Of Miracles and Interventions

Luke Glynn

ABSTRACT

In *Making Things Happen*, James Woodward influentially combines a causal modeling analysis of actual causation with an *interventionist* semantics for the counterfactuals encoded in causal models. This leads to circularities, since interventions are defined in terms of both actual causation and interventionist counterfactuals. Circularity can be avoided by instead combining a causal modeling analysis with a semantics along the lines of that given by David Lewis, on which counterfactuals are to be evaluated with respect to worlds in which their antecedents are realized by *miracles*. I argue, *pace* Woodward, that causal modeling analyses perform just as well when combined with the Lewisian semantics as when combined with the interventionist semantics. Reductivity therefore remains a reasonable hope.

1. Introduction

It has become common to deploy causal models in attempting to analyze actual (or token) causation.[1] I will call such accounts *causal modeling analyses of actual causation* (CMAACs). Causal models are systems of structural equations, which encode patterns of counterfactual dependence between variables.[2] The variables in question can be used to represent the relata of actual causal relations, which are usually taken to be token events. In virtue of the fact that causal models encode counterfactuals, CMAACs are a variety of *counterfactual analysis of actual causation* (CAAC).

Whether CMAACs are reductive turns (in part) upon the semantics the counterfactuals encoded in causal models are given. Woodward (2003) has influentially combined a CMAAC (ibid. p. 84) with an *interventionist* semantics for such counterfactuals (ibid. e.g. pp. 42-44). *Interventions* are defined by Woodward (ibid. p. 98) in terms of (among other things) (a) *actual causation*; and (b) *directed paths in causal models*.[3] The appeal to *each* of these notions stands in the way of a reduction of actual causation to counterfactual dependence, where the latter is reductively analyzed in terms of more basic, non-causal facts.

Woodward argues that CAACs – including CMAACs – founded upon the interventionist semantics perform better than those founded on the most well-known rival semantics, due to Lewis (1979). Lewis's proposal is that the counterfactuals relevant to actual causation be evaluated

---

[1] See, for example, Glymour and Wimberly (2007), Halpern (2008), Halpern and Hitchcock (ms.), Halpern and Pearl (2005), Hitchcock (2001a, 2007), Menzies (2004a), Pearl (2009, Ch. 10), and Woodward (2003, pp. 74-86).

[2] See Halpern and Hitchcock (2010, p. 385; ms.), Halpern and Pearl (2005, pp. 846-851), Hitchcock (2001a, pp. 279-284; 2007, pp. 499-504), Menzies (2004a, pp. 821-822), Pearl (2009, pp. 202-215), and Woodward (2003, pp. 42-43).

[3] Woodward thus makes a *twofold* appeal to causal models. *First*, he appeals to them at the level of his analysis of actual causation, which is a CMAAC. *Second*, he appeals to them at the level of his *semantics* for the counterfactuals encoded in causal models.

with respect to worlds in which their antecedents are realized by *miracles*. Since Lewis's semantics does not appeal to (a) or (b) (or to any other causal or counterfactual notion), it is compatible with reduction. But Woodward (*op. cit.* pp. 133-145) argues that CAACs founded on Lewis's semantics are subject to counterexample, which can be avoided only by adopting the interventionist semantics.

I will argue that this is not so. Woodward's putative counterexamples tell only against the combination of the Lewisian semantics with a naïve CAAC, which takes counterfactual dependence to be necessary for actual causation. They do not tell against the combination of the Lewisian semantics with sophisticated CMAACs. Since CMAACs are independently attractive, a reductive analysis of actual causation remains on the cards.

The plan is as follows. After introducing the notion of a causal model and briefly describing standard CMAACs (§2), I outline the interventionist semantics, and exhibit the circularities it introduces (§3). I then compare the interventionist semantics to the Lewisian semantics (§4). I outline Woodward's alleged counterexample to CAACs founded upon Lewis's semantics (§5). I argue that such examples are not telling against the combination of the Lewisian semantics with standard CMAACs. I then advance a counterexample of my own (§6), but argue that a slight modification of the Lewisian semantics allows this to be handled too. The conclusion (§7) is that a reductive CMAAC remains a reasonable hope.

Two preliminary points. First: my concern is not with whether either semantics is adequate to the ordinary language counterfactual, but rather with which best meets the needs of leading CAACs (in particular, CMAACs). Even if the semantics needed by CAACs bears only a loose connection to those appropriate to the ordinary language counterfactual, this does not tell against CAACs.[4] As Collins *et al.* (2004, p. 9) point out, "[t]he counterfactual analyst should not worry: After all, she is doing the metaphysics of causation, and not the semantics of some fragment of English."

---

[4] At most it tells against the terminological decision to call them 'counterfactual analyses' of actual causation (cp. Collins *et al.* 2004, p. 9).

Second: I do *not* seek to argue that the combination of (i) *standard* CMAACs with (ii) *an interventionist semantics for the counterfactuals encoded in casual models* yields an *extensionally inadequate* theory of causation. My aim is, rather, to defend (in the face of Woodward's putative counterexamples) the *extensional adequacy* of the combination of (i) *standard* CMAACs with (iii) *a Lewisian semantics for the counterfactuals encoded in causal models*. For the most part, I will be arguing that the combination of (i) with (iii) *yields the same verdicts* about causation as the combination of (i) with (ii) does. The tie-breaker between (ii) and (iii) is that adopting (iii) (but not (ii)) maintains the possibility of a reduction of causation to non-causal facts.

2. Causal Models

A causal model **M** is an ordered pair $\langle \mathbf{V}, \mathbf{E} \rangle$ where **V** is a set of variables, and **E** is a set of structural equations that encode the (non-backtracking) counterfactual dependencies among the variables in **V**.[5] Specifically, each equation in **E** specifies the value of some variable $V \in \mathbf{V}$, with V appearing on the left-hand side of the equation in question. A variable $W \in \mathbf{V} \setminus V$ appears on the right-hand side of the equation for V just in case the value of V counterfactually depends upon that of W for some assignment of values to all variables in $\mathbf{V} \setminus V$, W. If no variable appears on the right-hand side of the equation for V, then V is *exogenous*. In that case **E** simply contains an equation $V = v$ setting V equal to its actual value.

A model $\mathbf{M} = \langle \mathbf{V}, \mathbf{E} \rangle$ can be given a graphical representation by taking the variables in **V** as the vertices of the graph and drawing a directed edge (or 'arrow') from one variable to another just in case the value of the second is a function of the value of the first according to the structural equations in **E**. An ordered sequence $\langle X, V_1, ..., V_m, Y \rangle$ is a *directed path* in **M** just in case $X, V_1, ..., V_m, Y \in \mathbf{V}$ and there is a directed edge from X to V_1, a directed edge from V_1 to ...

---

[5] Thus, although they are standardly called 'causal' models, such models don't encode *causal* relations, but rather relations of (non-backtracking) counterfactual dependence. (Their alternative name – 'Structural Equations Models' – is less misleading in this regard.)

V_{m}, and a directed edge from V_{m} to Y.

CMAACs appeal to the notion of *counterfactual dependence under a contingency*.[6] A *contingency* is defined relative to a directed path $\mathbf{P} = \langle X, V_1, ..., V_{\mathrm{m}}, Y \rangle$ from X to Y in a model $\mathbf{M} = \langle \mathbf{V}, \mathbf{E} \rangle$ containing $\mathbf{P}$. Specifically, a contingency γ: $Z_1 = z_1^*$, ..., $Z_{\mathrm{n}} = z_{\mathrm{n}}^*$ is an assignment of values to the variables $Z_1, ...,$ and Z_{n} in $\mathbf{V} \backslash X, V_1, ..., V_{\mathrm{m}}, Y$: that is, to the 'off path' variables. Suppose that x and y are among the possible values of X and Y respectively. Then, $Y = y$ *counterfactually depends upon $X = x$ under the contingency γ* just in case there is a possible value x' ($x' \neq x$) of X and a possible value y' ($y' \neq y$) of Y such that the following pair of counterfactuals holds:

(†) $X = x \wedge Z_1 = z_1^* \wedge ... \wedge Z_{\mathrm{n}} = z_{\mathrm{n}}^* \ \square\!\!\rightarrow Y = y$;

(‡) $X = x' \wedge Z_1 = z_1^* \wedge ... \wedge Z_{\mathrm{n}} = z_{\mathrm{n}}^* \ \square\!\!\rightarrow Y = y'$.

That is, $Y = y$ counterfactually depends upon $X = x$ under the contingency γ just in case $Y = y$ counterfactually depends upon $X = x$ when all off-path variables are held fixed at the values they receive in γ.

The truth-values of (†) and (‡) can be evaluated with respect to a model $\mathbf{M} = \langle \mathbf{V}, \mathbf{E} \rangle$ such that $X, Z_1, ..., Z_{\mathrm{n}}, Y \in \mathbf{V}$. Counterfactual (†) is evaluated by replacing the equations for $X, Z_1, ...,$ and Z_{n} in $\mathbf{E}$ with the equations $X = x$, $Z_1 = z_1^*$, ..., and $Z_{\mathrm{n}} = z_{\mathrm{n}}^*$, while leaving all other equations intact. This yields a new set of equations $\mathbf{E'}$. Counterfactual (†) is true in the original model $\mathbf{M} = \langle \mathbf{V}, \mathbf{E} \rangle$ just in case, in the solution to $\mathbf{E'}$, $Y = y$. Counterfactual (‡) is evaluated analogously, but with the equation for X instead replaced with $X = x'$. Hitchcock (2001a, p. 275) dubs counterfactuals like (†) and (‡) *explicitly non-foretracking (ENF) counterfactuals*. The idea is that, because the values of off-path variables are held fixed, the counterfactual influence of X is not allowed to foretrack to Y along paths other than $\mathbf{P}$.

With this background in place, we are now in a position to describe *standard* CMAACs,

---

[6] Terminology due to Halpern and Pearl (2005, p. 844).

which have the form SCMAAC (I will call analyses of this form 'SCMAACs' in what follows).

(SCMAAC) Let c and e be distinct, actual events and let $X = x$ and $Y = y$ represent the occurrence of c and e respectively, with all other possible values of X and Y representing states of affairs incompatible with c and e. Then c is an *actual cause* of e just in case there is at least one *appropriate* causal model $\mathbf{M} = \langle \mathbf{V}, \mathbf{E} \rangle$ containing a directed path $\mathbf{P} = \langle X, V_1, \ldots, V_m, Y \rangle$ from X to Y such that $Y = y$ counterfactually depends upon $X = x$ in $\mathbf{M}$ under some *permissible* contingency γ relative to $\mathbf{P}$. That is, there is a permissible assignment γ: $Z_1 = z_1^*, \ldots, Z_n = z_n^*$ of values to the variables $Z_1, \ldots,$ and Z_n in $\mathbf{V} \backslash X, V_1, \ldots, V_m, Y$ (i.e. to the 'off-path' variables), some value x' ($x' \neq x$) of X, and some value y' ($y' \neq y$) of Y such that the ENF counterfactuals (†) and (‡) are true in $\mathbf{M}$.

This analysis schema contains two as yet unexplicated notions: an *appropriate* causal model, and a *permissible* contingency. For a model $\mathbf{M} = \langle \mathbf{V}, \mathbf{E} \rangle$ to count as *appropriate*, it must at least be the case that the equations in $\mathbf{E}$ entail no false counterfactuals (Hitchcock 2001a, p. 287) and that the values of distinct variables in $\mathbf{V}$ represent distinct events (ibid.; also Hitchcock 2007, pp. 502-503).[7]

The simplest account of what constitutes a *permissible* contingency (cp. Hitchcock 2001a, pp. 286-287; and Woodward 2003, p. 77) is that γ: $Z_1 = z_1^*, \ldots, Z_n = z_n^*$ is a permissible contingency (relative to a path $\mathbf{P} = \langle X, V_1, \ldots, V_m, Y \rangle$ in a model $\mathbf{M} = \langle \mathbf{V}, \mathbf{E} \rangle$) iff $Z_1 = z_1^*, \ldots,$ and $Z_n = z_n^*$ are the *actual* values of the off-path variables $Z_1, \ldots,$ and Z_n (that is, the values that they receive in the solution to $\mathbf{E}$). Call this the 'actuality' account of permissibility.

According to a second account, described by Glymour and Wimberly (2007, esp. p. 58) and Hitchcock (2001a, pp. 289-290), γ: $Z_1 = z_1^*, \ldots, Z_n = z_n^*$ is a permissible contingency (relative to a path $\mathbf{P} = \langle X, V_1, \ldots, V_m, Y \rangle$ in a model $\mathbf{M} = \langle \mathbf{V}, \mathbf{E} \rangle$) just in case, if X were to take its actual value

---

[7] For further discussion of model 'appropriateness', see Halpern and Hitchcock (2010).

$X = x$, and Z_1, …, and Z_n were to take the values $Z_1 = z_1^*$, …, and $Z_n = z_n^*$ (so that the antecedent of (†) is realized), then all remaining *on*-path variables (namely, V_1, …, V_m, and Y) would take their actual values. Intuitively this means that setting $Z_1 = z_1^*$, …, and $Z_n = z_n^*$ doesn't interfere with the causal process from X to Y (see Halpern and Pearl 2005, p. 854). Call this the 'non-interference' account of permissibility. What is important for present purposes is that (assuming the automatic truth of counterfactuals with true antecedents and true consequents[8]) *the non-interference account agrees with the actuality account in implying that the setting of off-path variables to their actual values always constitutes a permissible contingency.*[9]

The third prominent account of permissibility, developed by Halpern (2008, pp. 203-205) and Halpern and Hitchcock (2010, pp. 400-403),[10] adds to the non-interference requirement the further requirement that a possible world (presumably the closest) at which $X = x'$, and $Z_1 = z_1^*$, …, and $Z_n = z_n^*$ (so that the antecedent of (‡) is realized) is *at least as normal as the actual world.* Call this the 'normality' account of permissibility. To this point, advocates of the normality account have yet to furnish a precise normalcy metric over possible worlds. For this reason, I regard it as less satisfactory than the others. Still, for present purposes, the details of any such metric won't matter.

3. The Interventionist Semantics

Woodward (2003, p. 145) claims that the sort of counterfactuals relevant to causal claims should

---

[8] This assumption is implicit in the standard rule – described three paragraphs ago – for evaluating counterfactuals with respect to a model. It is a straightforward consequence of the Lewisian semantics to be described in §4, below. Woodward (2003, p. 83-84) also seemingly takes it as true in developing his own CMAAC.

[9] Variants on this second account of permissibility are described by Woodward (2003, pp. 83-84) and by Halpern and Pearl (2005, p. 853). These variants also imply that the actual contingency is always permissible (see Woodward 2003, p. 83; and Halpern and Pearl 2005, p. 855).

[10] For related approaches, see Menzies (2004a, 2004b) and Hitchcock (2007).

be evaluated with respect to possible worlds in which their antecedents are realized by *interventions*.[11] We have now seen that, according to advocates of SCMAACs, including Woodward, the counterfactuals relevant to causal claims are ENFs of the form (†) and (‡).[12]

On Woodward's account, an intervention on a variable V_1 is characterized *with respect to a second variable*, V_2. So Woodward's proposal concerning the evaluation of ENFs like (†) and (‡), stated more precisely, is that the relevant worlds to consider are those in which each of X, Z_1, …, and Z_n is set to the value specified in the antecedent of the counterfactual by an intervention upon it *with respect to Y*.

Woodward's definition of an intervention is as follows:

"(**IN**) I's assuming some value $I = i$, is an intervention on V_1 with respect to V_2 if and only if I is an intervention variable for V_1 with respect to V_2 and $I = i$ is an actual cause of the value taken by V_1." (Woodward 2003, p. 98; notation modified)

Definition **IN** of an intervention makes essential appeal to the notion of an *intervention variable*. Woodward provides four conditions (which he labels I1-I4) intended as individually necessary and jointly sufficient for a variable I's counting as an intervention variable for V_1 with respect to V_2. He makes use of only one of these conditions – I3 – in his critique of Lewis.

"I3.   Any directed path from I to V_2 goes through V_1. [...]" (Woodward 2003, p. 98; notation modified).

---

[11] Note that Woodward is indeed happy to express his interventionist semantics in terms of possible worlds (see ibid., e.g. pp. 139-145).

[12] Simple-antecedent counterfactuals of the form ($) $X = x \,\square\!\!\rightarrow Y = y$ and ($$) $X = x' \,\square\!\!\rightarrow Y = y'$ can be taken as limiting cases of ENFs of form (†) and (‡). The truth of such a pair of counterfactuals indicates counterfactual dependence of $Y = y$ upon $X = x$ under the null contingency relative to the direct path $\langle X, Y \rangle$ in a model constructed from the variable set $\{X, Y\}$, in which there are no off-path variables to hold fixed.

Woodward's interventionist semantics introduces circularity into his SCMAAC. In fact there are two circularities, one fairly obvious, the other more subtle. The first, more obvious, circularity is that Woodward's definition **IN** of an intervention appeals to the notion of *actual causation*. Since Woodward (2003, pp. 77-84) adopts a SCMAAC, and gives the ENF counterfactuals appealed to by that analysis an interventionist semantics (e.g. ibid. p. 84), defining interventions in terms of actual causation means that actual causation is ultimately analyzed, at least in part, in terms of actual causation. While Woodward insists that this circularity is not vicious (ibid. pp. 104-105), he readily admits the resulting analysis is not fully reductive (ibid. p. 106).

The second, more subtle, circularity arises from the fact that interventions are defined (indirectly) in terms of the notion of a directed path between variables in a model. Specifically, clause I3 of Woodward's definition of an intervention variable (a concept appealed to in his definition **IN** of an intervention) makes reference to the notion of a directed path. But the notion of a directed path in a model is defined in terms of the associated structural equations, which encode counterfactuals, which Woodward gives an interventionist semantics (ibid. pp. 42-44). So Woodward defines the notion of an intervention in terms of that of a directed path, which is a causal modeling notion that he (ultimately) defines in terms of interventionist counterfactuals. While this circularity does not in itself prevent a reductive analysis of actual causation in terms of facts about counterfactual dependence, it does ensure that the interventionist semantics does not deliver a reduction of the counterfactual dependencies in terms of which actual causation is analyzed to more basic facts.

Given that the interventionist semantics involves us in these circularities and consequently is not an apt foundation for a fully reductive CAAC (whether a SCMAAC or otherwise), one might wonder why Woodward adopts it. This might seem especially puzzling given the existence of a well-known non-causal semantics developed by Lewis (1979).

4. The Lewisian Semantics

Lewis's semantics can be explicated as follows (1979, p. 462).[13] Suppose that $X = x$ is the actual value of X and $Y = y$ is the actual value of Y, and consider a counterfactual of the form (*) $X = x^*$ $\square\!\!\rightarrow Y = y^*$ (it is left open whether $x^* = x$ and $y^* = y$). Suppose, moreover, that $X = x$ represents an event that occurs in the actual world (hereafter '@') through an interval of time, $t_{X=x}$. Then Lewis's idea is that, in evaluating (*), we should consider those worlds w that exactly match @ in matters of particular fact up to some time $t_{X=x}^-$ shortly before $t_{X=x}$, which evolve according to the actual laws at all times after $t_{X=x}$ and which, during the transition period from $t_{X=x}^-$ until the end of $t_{X=x}$, differ from @ just enough to permit $X = x^*$ to obtain. The counterfactual (*) is true just in case $Y = y^*$ holds at all such worlds.[14]

If @ is deterministic, the divergence of the relevant worlds w from @ during the transition period from $t_{X=x}^-$ until the end of $t_{X=x}$ will be miraculous (relative to the laws of @), given that the worlds w exactly match @ prior to $t_{X=x}^-$.

Woodward (2003, pp. 127-133) observes that, in making this appeal to possible worlds involving miracles (that is, violations of the laws of @), the Lewisian semantics *does not differ* from the interventionist semantics. The interventionist semantics requires that the consequent of a counterfactual be true in those worlds where its antecedent is made true by an intervention. Such hypothetical interventions obviously needn't actually occur (where the antecedent of the counterfactual in question is false, they typically don't). If @ is deterministic, then a possible (but

---

[13] The semantics for counterfactuals that I describe here corresponds to Lewis's (1979, p. 462) 'Analysis 1' of counterfactuals. Lewis attempts to subsume this analysis under a general 'closest-worlds' semantics ('Analysis 2') that he regards as fully adequate to the ordinary language counterfactual (Lewis 1979, p. 465). For the reason noted at the end of §1 (the first of the two 'preliminary points'), this latter project needn't concern us here.

[14] If $x^* = x$, so that (*) expresses a true-antecedent counterfactual, *no* divergence from @ during the transition period starting at $t_{X=x}^-$ is needed for $X = x^*$ to obtain, and so (assuming determinism) (*) is true just in case $y^* = y$.

non-actual) intervention must either be miraculous (relative to the laws of @) or result from the worlds in question having different initial conditions to @. But even a slight difference in the initial conditions of two worlds may result in widespread effects throughout time and space. So, for a pair of variables, X and Y, bringing about a difference in the value of X by means of a difference in initial conditions may well result in influence on Y over and above that brought about by the change in X.[15] On the other hand, bringing about a difference in the value of X by means of a miracle occurring shortly before hand would seem to ensure that any disturbance to factors independent of X ramifies as little as possible, minimizing the opportunity for Y to be independently affected. It would thus seem that condition I3 of Woodward's definition of an intervention variable is more likely to be satisfied where the change in X is brought about by means of such a miracle. As Woodward observes, the interventionist semantics "*leads unavoidably*" to the consideration of "*what would happen under conditions that may involve violations of physical law*" (ibid. p. 132; italics original).

Indeed, Woodward is insistent upon this point, which will be important in what follows. He says:

"One of the features of Lewis's theory that has been most resisted by commentators is

---

[15] As a referee pointed out to me, Loewer (2007, esp. pp. 300-302) observes that a difference in the initial microstates of two worlds may be reflected in a difference in macrostate only much later in time. If this is correct then – where the possible values of X represent macroscopic events or states of affairs – it may be possible to set X to some target value $X = x^*$ via a difference in initial microconditions that does not manifest itself macroscopically until at (or shortly before) the time that X comes to take the value $X = x^*$. Hence – where Y is also a macrovariable – this difference in initial conditions *may* not independently influence Y. It is, however, at best unclear whether implementing counterfactual antecedents in this way will always be possible (especially without an independent influence on Y). In the absence of a proof that it is, I shall follow Woodward and Lewis in supposing that miracles are needed to implement counterfactual antecedents.

his insistence that in many cases, the right counterfactuals to consider in explicating causal claims will be those whose antecedents require the occurrence of a 'miracle,' a violation of some law. Although my version of a manipulability theory differs in various respects from Lewis's theory … it agrees with his on this important point." (2003, p. 133)

One way of thinking about a hypothetical intervention, then, is as a miracle that meets certain constraints: in particular, those captured by **IN** and I3 (see Woodward 2003, pp. 132-133, 144-145). As already noted, those constraints involve problematic (from the reductionist's perspective) appeals to the notions of (a) actual causation, and (b) a directed path in a causal model. If we strip away those constraints, then we are left with something along the lines of the Lewisian semantics. The central issue is therefore whether the constraints are necessary. Woodward's counterexample to the Lewisian semantics (to be considered in the next section) is advanced in a bid to show that I3 at least *is* necessary.

While Lewis focuses on simple-antecedent counterfactuals (presumably because his own CAAC appeals exclusively to these), a proposal that seems Lewisian in spirit is to evaluate complex-antecedent counterfactuals of form (**) $X_1 = x_1^{\ *} \wedge \ldots \wedge X_n = x_n^{\ *} \ \Box\!\!\rightarrow Y_1 = y_1^{\ *}$ with respect to those worlds w^* which have the same initial conditions as @ and in which each variable mentioned in the antecedent is set to the specified value by means of the smallest necessary miracle occurring during a transition period starting shortly beforehand.[16] Otherwise, the worlds

---

[16] More precisely, suppose there is a pair of times, $\langle t_{X_i=x_i^{**}}^{=}, t_{X_i=x_i^{**}}^{-} \rangle$, and an interval of time, $t_{X_i=x_i^{**}}$, such that $t_{X_i=x_i^{**}}^{=}$ is shortly before $t_{X_i=x_i^{**}}^{-}$, which is shortly before the beginning of $t_{X_i=x_i^{**}}$, and which is also such that the actual laws together with the total state of w^* at $t_{X_i=x_i^{**}}^{=}$ entail that, throughout $t_{X_i=x_i^{**}}$, the event represented by $X_i = x_i^{**}$ ($x_i^{**} \neq x_i^{*}$) will occur. Then w^* contains a miracle during the period starting at $t_{X_i=x_i^{**}}^{-}$ and ending at the end of $t_{X_i=x_i^{**}}$ which is just large enough to ensure that X_i instead takes value $X_i = x_i^{*}$ (as required to implement the

w* conform to the laws of @. A counterfactual of the form (**) comes out true just in case its consequent is true in all such worlds *w**. The possible need to appeal to more than one miracle to realize the antecedent of counterfactuals like (**) is in analogy to the interventionist semantics, on which more than one intervention may be required to realize the antecedent of such a counterfactual.

The Lewisian semantics just outlined does not make any use of causal or counterfactual notions. In specifying how the relevant counterfactuals are to be evaluated, appeal is made to possible worlds, to times, and to facts about laws and matters of particular fact. None of these notions is (obviously) to be analyzed in terms of causation or counterfactuals. Adopting the Lewisian semantics therefore seems compatible with maintaining the potential reductivity of CAACs, including SCMAACs.

In order to motivate his preference for the interventionist semantics – which is *in*compatible with maintaining potential reductivity – Woodward advances a counterexample to CAACs founded upon the Lewisian semantics. In the next section, I discuss this example and argue that it does not tell against the combination of the Lewisian semantics with sophisticated, SCMAACs.

5. Woodward's Counterexample to the Lewisian Semantics

Woodward uses the following example to motivate his rejection of the Lewisian semantics in favor of his own interventionist semantics:

> "You are driving on an unfamiliar freeway in the left-hand lane when, unexpectedly, the exit you need to take appears on the right. You are unable to get over in time to exit and as a result are late for your appointment. There are only two lanes, left and right. Driving in the left-hand lane (rather than the right) caused you to be late." (Woodward 2003, p. 142)

Let binary variable *L* take value 1 if you're in the left-hand lane at the relevant time, 0 if you're in

---

antecedent of (**)).

the right; and let T take value 1 if you're late, 0 if you're not. And consider the simple-antecedent counterfactual conditional (*CC*1).

(*CC*1) $L = 0 \;\square\!\!\rightarrow\; T = 0$.

According to the Lewisian semantics, we should evaluate (*CC*1) with respect to those worlds that match @ until the beginning of a transition period beginning shortly before you arrive at the exit, with the smallest necessary miracle occurring during this transition period to ensure that you are in the right-hand lane at the relevant time ($L = 0$). The relevant worlds evolve lawfully thereafter. Perhaps the needed miracle involves a few extra neurons firing in your brain a quarter of a mile prior to the exit, with the result that you indicate, check your mirrors, and steer into the right-hand lane. But Woodward worries that such a miracle may have effects that interfere with your exiting. It may, for example, be that in some worlds in which such a miracle occurs:

> "I [sic] move into the right-hand lane [...], hit a pothole that damages my car, and continue in the right lane until just before the exit, at which point my car stalls from the damage, and I am unable to exit." (Ibid. p. 144)

If this is correct, then (*CC*1) comes out false on the Lewisian semantics.

One might be tempted to locate the problem in the fact that the Lewisian miracle (the extra neurons firing in your brain) is allowed to occur during a transition period beginning shortly before the event that it is intended to secure (namely, your being in the right lane when you arrive at the exit). After all, your hitting a pothole in the right-hand lane is an event that occurs during this transition period. Perhaps we could avoid the problem by modifying the Lewisian semantics to avoid transition periods (cp. Lewis 1979, p. 463). We might require that the miracle implementing the antecedent of (*CC*1) occur *when* you arrive at the exit, rather than in an interval starting *shortly before* you arrive at the exit. That is, we might require that it be (what we might call) a 'late miracle', rather than an 'early miracle'.

But Woodward (2003, p. 142) argues that such a modification of the Lewisian semantics will

not help. The occurrence of a late miracle implementing the antecedent of (*CC*1) means that "your car dematerializes and reappears instantaneously in the right-hand lane just before the exit" (ibid.). The trouble is "it is not at all unlikely that the very occurrence of this miracle will produce effects that will interfere with your exiting" (ibid.). For example, Woodward suggests, it may result in your car colliding with other cars. So whether we go for the 'early miracles' or the 'late miracles' version of the Lewisian semantics, (*CC*1) may well come out false.[17] Since, nevertheless, your being in the left-hand lane ($L = 1$) is intuitively a cause of your being late ($T = 1$), the worry is that founding a CAAC upon the Lewisian semantics will result in *under-generation*.

The interventionist semantics avoids this problem because it is a consequence of the definition of an intervention that an intervention realizing the antecedent of (*CC*1) satisfy condition I3: it must not be such that a variable representing it lies on a path to T that bypasses L.[18] As Woodward (2003, pp. 144-145) says:

---

[17] Lewis (1979, p. 463) himself prefers early miracles on the grounds that late miracles make for "abrupt discontinuities" (ibid.). Perhaps it was in part because of worries about the effects of such discontinuities that he rejected late miracles. Still, if Woodward is correct, the early miracles version of the Lewisian semantics doesn't fare any better.

[18] To avoid introducing an undesirable model-relativity into the interventionist semantics, one might demand that I3 be fulfilled with respect to *all* appropriate models. Alternatively, one might require that it be fulfilled with respect to any model that has a variable set that is 'sufficiently rich' in the sense defined by Hitchcock (2001b, p. 394n): namely (where $I_{L,T}$ is a putative intervention variable for L with respect to T), "[a] variable set would be sufficiently rich if the addition of new variables would not create any new directed paths between variables $[I_{L,T}]$ and $[T]$, but only interpolate variables along existing paths."

For discussion of whether the interventionist semantics renders the truth-values of counterfactuals model-relative and, if so, how this relativity may be overcome, see Strevens (2007, 2008) and Woodward (2008).

"Recall that one feature an intervention on V_1 for the purpose of determining whether V_1 causes V_2 should possess is that the intervention should not directly affect V_2 [requirement I3]: the intervention should change V_1, and it should change V_2 if at all only through V_1. [...] This requirement implies that the correct possible world to consider in evaluating [(CC1)] is [...] the world in which there is no transition period and in which all of the independent effects that the intervention that places my car in the right lane would otherwise have on my exiting are cancelled or removed. This is so despite the fact that [such a world] involves many miracles [...]. In other words, in contrast to Lewis, the interventionist account tells us that we should avoid transition periods entirely (because they may introduce factors that affect the effect independently of the putative cause), even if the cost of doing so is the postulation of interventions whose occurrence requires many diverse miracles." (Notation modified)

Woodward claims that the interventionist account tells us to *avoid transition periods entirely*. This suggests that he conceives of an intervention on L with respect to T that sets $L = 0$ (let us denote this intervention $I_{L,T} = i_{L=0}$) as consisting of a 'late miracle' that ensures that you are in the right-hand lane when you arrive at the exit. After all, an early miracle or a difference in initial conditions would result in a difference during a transition period prior to your arrival at the exit. Perhaps this late miracle consists in a sudden shift of your car from the left-hand lane to the right-hand lane.

But, *crucially*, an intervention on L with respect to T that sets $L = 0$ might not consist of such a late miracle *alone*. Rather, it might have to be a *complex* miracle (that is, a complex event consisting of simpler events, each of which violates the laws of @ – cp. Lewis 1986, p. 56). Specifically, as Woodward observes in the passage just quoted, the intervention $I_{L,T} = i_{L=0}$ will have to consist *not only* of a late miracle placing you in the right-hand lane when you arrive at the exit, *but also* of further miracles to 'cancel or remove' any independent effects that this first miracle would otherwise have on your exiting.

The idea is as follows. Let M be a binary variable whose value $M = 1$ represents the

occurrence of a late miracle that sets $L = 0$. That is, $M = 1$ represents the (miraculous) sudden shift of your car from the left-hand lane to the right-hand lane when you arrive at the exit. Perhaps (as Woodward suggests) this event would have led to a crash that would interfere with your exiting. Let C be a binary variable that takes value 1 if such a crash occurs, 0 if it doesn't. Then the variable M might not only lie on a path $M \rightarrow L \rightarrow T$, but also on a path $M \rightarrow C \rightarrow T$, reflecting the fact that the (miraculous) sudden shift of your car may influence whether you're late not just by influencing whether you're in the right lane, but also by influencing whether or not there's a crash. If so, the variable M violates Woodward's condition I3 and does not count as an *intervention variable* for L with respect to T. The late miracle $M = 1$ therefore does not (when taken alone) constitute an *intervention* on L with respect to T.

Woodward's idea (expressed in the above quotation) is that a genuine intervention must be a more complex event than that represented by $M = 1$ (namely, your sudden lane-shifting). If the latter would cause a collision – $C = 1$ – that would interfere with your exiting, then a genuine intervention $I_{L,T} = i_{L=0}$ must, in addition to your miraculous lane-shifting, also comprise a second miracle which ensures that such a collision fails to occur – so that C remains at its actual value $C = 0$ – in spite of your sudden lane-shifting. That is to say, a genuine intervention $I_{L,T} = i_{L=0}$ must be a complex miracle, which has both your miraculous lane-shifting *and* a further, crash-preventing, miracle as constituents.

The fact that $I_{L,T} = i_{L=0}$ has this crash-preventing miracle as a constituent means that the variable $I_{L,T}$ (unlike the variable M) *does not* lie on a path to C which then proceeds to T (and which bypasses L, in violation of requirement I3). The reason is that $C = 0$ *whether or not* the intervention occurs.[19]

Perhaps the miracle needed to prevent a crash (despite your suddenly changing lanes) would

---

[19] It is important here that the non-occurrence of the intervention should be construed as involving the non-occurrence of *all* of its constituents and *not*, for example, merely the non-occurrence of the crash-preventing miracle. But this just reflects a requirement that is quite generally essential to the success of CAACs (see Lewis 2004, p. 90).

itself have had some further effect that would have prevented your exit. For example, perhaps it would have startled a deer on the embankment, which would then have run into the slip road impeding your exit. Let us represent the occurrence of this event as the binary variable D_1's taking value $D_1 = 1$. Then the relevant intervention $I_{L,T} = i_{L=0}$ must, in addition to a first miracle setting $L = 0$ and a second miracle ensuring that (in spite of the first miracle) $C = 0$, also comprise a third miracle to ensure that (in spite of the second miracle) D_1 remains at its actual value, $D_1 = 0$ (that is, that the deer doesn't impede your exit). The intervention must comprise this third miracle if it is to be the case that $D_1 = 0$ *whether or not* the intervention occurs. This is necessary to ensure that the variable $I_{L,T}$ representing the intervention does not lie on a path to D_1 (a path which then proceeds to T and which bypasses L, in violation of I3).

But suppose that this deer-inhibiting miracle would itself induce some factor that would independently influence T. Let us represent this effect of the deer-inhibiting miracle by a variable D_2's taking the non-actual value $D_2 = d_2'$ (the value must be non-actual, otherwise the factor is there anyway, and is not one that is introduced by the deer-inhibiting miracle). Then the intervention $I_{L,T} = i_{L=0}$ must comprise a fourth miracle to ensure that (in spite of the deer-inhibiting miracle), D_2 remains at its actual value $D_2 = d_2$. The intervention must comprise this fourth miracle if it is to be the case that $D_2 = d_2$ *whether or not* the intervention occurs. This is necessary to ensure that the variable $I_{L,T}$ representing the intervention does not lie on a path to D_2 (a path which then proceeds to T and which bypasses L, in violation of I3).

And so on.[20]

An intervention $I_{L,T} = i_{L=0}$ thus, as Woodward observes, must comprise a complex of miracles, consisting of a (late) miracle that sets $L = 0$ and also further miracles that are such as to 'cancel or remove' any factors that are (i) introduced by some miracle in the combination and (ii)

---

[20] For simplicity, I'm supposing that the first miracle (and each subsequently introduced miracle) would have at most one effect that threatens to interfere with your exiting. Relaxing this assumption simply means allowing that the intervention $I_{L,T} = i_{L=0}$ may have to comprise a still more complex miracle.

threaten to have an independent influence upon T. We can suppose the factors that end up getting 'cancelled or removed' to be represented by the variables C, D_1, D_2, …, and D_n taking the non-actual values $C = 1$, $D_1 = 1$, $D_2 = d_2'$, …, and $D_n = d_n'$. The need to cancel or remove these factors means that the intervention $I_{L,T} = i_{L=0}$, in addition to a miracle that sets $L = 0$, comprises further miracles that serve to hold fixed the variables C, D_1, D_2, …, and D_n at their actual values $C = 0$, $D_1 = 0$, $D_2 = d_2$, …, and $D_n = d_n$ (representing the non-occurrence of the crash, the non-occurrence of the deer running into the slip road, etc.). It is for this reason that Woodward (ibid.) says that the interventionist semantics requires "the postulation of interventions whose occurrence requires many diverse miracles" and that it involves considering worlds that involve "many miracles". In a world in which such a complex of miracles occurs, you exit without inhibition and are not late. Consequently, (*CC*1) comes out true on the interventionist semantics.

So (*CC*1) comes out true on the interventionist semantics, but may well be false on the Lewisian semantics. One consequence of this is that a causal model constructed from a variable set including L and T might differ depending on whether the counterfactuals it encodes are those that come out true on the interventionist semantics, or those that come out true on the Lewisian semantics. Define a *Lewisian model* $\mathbf{M^L} = \langle \mathbf{V}, \mathbf{E^L} \rangle$ to be a causal model such that the structural equations in $\mathbf{E^L}$ encode precisely those counterfactual dependencies that obtain among the values of variables in $\mathbf{V}$ when all counterfactuals are given a Lewisian semantics. By contrast, define an *interventionist model* $\mathbf{M^I} = \langle \mathbf{V}, \mathbf{E^I} \rangle$ to be a model such that the structural equations in $\mathbf{E^I}$ encode precisely those counterfactual dependencies that obtain among the values of variables in $\mathbf{V}$ when all counterfactuals are given an interventionist semantics.

If Woodward is right that (*CC*1) comes out true on the interventionist semantics, but false on the Lewisian semantics, then the interventionist model constructed from the variable set $\{L, T\}$ is $\mathbf{D^I} = \langle \{L, T\}, \{L = 1, T = L\} \rangle$, while the Lewisian model is $\mathbf{D^L} = \langle \{L, T\}, \{L = 1, T = 1\} \rangle$. In the interventionist model $\mathbf{D^I}$ there is a directed path $\langle L, T \rangle$, and $T = 1$ counterfactually depends upon $L = 1$ along this path under the null contingency. By contrast, in the Lewisian model, there is no directed path $\langle L, T \rangle$, and hence no contingent dependence of $T = 1$ upon $L =$

1.

But note that this does not directly tell against the ability of the Lewisian semantics to serve as a foundation for an extensionally adequate CAAC.[21] In particular, according to SCMAACs, the truth of (*CC1*) isn't necessary for $L = 1$ to be a cause of $T = 1$. What *is* necessary is that, in *some* appropriate causal model (which may have a variable set that is richer than $\{L, T\}$), $T = 1$ counterfactually depends upon $L = 1$ under a permissible contingency.

But the existence of counterfactual dependence of $T = 1$ upon $L = 1$ in an appropriate model under a permissible contingency, when all counterfactuals are evaluated using the Lewisian semantics, follows straightforwardly from the truth of (*CC1*) when evaluated using the interventionist semantics.

To illustrate this, suppose that we adopt the 'late miracles' variant of the Lewisian semantics. Then, although there may be no path $\langle L, T \rangle$ in the Lewisian model constructed from the variable set $\{L, T\}$, there is a path $\langle L, T \rangle$ in the Lewisian model constructed from the variable set $\{L, T, C, D_1, D_2, \ldots, D_n\}$. That this is so can be seen by considering the counterfactual (*CC2*).

(*CC2*) $L = 0 \wedge C = 0 \wedge D_1 = 0 \wedge D_2 = d_2 \wedge \ldots \wedge D_n = d_n \: \square\!\!\rightarrow T = 0$.

Evaluating (*CC2*) using the 'late miracles' variant of the Lewisian semantics means considering a world in which a late miracle occurs to set $L = 0$, and further miracles occur to set $C, D_1, D_2, \ldots,$ and D_n equal to their (actual) values $C = 0, D_1 = 0, D_2 = d_2, \ldots,$ and $D_n = d_n$. That is, in evaluating (*CC2*) according to the Lewisian semantics, we are led to consider *precisely the same world* as we consider in evaluating (*CC1*) according to the interventionist semantics. Since (*CC1*) comes out true on the interventionist semantics, (*CC2*) comes out true on the Lewisian semantics.

In virtue of having a true antecedent and consequent, (*CC3*) automatically comes out true

---

[21] The falsity of (*CC1*) on the Lewisian semantics means only that it delivers the wrong results about causation when combined with the most naïve of CAACs (which take the truth of (*CC1*) to be *necessary* for $L = 1$ to be a cause of $T = 1$).

on the Lewisian semantics.

(CC3) $L = 1 \land C = 0 \land D_1 = 0 \land D_2 = d_2 \land \ldots \land D_n = d_n \,\square\!\!\rightarrow T = 1$.

It follows directly from the truth of (CC2) and (CC3) on the Lewisian semantics that there exists the path $\langle L, T \rangle$ in the Lewisian model constructed from the variable set $\{L, C, D_1, D_2, \ldots, D_n, T\}$. In such a model, $\{C, D_1, D_2, \ldots, D_n\}$ are the variables off the path $\langle L, T \rangle$. So, provided that the assignment $C = 0$, $D_1 = 0$, $D_2 = d_2$, $\ldots$, $D_n = d_n$ constitutes a permissible contingency, SCMAACs (when combined with the Lewisian semantics) yield the result that $L = 1$ is a cause of $T = 1$.

The *actuality* and *non-interference* accounts of permissibility count $C = 0$, $D_1 = 0$, $D_2 = d_2$, $\ldots$, $D_n = d_n$ as a permissible contingency because these are the actual values of the off-path variables $C, D_1, D_2, \ldots,$ and D_n. Finally, on the *normality* account, $C = 0$, $D_1 = 0$, $D_2 = d_2$, $\ldots$, $D_n = d_n$ is permissible just in case a world in which the antecedent of (CC2) holds is at least as normal as @. But, as already noted, in evaluating (CC2) on the Lewisian semantics, we are led to consider *precisely the same world* as we are led to consider in evaluating (CC1) on the interventionist semantics. Consequently, if the null contingency is to count as permissible by the interventionist's lights – so that counterfactual dependence of $T = 1$ upon $L = 1$ under the null contingency in the interventionist model $\mathbf{D^I} = \langle \{L, T\}, \{L = 1, T = L\}\rangle$ (revealed by the truth of (CC1) on the interventionist semantics) is sufficient for $L = 1$ to count as a cause of $T = 1$ – then the contingency $C = 0$, $D_1 = 0$, $D_2 = d_2$, $\ldots$, $D_n = d_n$ must be permissible by the Lewisian's lights.

All three accounts of permissibility thus count the contingency $C = 0$, $D_1 = 0$, $D_2 = d_2$, $\ldots$, $D_n = d_n$ as permissible. But we have already seen that $T = 1$ counterfactually depends upon $L = 1$ under this contingency in the Lewisian model constructed from the variable set $\{L, C, D_1, D_2, \ldots, D_n, T\}$. Consequently the combination of a SCMAAC with the Lewisian semantics correctly counts $L = 1$ as a cause of $T = 1$.[22]

---

[22] It has just been shown that there is counterfactual dependence of $T = 1$ upon $L = 1$ under a permissible contingency in the Lewisian model constructed from the variable set $\{L, C, D_1, D_2,$

…, D_n, T}. Since SCMAACs existentially quantify over models – $L = 1$ is a cause of $T = 1$ just in case there is counterfactual dependence of $T = 1$ upon $L = 1$ under a permissible contingency in *at least one* (appropriate) model – it follows that the combination of a SCMAAC with the Lewisian semantics delivers the verdict that $L = 1$ is a cause of $T = 1$.

It might nevertheless be objected that the strategy that I have employed for defending the Lewisian semantics is somehow parasitic upon the interventionist semantics. The objector might say something along the following lines:

> "While we can hit upon a set of variables – namely {L, C, D_1, D_2, …, D_n, T} – such that there is contingent dependence of $T = 1$ upon $L = 1$ in the Lewisian model constructed from that variable set, the Lewisian semantics doesn't tell us what the relevant variables to include in our variable set are (and hence what to hold fixed in looking for a relation of contingent dependence of $L = 1$ upon $T = 1$). The interventionist semantics, by contrast, tells us what the relevant factors to suppress are: namely, those by which a miracle setting $L = 0$ threatens to independently influence the value of T."

Such an objector is looking for is a criterion for identifying the variables that must be included in our variable set (if there is to be contingent counterfactual dependence in the Lewisian model constructed from it) which does not simply involve adopting a causal, interventionist-inspired rule along the lines of "include (and hold fixed by miracles) variables representing those factors by introducing which a miracle setting $L = 0$ threatens to independently influence T".

One plausible criterion for identifying the relevant variables without falling back upon such a rule appeals to the technical notion of a 'sufficiently rich' variable set (see fn. 18 above). No Lewisian model with a variable set that doesn't include at least the variables L, C, D_1, D_2, …, D_n, and T would count as sufficiently rich (with respect to the variables L and T). The reason is that only in a model that includes all of these variables does the direct path $\langle L, T \rangle$ appear (because only by holding fixed the variables C, D_1, D_2, …, D_n at their actual values can we recover a

The point generalizes. Suppose that $X = x$ and $Y = y$, and some pair of counterfactuals of the form (†) and (‡) comes out true on the interventionist semantics, thus indicating the existence of counterfactual dependence of $Y = y$ upon $X = x$ relative to the direct path $\langle X, Y \rangle$ under the contingency $Z_1 = z_1^*$, …, and $Z_n = z_n^*$ in an interventionist model constructed from the variable set $\{X, Z_1, …, Z_n, Y\}$. Evaluating (†) and (‡) using the interventionist semantics requires considering a world in which a set of interventions (with respect to Y) occur that set $Z_1 = z_1^*$, …, and $Z_n = z_n^*$. These interventions may be denoted $I_{Z_1,Y} = i_{Z_1=z_1^*}$, …, and $I_{Z_n,Y} = i_{Z_n=z_n^*}$ respectively. The interventions in question must also be such as to suppress those factors $U_1 = u_1'$, …, and $U_m = u_m'$ that threaten (i) to be introduced by a constituent miracle of one of the interventions; and (ii) to independently influence Y. This can be achieved by their including constituent miracles that serve to hold fixed U_1, …, and U_m at their actual values $U_1 = u_1$, …, and $U_m = u_m$.

In addition to $I_{Z_1,Y} = i_{Z_1=z_1^*}$, …, and $I_{Z_n,Y} = i_{Z_n=z_n^*}$, one further intervention apiece is required to realize the antecedents of (†) and (‡). An intervention $I_{X,Y} = i_{X=x}$ setting $X = x$ is required to realize the antecedent of (†), while an intervention $I_{X,Y} = i_{X=x'}$ instead setting $X = x'$ is required to realize the antecedent of (‡). Suppose that, in addition to a miracle setting $X = x$, $I_{X,Y} = i_{X=x}$ comprises further miracles to suppress (by holding fixed the relevant variables at their actual values) the factors $T_1 = t_1'$, …, and $T_k = t_k'$ which certain of its constituent miracles threaten to independently influence Y by introducing. And suppose an intervention $I_{X,Y} = i_{X=x'}$ setting $X = x'$ comprises further miracles to suppress (again, by holding fixed the relevant variables at their actual values) the factors $S_1 = s_1'$, …, and $S_j = s_j'$, which certain of *its* constituent miracles threaten to independently influence Y by introducing.

Then, if counterfactuals (†) and (‡) come out true on the *interventionist* semantics,

---

counterfactual dependence of $T = 1$ upon $L = 1$). The Lewisian can thus say that $L = 1$ is a cause of $T = 1$ iff there is contingent counterfactual dependence of $T = 1$ upon $L = 1$ in a Lewisian model constructed from a variable set that is 'sufficiently rich'.

counterfactuals (!) and (!!) come out true on the *Lewisian* semantics:

(!) $X = x \wedge Z_1 = z_1^* \wedge \ldots \wedge Z_n = z_n^* \wedge U_1 = u_1 \wedge \ldots \wedge U_m = u_m \wedge T_1 = t_1 \wedge \ldots \wedge T_k = t_k \wedge S_1 = s_1 \wedge \ldots \wedge S_j = s_j \mathbin{\square\!\!\rightarrow} Y = y$;

(!!) $X = x' \wedge Z_1 = z_1^* \wedge \ldots \wedge Z_n = z_n^* \wedge U_1 = u_1 \wedge \ldots \wedge U_m = u_m \wedge T_1 = t_1 \wedge \ldots \wedge T_k = t_k \wedge S_1 = s_1 \wedge \ldots \wedge S_j = s_j \mathbin{\square\!\!\rightarrow} Y = y'$.

The reason is that the interventionist semantics has us consider *exactly the same worlds* in evaluating (†) and (‡) as the Lewisian semantics has us consider in evaluating (!) and (!!). First, the interventionist semantics has us consider the same worlds in evaluating (†) as the Lewisian semantics has us consider in evaluating (!): namely those in which such a combination of miracles occurs as, not only to set $X = x$, $Z_1 = z_1^*$, …, and $Z_n = z_n^*$, but also to set $U_1 = u_1$, …, $U_m = u_m$, $T_1 = t_1$, …, and $T_k = t_k$, thus suppressing the independent influence on Y of any of the miracles introduced (no *further* miracles are required to set $S_1 = s_1$, …, and $S_j = s_j$ since these are the factors that threaten to be disturbed by a miracle setting $X = x'$, rather than one setting $X = x$).

Second, the interventionist semantics has us consider the same worlds in evaluating (‡) as the Lewisian semantics has us consider in evaluating (!!): namely those in which such a combination of miracles occurs as, not only to set $X = x'$, $Z_1 = z_1^*$, …, and $Z_n = z_n^*$, but also to set $U_1 = u_1$, …, $U_m = u_m$, $S_1 = s_1$, …, and $S_j = s_j$, thus suppressing the independent influence on Y of any of the miracles introduced (no *further* miracles are required to set $T_1 = t_1$, …, and $T_k = t_k$ since these are the factors that threaten to be disturbed by a miracle setting $X = x$, rather than one setting $X = x'$).

The truth of (!) and (!!) implies the existence of a path $\langle X, Y \rangle$ in the Lewisian model constructed from the variable set $\{X, Z_1, \ldots, Z_n, U_1, \ldots, U_m, S_1, \ldots, S_j, T_1, \ldots, T_k, Y\}$.[23] Provided

---

[23] More impoverished variable sets, for which the associated Lewisian model does not contain the path $\langle X, Y \rangle$, are not 'sufficiently rich' with respect to X and Y.

that the contingency $Z_1 = z_1^*, \ldots, Z_n = z_n^*, U_1 = u_1, \ldots, U_m = u_m, S_1 = s_1, \ldots, S_j = s_j, T_1 = t_1, \ldots, T_k$ $= t_k$, is permissible, we therefore get the result that $X = x$ is a cause of $Y = y$ when we combine SCMAACs with the Lewisian semantics.

Does the permissibility of the contingency $Z_1 = z_1^*, \ldots, Z_n = z_n^*, U_1 = u_1, \ldots, U_m = u_m, S_1 = s_1, \ldots, S_j = s_j, T_1 = t_1, \ldots, T_k = t_k$, follow from that of $Z_1 = z_1^*, \ldots, Z_n = z_n^*$ (the latter must be permissible if a SCMAAC founded upon the interventionist semantics is to imply that the truth of (†) and (‡) indicates actual causation between $X = x$ and $Y = y$)? It does on the *actuality* account of permissibility. If $Z_1 = z_1^*, \ldots, Z_n = z_n^*$ is permissible on this account, these must be the actual values of $Z_1, \ldots,$ and Z_n. But $U_1 = u_1, \ldots, U_m = u_m, S_1 = s_1, \ldots, S_j = s_j, T_1 = t_1, \ldots,$ and $T_k = t_k$ are also the actual values of $U_1 \ldots, U_m, S_1, \ldots, S_j, T_1, \ldots,$ and T_k.

It also counts as permissible on the *non-interference* account. On that account it is required that, when the off-path variables take the values specified by the contingency and X takes its actual value $X = x$, the remaining on-path variables (in this case just Y) take their actual values (in this case, $Y = y$). That this is so is indicated by the truth of (!). And the truth of (!) on the Lewisian semantics follows (as shown above) from the truth of (†) on the interventionist semantics.

Finally, on the *normality* criterion of permissibility, what matters is whether a world in which the antecedent of (!!) is realized is at least as normal as @. But since the interventionist evaluates (‡) with respect to exactly the same world as the Lewisian evaluates (!!) with respect to, the contingency the Lewisian considers (in evaluating (!!)) must count as permissible (on the normality account) if the contingency the interventionist considers (in evaluating (‡)) is.

So far I have been concerned to show that, if there is counterfactual dependence of $Y = y$ upon $X = x$ under a permissible contingency in an interventionist model, then there is also counterfactual dependence of $Y = y$ upon $X = x$ under a permissible contingency in a Lewisian model. This shows that combining SCMAACs with a Lewisian semantics does not result in *under-generation* relative either to intuition or to the combination of SCMAACs with the interventionist semantics. In particular, Woodward's counterexample to the Lewisian semantics, which was designed to establish such under-generation (by showing that the former combination doesn't count $L = 1$ as a cause of $T = 1$), fails.

But might employing the Lewisian semantics result in *over-generation* of causal relations relative to intuition and the interventionist semantics? Woodward does not advance any examples designed to show this, but perhaps we can describe such a case.

6. Does the Lewisian Semantics Lead to Over-Generation?

Add to Woodward's example (considered in the previous section) the supposition that a team of scientists is monitoring your progress as you drive down the road. The scientists will tick a box labeled 'N' if everything is normal, or one labeled 'A' if they observe some anomaly.

Let N be a variable that takes value $N = 1$ if the scientists tick box 'N' (which they do in @), and $N = 0$ otherwise. Evaluated with respect to the Lewisian semantics, the counterfactual (*CC*4) seemingly comes out true.

(*CC*4) $L = 0 \:\square\!\!\rightarrow N = 0$.

The Lewisian semantics requires a miracle to implement the antecedent of (*CC*4). In a world in which such a miracle occurs, the scientists (who, let's suppose, are monitoring the situation very carefully and in minute detail) notice this miracle and so tick box A instead of box N. The counterfactual consequently comes out true.

But if (*CC*4) comes out true, then there is counterfactual dependence of $N = 1$ on $L = 1$ under the null contingency relative to the direct path $\langle L, N \rangle$ in the Lewisian model $\langle \{L, N\}, \{L = 1, N = L\} \rangle$. When combined with the Lewisian semantics, SCMAACs will thus yield the result that your being in the left-hand lane was a cause of box N being ticked.

Is this the correct result? I think the Lewisian can make a plausible case for an affirmative answer. She can argue that, in the circumstances, which include your having been in the left-hand lane at all times previously, your *not being* in the left-hand lane when you arrived at the exit would be anomalous, and would cause the scientists to tick box A rather than box N. In the circumstances, then, your *being* in the left-hand lane is a cause of the scientists' ticking box N (as is every other lawful event that the scientists observe whose non-occurrence, in the circumstances,

would have been anomalous).

But note that the general success of this sort of argument depends upon the adoption of the 'late miracles' version of the Lewisian semantics. Suppose the Lewisian instead adopts the 'early miracles' variant of that semantics, taking the relevant miracle to be some event such as the firing of extra neurons in your brain a quarter of a mile prior to the exit. This version of the Lewisian semantics overgenerates. To see this, suppose that the scientists, as part of their monitoring efforts, are examining (by means of an MRI hooked up to a transmitter) what is happening in your brain as you travel down the freeway. Suppose that they are very quick at analyzing and recording their data: if there is a miraculous event before you reach the exit, then they will record this by ticking box A before you reach the exit.[24] Then, on the 'early miracles' version of the Lewisian semantics, it will be true that if you hadn't been in the left-hand lane when you reached the exit, then the scientists would have *earlier* ticked box A. It follows that your being in the left-hand lane when you reached the exit prevented them from doing so: a case of spurious backwards-in-time causation.[25]

Such examples can be avoided by adopting a 'late miracles' version of the Lewisian semantics. Indeed, since *any* temporal gap (however small) between the state of affairs represented by $L = 0$ and the miracle that secures this state of affairs threatens over-generation (just take the ticking of box A in response to the miracle to occur sufficiently quickly), it seems that the Lewisian should require that the 'late miracle' in question be the *latest possible*: an instantaneous shift *when* you arrive at the exit.

Such a miracle is plausibly regarded as not being distinct from the state of affairs that it secures (namely that represented by $L = 0$). The miracle *just is* your being in the right-hand lane

---

[24] Where $t_{L=1}$ is the time at which you reach the exit, then for any $t_{L=1}^-$ ($< t_{L=1}$), if the miracle occurs at $t_{L=1}^-$, then those clever scientists will manage to tick box A by $t_{L=1}^- + (t_{L=1} - t_{L=1}^-)/2$.

[25] I follow most philosophers in the causal modeling (and wider counterfactual) tradition in assuming that absences (such as the scientists' not ticking box A) can act as (causes and) effects. It would be straightforward to modify the example so that it involved only positive events.

when you reach the exit, despite the fact that all the earlier determinants of $L = 1$ (your being in the left-hand lane when you reach the exit) remain intact.[26]

There is good reason to think that, on this version of the Lewisian semantics, over-generation is quite generally avoided. Late miracles are not distinct from the events they 'bring about'. A late miracle simply *is* the occurrence of a hypothetical event, in spite of the fact that all antecedent determinants of some incompatible (actual) event remain intact. Any effects that the late miracle would have are plausibly regarded as effects that the hypothetical event would have. It consequently seems reasonable to regard the incompatible actual event as causing whatever is caused by the non-occurrence of the miracle. For example, in the circumstances in which you were in the left-hand lane at all times prior to arriving at the exit, your (miraculously) not being there upon arrival would have caused the scientists *not* to tick box N. The scientists' *ticking* box N is therefore plausibly an effect of your being in the left-hand lane at the relevant time. In general, over-generation would result only if late miracles had effects whose absences were not plausibly caused by the events that those late miracles eliminate.

This raises a question for the interventionist. We saw that, in the passage quoted in the previous section, Woodward says that the interventionist semantics tells us to avoid transition periods altogether. This suggests that an intervention on X with respect to Y that sets $X = x'$ will comprise a late miracle, together with further miracles to 'cancel or remove' any independent effects that the intervention would otherwise have on Y. But it has just been suggested that if we really *avoid transition periods altogether* (even vanishingly small ones), by having the miracle occur no earlier than $X = x'$, then the miracle is no longer distinct from the event – $X = x'$ – that it secures. But if it is not distinct, then (I) it cannot be an *actual cause* of $X = x'$ (as the definition **IN** of an intervention demands); and (II) it seems that it cannot have any effects upon Y that are *independent* of those that it has via setting $X = x'$. So if an intervention on X with respect to Y that sets $X =$

---

[26] Taking the relevant miracle to be non-distinct from $L = 0$ makes no difference to the Lewisian's response (discussed in the previous section) to the threat of *under*-generation, which goes through just as before.

x′ avoids transition periods altogether by comprising the latest possible miracle setting $X = x′$, then (trivially) it will not have to comprise any further miracles, since there can be no effects upon Y of this first miracle that are independent of its influence on $X = x′$ (*a fortiori* there are no independent effects to cancel or remove).

It would therefore be more natural to construe Woodward, not as claiming that we should really avoid transition periods *altogether*, but rather as claiming that we should avoid *gratuitous* transition periods. In other words, an intervention on X with respect to Y that sets $X = x′$ should not comprise the latest possible miracle. That is, it should not be construed as comprising a miracle that simply *is* the fact that $X = x′$. Rather, it should be construed as comprising a *late-ish* miracle which, while not occurring too long before the event represented by $X = x′$, nevertheless remains distinct from that event. As such it is capable of acting as an *actual cause* of $X = x′$ as well as potentially having an independent influence on Y that the further constituent miracles of the intervention must suppress.

So what does the interventionist semantics imply about the counterfactual (*CC4*)? We are to suppose the antecedent $L = 0$ to be realized by a late-ish miracle (perhaps an instantaneous shift of your car *very shortly before* you arrive at the exit, rather than *when* you arrive at the exit). This miracle is distinct from the event represented by $L = 0$ and consequently is able to exert an influence on whether or not box N is ticked independently of its influence on whether or not you're in the left-hand lane at the relevant time. The interventionist semantics tells us that, in evaluating (*CC4*), the relevant worlds to consider are those in which this influence is suppressed (miraculously the scientists fail to observe the anomaly, or fail to record it by ticking box A instead of N). If this is correct, then the counterfactual (*CC4*) will come out *false* on the interventionist semantics. Consequently, there is *no* counterfactual dependence of $N = 1$ on $L = 1$ under the null contingency relative to the direct path $\langle L, N \rangle$ in the interventionist model constructed from the variable set $\{L, N\}$. Unless there is some richer variable set relative to which there *is* contingent dependence of $N = 1$ on $L = 1$ (and it is at least unobvious that there is), SCMAACs, when combined with the interventionist semantics, will yield the result that your being in the left-hand lane was *not* a cause of box N being ticked. This result is the opposite of the

result derived from combining SCMAACs with the Lewisian semantics, on its late miracles variant. This shows that the two semantics do not provide equivalent foundations for SCMAACs.

Which result is intuitively correct? And which semantics is therefore to be favored? As I've argued it seems to me that your being in the left-hand lane when you arrive at the exit *is* a (partial) cause of the scientists' ticking box N. Still, I don't think it's necessary to set too much store by this intuition. Perhaps the fact that the intuition is elicited by emphasizing the circumstances that obtained (that in fact you were in the left-hand lane at all times previously) suggests that the intuition about causation is sensitive to how we evaluate the counterfactual: whether, as the late miracles variant of the Lewisian semantics requires, we hold fixed the fact that you were in the left-hand lane at all times previously; or whether, as the interventionist semantics requires, we don't. If so the intuition perhaps reflects the existence of two reasonable ways to evaluate the counterfactuals relevant to causation (though, as I have been arguing, we must insist upon one of these ways – the Lewisian way – if we are to effect a reduction of causation to non-causal facts).

7. Conclusion

Woodward influentially combines a SCMAAC with an interventionist semantics for the counterfactuals encoded in causal models. Woodward's formulation of the interventionist semantics appeals to (a) actual causation; and (b) the notion of a directed path in a causal model. In virtue of its appeal to these notions, the adoption of this semantics is incompatible with the project of reductively analyzing actual causation in terms of counterfactual dependence, where counterfactual dependence is itself reductively analyzed in other (non-causal and non-counterfactual) terms. Yet Woodward argues that adopting the interventionist semantics is necessary in order to overcome counterexamples that afflict CAACs when they are combined with the rival, Lewisian semantics.

I have argued that this is not so. The combination of SCMAACs (which are independently attractive) with the Lewisian semantics is perfectly defensible. The putative counterexample Woodward advances to show that CAACs under-generate when founded upon the Lewisian semantics is not telling when the CAACs in question are SCMAACs. Indeed, I have argued that

whenever the interventionist semantics implies that $Y = y$ counterfactually depends upon $X = x$ under a permissible contingency in an appropriate causal model (a condition that SCMAACs take to be necessary and sufficient for $X = x$ to be an actual cause of $Y = y$), the Lewisian semantics implies this too.

The Lewisian semantics, on its 'early miracles' variant, is susceptible to the converse charge: namely that it results in over-generation. But this needn't force us to accept the interventionist semantics and to abandon the project of reduction. By going for the 'late miracles' variant of the Lewisian semantics, over-generation is avoided.

My strategy has *not* been to argue that the combination of SCMAACs with the interventionist semantics yields an *extensionally* in*adequate* theory of causation. My aim, rather, has been to defend the *extensional adequacy* of the combination of SCMAACs with a Lewisian semantics. Indeed, in §5, I argued that the latter combination yields equivalent results to the former in a certain class of cases. And while, in §6, I argued that a suitable variant of the Lewisian semantics actually appears (in combination with SCMAACs) to yield superior results to the interventionist semantics in another class of cases, I acknowledged that either result might be regarded as acceptable.

My conclusion, then, is that since plausible SCMAACs perform just as well when combined with the Lewisian semantics (on its 'late miracles' variant) as when combined with the interventionist semantics, and since the Lewisian semantics does not make reference to causal or counterfactual notions, a reductive counterfactual analysis of actual causation remains a reasonable hope.

Acknowledgements

McDonnell Foundation Causal Learning Collaborative.

References

Collins, J., Hall, N., & Paul, L. A. (Eds.) (2004). *Causation and Counterfactuals.* (Cambridge, MA: MIT Press)

Glymour, C. & Wimberly, F. (2007). Actual Causes and Thought Experiments. (In J. K. Campbell, M. O'Rourke, & H. Silverstein (Eds.), *Causation and Explanation* (pp. 43-67). Cambridge, MA: MIT Press.)

Halpern, J. Y. (2008). Defaults and Normality in Causal Structures. (In G. Brewka & J. Lang (Eds.), *Proceedings of the Eleventh International Conference on Principles of Knowledge Representation and Reasoning* (pp. 198-208). Menlo Park, CA: AAAI Press.)

Halpern, J. Y. & Hitchcock, C. (2010). Actual Causation and the Art of Modeling. (In R. Dechter, H. Geffner, & J. Y. Halpern (Eds.), *Heuristics, Probability and Causality: A Tribute to Judea Pearl* (pp. 383-406). London: College Publications.)

Halpern, J. Y. & Pearl, J. (2005). Causes and Explanations: A Structural-Model Approach, Part I: Causes. *British Journal for the Philosophy of Science*, 56, 843-887

Hitchcock, C. (2001a). The Intransitivity of Causation Revealed in Equations and Graphs. *Journal of Philosophy*, 98, 273-299

Hitchcock, C. (2001b). A Tale of Two Effects. *Philosophical Review*, 110, 361-396.

Hitchcock, C. (2007). Prevention, Preemption, and the Principle of Sufficient Reason. *Philosophical Review*, 116, 495-532

Lewis, D. (1979). Counterfactual Dependence and Time's Arrow. *Nous*, 13, 455-476

Lewis, D. (1986). Postscripts to "Counterfactual Dependence and Time's Arrow". (In D. Lewis, *Philosophical Papers*, Vol. II (pp. 52-66). New York: Oxford University Press.)

Lewis, D. (2004). Causation as Influence. (In Collins *et al.* 2004, pp. 75-106).

Loewer, B. (2007). Counterfactuals and the Second Law. (In H. Price and R. Corry (Eds.), *Causation, Physics and the Constitution of Reality: Russell's Republic Revisited* (pp. 293-326). Oxford: Oxford University Press.)

Menzies, P. (2004a). Causal Models, Token Causation, and Processes. *Philosophy of Science*, 71, 820-832.

Menzies, P. (2004b). Difference-making in Context. In Collins *et al.* (2004), 139-180.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference.* 2nd Ed. (Cambridge: Cambridge University Press)

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search.* 2nd Edition. (Cambridge, MA: MIT Press)

Strevens, M. (2007). Review of Woodward, *Making Things Happen. Philosophy and Phenomenological Research*, 74, 233-249

Strevens, M. (2008). Comments on Woodward, *Making Things Happen. Philosophy and Phenomenological Research*, 77, 171-192

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

Woodward, J. (2008). Response to Strevens. *Philosophy and Phenomenological Research*, 77, 193-212