusion

Interventionists require that interventions be merely logically possible. I have argued against this requirement by establishing two claims: *first*, merely logically possible interventions are dispensable for stating the truth conditions of causal claims. If this is true, then the interventionist theory, as it stands, collapses. *Second*, counterfactuals involving merely logically possible interventions lead to the fatal result that interventionist counterfactuals are inadequately false. This is likewise an unwelcome result for interventionists. I conclude that if we attempt to master the tasks of explicating causal concepts and stating the truth conditions of causal claims we best get rid of Woodwardian interventions.

Acknowledgements

References

Adams, Ernest. 1975. *A Theory of Conditionals. An Application of Probability to Deductive Logic*. Dordrecht: Reidel.

Campbell, John. 2007. "An Interventionist Approach to Causation in Psychology." In *Causal Learning: Psychology, Philosophy and Computation*, Alison Gopnik and Laura Schulz (eds.), Oxford: Oxford University Press, pp. 58-66.

Craver, Carl. 2007. *Explaining the Brain. Mechanisms and the Mosaic Unity of Neuroscience*, Oxford: Clarendon Press.

Edgington, Dorothy. 2008. "Counterfactuals." *Proceedings of the Aristotelian Society* 108: 1-21.

Field, Hartry. 1980. *Science Without Numbers: A Defence of Nominalism*, Oxford: Blackwell.

Field, Hartry. 2003. "Causation in a Physical World." In M. Loux and D. Zimmerman (eds.), *The Oxford Handbook of Metaphysics*, Oxford: Oxford University Press, 435-460.

Goodman, Nelson. 1983. *Fact, Fiction and Forecast*, Fourth Edition, Cambridge, MA: Cambridge University Press.

Halpern, Joseph, and Judea Pearl. 2005. "Causes and Explanations: A Structural-Model Approach. Part I: Causes." *British Journal for the Philosophy of Science* 56, 843-887.

Hausman, Daniel. 2009. "Laws, Causation, and Economic Methodology.", In Kincaid, H., and D. Ross (eds.), 2009, *The Oxford Handbook of Economics*, Oxford: Oxford University Press, 35-54.

Hüttemann, Andreas. 2004. *What's Wrong with Microphysicalism?*, London: Routledge.

Leitgeb, Hannes. forthcoming. "Probabilistic Semantics for Counterfactuals. Part A." to appear in *Review of Symbolic Logic*.

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

Lewis, David. 1979. "Counterfactual Dependence and Time's Arrow." In David Lewis (1986), *Philosophical Papers II*, Oxford: Oxford University Press, 32-51.

Maudlin, Tim. 2002. *Quantum Non-Locality and Relativity*, Oxford Blackwell.

Menzies, Peter, and Huw Price. 1993. "Causation as a secondary quality." *British Journal of the Philosophy of Science* 44: 187-203.

Menzies, Peter. 2007. "Causation in Context." In Huw Price and Richard Corry (eds.) (2007), pp. 191-223.

Norton, John. 2007. "Causation as Folk Science.", In Huw Price & Richard Corry (eds) (2007): *Causation, Physics and the Constitution of Reality*, Oxford: Oxford University Press, 11–44.

Pearl, Judea. 2000. *Causality: Models, Reasoning and Inference*, Cambridge: Cambridge University Press.

Putnam, Hilary. 1975. "Philosophy of Logic." reprinted in *Mathematics Matter and Method: Philosophical Papers, Volume 1*, 2nd edition, Cambridge: Cambridge University Press, pp. 323–357.

Ramsey, Frank P. 1929. "General Propositions and Causality." In F. P. Ramsey: *Philosophical Papers*, D.H. Mellor (ed), pp. 237-255.

Reutlinger, Alexander. 2011. *The Trouble with Interventions. A Theory of Causation in the Special Sciences*, University of Köln, unpublished dissertation.

Reutlinger, Alexander. under review. "Woodward meets Russell." *British Journal for the Philosophy of Science*.

Schaffer, Jonathan. 2004. "Counterfactuals, Causal Independence, and Conceptual Circularity", *Analysis* 64.4: 299-309.

Shapiro, L., and Sober, E. 2007. "Epiphenomenalism. The dos and don'ts." In *Thinking about causes: From greek philosophy to modern physics*, edited by G. Wolters and P. Machamer, 235–264. Pittsburgh: University of Pittsburgh Press.

Skyrms, Brian. 1994. "Adams Conditionals." In *Probability and Conditionals. Belief Revision and Rational Decision*, Ellery Eells and Brian Skyrms (eds), Cambridge/UK: Cambridge University Press, pp. 13-26.

Waters, Ken. 2007. "*Causes that Make a Difference.*" *The Journal of Philosophy* CIV: 551-579.

Woodward, James. 2003. *Making Things Happen. A Theory of Causal Explanation*, Oxford: Oxford University Press.