A Theory of Structural Determination

J. Dmitri Gallow [*,†]

Abstract

While structural equations modeling is increasingly used in philosophical theorizing about causation, it remains unclear what it takes for a particular structural equations model to be correct. To the extent that this issue has been addressed, the consensus appears to be that it takes a certain family of causal counterfactuals being true. I argue that this account faces difficulties in securing the independent manipulability of the structural determination relations represented in a correct structural equations model. I then offer an alternate understanding of structural determination, and I demonstrate that this theory guarantees that structural determination relations are independently manipulable. The account provides a straightforward way of understanding hypothetical interventions, as well as a criterion for distinguishing hypothetical changes in the values of variables which constitute interventions from those which do not. It additionally affords a semantics for causal counterfactual conditionals which is able to yield a clean solution to a problem case for the standard 'closest possible world' semantics.

1 Introduction

As a rough approximation, regularity theories of causation hold that, given the circumstances, causes are nomically *sufficient* for their effects. As a matter of law, if the cause is present in these circumstances, then the effect will be present too. As a rough approximation, counterfactual theories of causation hold that, given the circumstances, causes are nomically *necessary* for their effects. As a matter of law, were the cause to have been absent in these circumstances, so too would the effect have been absent. I, like many, have been persuaded that counterfactual theories are roughly correct about the relation of singular, token, or actual causation. Here, I will argue that nomic sufficiency accounts are roughly correct about what I will call *structural*

* Final Draft; forthcoming in *Philosophical Studies*
✉: jdmitrig@nyu.edu
† Thanks to Gordon Belot, Allan Gibbard, Jim Joyce, Brian Weatherson, and an anonymous referee for helpful feedback on earlier versions of this material.

determination relations. On the theory advanced here, these structural determination relations provide truth conditions for causal counterfactual conditionals. For counterfactual theorists, this means that regularity accounts still have an important role to play in the metaphysics of causation.

Singular causal relations—the relations expressed by sentences of the form "*c*'s *F*-ing caused *e* to *G*", where *c*'s *F*-ing and *e*'s *G*-ing are particular events or facts—are familiar philosophical fare. Structural determination, less so. As I will explain in further depth below, structural determination relations link qualities or quantities of particular parts of the world. We can represent these qualities or quantities with variable values. When we do so, structural determination relations are representable as structural *equations*, which establish functional relationships between those variables.

Structural equations modeling has been used to provide novel semantics for causal counterfactual conditionals (Hiddleston, 2005, Shulz, 2011, and Briggs, 2012), to investigate traditional metaphysical questions about singular causation (Hitchcock, 2001, Woodward, 2003, ch. 2, Menzies, 2004, Halpern & Pearl, 2001, 2005a, Menzies, 2007, Handfield et al., 2008, Halpern, 2008, Hitchcock & Knobe, 2009, Glymour et al., 2010, Halpern & Hitchcock, 2010, forthcoming, Paul & Hall, 2013, Livengood, 2013, Baumgartner, 2013, and Weslake, forthcoming), to explicate the nature of causal enquiry and scientific explanation (Woodward, 1999, Halpern & Pearl, 2005b, Woodward & Hitchcock, 2003a,b, and Woodward, 2003, ch. 7), and to undergird novel statistical techniques for drawing inferences about the causal structure of the world on the basis of sample data (Pearl, 2000, 2009 and Spirtes et al., 2000). However, relatively little has been done to get clear about what exactly someone commits themselves to when they endorse one of these models—what exactly, that is, a structural equations model *says* about the world.[1]

To the extent that it has been addressed, the consensus view appears to be that structural equations represent patterns of causal counterfactual dependence amongst variable values (in particular, see Hitchcock, 2001, Woodward & Hitchcock, 2003a, Hall, 2007, Hitchcock, 2007, Menzies, 2008, Halpern & Hitchcock, 2010, and Glynn, 2013).[2] In §3 below, I will explain why causal counterfactuals are not well-suited to provide a semantics for structural determination relations. My contention there will be that the counterfactual account is not capable of securing the independent manipulability of the structural determination relations represented in a structural equations model—a property of structural equations models known as *modularity*. In contrast, I will suggest that one variable is structurally determined by others

[1] Notable exceptions include Handfield et al. (2008), Baumgartner (2013), and Glynn (2013).

[2] Not just any counterfactual is a *causal* counterfactual. While the question of which counterfactuals are causal is a question to be decided by theory rather than stipulation, at the least, causal counterfactuals must be *non-backtracking* (see Lewis, 1979) and they must relate *distinct* events (see Kim, 1973 and Lewis, 1986).

just in case, within a certain region of modal space, the values of the latter variables are sufficient for the value of the former (§4). This account will allow us to understand the technical notion of an *intervention*; it will guarantee the modularity of a correct structural equations model; and it will allow us to provide a semantics for causal counterfactual conditionals which neatly solves a problem case for the standard 'closest possible world' semantics of Stalnaker (1968) and Lewis (1973).

2 Structural Equations Models

A structural equations model $\mathcal{M}$ is a triple $(\mathcal{U}, \mathcal{V}, \mathcal{E})$ of a vector of exogenous variable $\mathcal{U} = (U_1 \ldots U_M)$, a vector of endogenous variables $\mathcal{V} = (V_1 \ldots V_N)$ and a vector of structural equations $\mathcal{E} = (\phi_{V_1} \ldots \phi_{V_N})$, one for each endogenous variable.[3] Formally, a variable is a partial function from a set of possibilities (or 'worlds') $\mathcal{W}$ to the real line $\mathbb{R}$. What makes the function *partial* is just that it needn't map each and every possibility $w \in \mathcal{W}$ to some real number. So, for instance, I might be interested in the variable S = *number appearing on the digital scale at t*. This variable assigns a value, s, to every world at which the scale displays a number at t. However, it will not assign any value to a world at which the scale does not display a number at t, or does not exist at t. Here's another (equivalent) way to understand a variable: it is an assignment of values to a set of pairwise inconsistent propositions $\{\mathrm{P}_i\} \subset \wp(\mathcal{W})$. Which value the variable takes on depends upon which of these propositions is true. For instance, the variable F could assign the value f to the proposition *you exert a force of f Newtons on the surface of the scale at t*, for every f in some specified range. In general, variables stand to their values as determinables stand to their determinants; just as being red is one way for an object to be colored, having the property represented by $V = v$ is one way for a part of the world to have the property represented by V. (A word on notation: I'll write '$V_w = v$' to mean that the value of w, under the function V, is v and I will often use '$V = v$' to denote the set of worlds w such that $V_w = v$.)

The structural equations in $\mathcal{E}$ establish functional relationships amongst the variables in $\mathcal{U} \cup \mathcal{V}$. For instance, suppose that the digital scale is accurate at t, zeroed out in the appropriate way (so that it reads '0' when subjected to the earth's gravitational force and the ambient air pressure), and nothing else (besides you) is exerting any force

[3] A word on notation: throughout, I'll be using uppercase letters $(A, B, C, \ldots, Z)$ to represent variables, and the corresponding lowercase letters $(a, b, c, \ldots, z)$ to stand for the values of those variables. Functions will be denoted with 'ϕ', with subscripts added to indicate which variable the function is associated with. Vectors of variables or variable values will be represented with boldface of calligraphic letters $(\mathcal{U}, \mathcal{V}, \mathbf{X}, \mathbf{x}, \mathbf{PA}(V)$, *etc.*). At times I will use the function name alone to denote the entire structural equation—for instance, I will write 'ϕ_Y' to denote the structural equation '$Y := \phi_Y(X_1, \ldots, X_N)$'. Propositions will be denoted with upright uppercase letters $(\mathrm{A}, \mathrm{B}, \mathrm{C}, \ldots, \mathrm{Z})$. I'll also be abusing set-theoretic notation, $\in, \cup, \subseteq, -$, and so on, by applying it to *vectors* of variables (or variable values).

FIGURE 1: A causal graph

upon the surface of the scale. Then, the value of S will be *determined by* the value of F. Since an object weighing 0.2248 pounds exerts 1 Newton on the surface of the Earth, if $F = f$, then $S = \lceil 0.2248 \cdot f \rceil$.[4] We can thus write down the *structural equation*

$$S := \lceil 0.2248 \cdot F \rceil$$

What makes this equation *structural* is that it is asymmetric; it matters which variable is to the left of the ':='. In addition to claiming that the value of S is a function of the value of F, the structural equation makes the further claim that that the value of S is *determined by* the value of F in a way that the value of F isn't determined by the value of S. Here's a way of getting at this asymmetry: if there were a certain kind of intervention on the value of F—if, for instance, you were to put one foot on the floor—then the value of S would have been different—the scale would have displayed a different number. If, however, there were an intervention on the value of S—if, for instance, the scale was re-zeroed to read '0'—then the value of F would *not* be different—you would not suddenly exert 0 Newtons on the surface of the scale. (To emphasize this asymmetry, I use ':=' to distinguish that asymmetrical relation from the symmetrical '='.)[5]

These variables and this structural equation together constitute a *structural equations model*, or a *causal model* (I'll use these terms interchangeably throughout). A *causal graph* provides an intuitive representation of much of the information contained in a causal model. If U appears on the right-hand-side of V's structural equation ϕ_V, then there will be an arrow with its tail at U and its head at V in the causal graph. For instance, the model of the scale's display and the force you exert on the scale's surface generates the causal graph shown in figure 1. This causal graph tells us that the value of F determines the value of S, without telling us exactly *how*. It tells us that the numbers on the scale's display are determined by the force you exert upon it, but it doesn't, for instance, tell us what units the scale's display is set to. For that information, we must look to S's structural equation in $\mathcal{E}$.

A causal model can involve many more variables and structural equations than this. Also, a single structural equation can relate more than two variables. Adapting an

[4] '$\lceil x \rceil$' is the function which rounds x up to the closest integer.

[5] It's worth noting that the functions ϕ_V must be *non-constant*. A constant function from one variable to another does not represent any kind of *determination of* of the latter variable by the former. We should also require that the domain of each structural equation include the entire image of their parent variables' structural equations, and *only* that image.

$$\mathcal{E}_2 = \left(\begin{array}{l} L := C \\ R := C \\ D := L \vee R \end{array} \right)$$

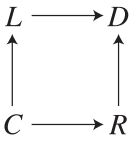

FIGURE 2: The system of structural equations $\mathcal{E}_2$.

example from PEARL (2000, ch. 7), suppose that there are two riflemen, one standing on the left, the other standing on the right, who have their rifles aimed at a deserter. If the captain gives the order, then both riflemen will fire, and the deserter will die. We can model the causal structure of this case with $\mathcal{M}_2 = ((C), (L, R, D), \mathcal{E}_2)$, where C is a binary variable which takes the value 1 if the captain gives the order to fire and takes the value 0 otherwise, L is a binary variable which takes the value 1 iff the left rifleman fires, R is a binary variable which takes the value 1 iff the right rifleman fires, and D is a binary variable which takes the value 1 iff the deserter dies. The structural equations in $\mathcal{E}_2$ are shown in figure 2. (There, $x \vee y$ is the truth function $\max\{x, y\}$.) These equations tell us that the left rifleman will fire iff the captain gives the order, and likewise for the right rifleman. And the deserter will die iff at least one of the riflemen fire. $\mathcal{M}_2$ tells us that the value of C determines the values of L and R and that the values of L and R jointly determine the value of D.

We may use the language of genealogy to talk about the structural relationships between variables. Thus, all of the variables in a causal model which appear on the right-hand-side of V's structural equation, ϕ_V, are called V's *structural parents*. I'll use '$\mathbf{PA}(V)$' to refer to a vector of V's structural parents.[6] (If U is exogenous, then $\mathbf{PA}(U)$ is the empty vector.) In the model shown in figure 2, *e.g.*, $\mathbf{PA}(D) = (L, R)$. In a similar fashion, we can define V's *structural descendants*—with the slight wrinkle that we stipulate that every variable V is one of its own descendants. I'll use '$\mathbf{DE}(V)$' to refer to a vector of V's structural descendants. In the model shown in figure 2, $\mathbf{DE}(L) = (L, D)$.

A few paragraphs back, I invoked the notion of an *intervention*. Formally, an intervention is a way of setting the values of some of the variables in $\mathcal{U} \cup \mathcal{V}$ without directly affecting any of the other variables in $\mathcal{U} \cup \mathcal{V}$, or their determination structure. To illustrate, suppose that in the model shown in figure 2, the value of L is set to 1 via an intervention. Suppose, that is, that we perform an intervention to make the left rifleman fire. The way this is modeled is by replacing L's structural equation, $L := C$, with $L = 1$ (indicating that L has been set to 1 via an intervention) and leaving all other structural equations unchanged. We thus get the mutilated model

[6] In general, there will be many such vectors. It won't matter for my purposes which one '$\mathbf{PA}(V)$' refers to. Pick one. Likewise for the other vectors of variables or variable values I discuss here.

$$\mathcal{E}_{2,\,L=1} = \left(\begin{array}{l} L = 1 \\ R := C \\ D := L \vee R \end{array} \right)$$

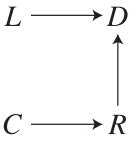

FIGURE 3: The system of equations $\mathcal{E}_{2,\,L=1}$ models an intervention on the variable L.

$\mathcal{M}_{2,\,L=1} = ((C), (L, R, D), \mathcal{E}_{2,\,L=1})$, shown in figure 3. In $\mathcal{M}_{2,\,L=1}$, whether the left rifleman fires is no longer determined by whether the captain gives the order. However, whether the deserter dies is still determined by whether the left rifleman fires. In general, the graphical result of an intervention on a variable V is to remove all of the arrows leading *into* V (if such there be), to destroy all of the structural determination relations between V and $\mathbf{PA}(V)$, while leaving all other structural determination relations intact.

This property of a structural equations model—that there are in-principle hypothetical interventions upon the variables which leave all the other structural determination relations intact—is known as *modularity*. Without modularity, structural equations models do not tell us anything about the results of hypothetical interventions, since without the assumption that the structural equations other than ϕ_V remain in place post-intervention, we cannot calculate the down-stream effects of setting the value of V.

Notice that not every way of setting the value of V will have this result. Some ways of setting the value of V will affect other variables in the graph as well. For instance, one way of setting S to 0, one way of making the scale read '0', is to simply lift you off of the scale. But this wouldn't count as an *intervention* on the value of S, since it wouldn't alter the manner in which the value of F determines the value of S. It wouldn't be correct to model this way of setting S to 0 by replacing $S := \phi_S(F)$ with $S = 0$, since the determination relation represented by $S := \phi_S(F)$ would still be in force. It would be this very determination relation that we would exploit in order to affect the value of S. Additionally, we could set the value of F in such a way that we affect the manner in which the value of F determines the value of S—*i.e.*, our meddling could have the result of *changing* the structural equation ϕ_S. For instance, we might decide to keep you from stepping onto the scale by placing a dead five-pound rat on the scale. In that case, our method for setting the value of F would alter S's structural equation, replacing it with

$$S := \lceil 0.2248 \cdot F + 5 \rceil$$

So only certain methods of setting the value of a variable in a causal model will count

as *interventions* on the value of that variable, in our technical sense.[7]

Once we have this method for modeling interventions, a method for evaluating causal counterfactual conditionals comes along for free. On this account, the counterfactual A $\boxrightarrow$ C is true at a world w according to the model $\mathcal{M}$ just in case $\mathcal{M}_{\mathrm{A}}, \mathcal{U}_w \models \mathrm{C}$. That is: the counterfactual A $\boxrightarrow$ C is true at a world w, according to the model $\mathcal{M}$, given the exogenous variable assignment $\mathcal{U} = \mathcal{U}_w$, iff C is true in the model that we get by minimally mutilating $\mathcal{M}$ so as to make A true. To illustrate: suppose that, in the causal model shown in figure 2, the actual value of C is 0. Suppose, that is, the the captain doesn't actually give the order to fire.[8] Then, neither the left nor the right rifleman fires, and the deserter does not die. And suppose that we want to evaluate the causal counterfactual 'If the left rifleman were to have fired, then the deserter would have died'—or '$L = 1 \boxrightarrow D = 1$'. To evaluate this causal counterfactual, we simply perform an intervention on the value of L so as to make the antecedent true; we mutilate the model, so that the value of L is no longer determined by the value of C, we set L to 1, and then we calculate the values of R and D in the mutilated model in accordance with their structural equations and the values for the exogenous variables. If the consequent comes out true in the mutilated model $\mathcal{M}_{2,\,L=1}$, then the counterfactual was true in the original model $\mathcal{M}_2$. According to $\mathcal{M}_2$, then, '$L = 1 \boxrightarrow D = 1$' is true. If the left rifleman were to fire, then the deserter would have died. Note that, without modularity, we would not be able to evaluate these counterfactual conditionals, since, without modularity, there is no guarantee that the downstream structural determination relations would remain intact post-intervention.

A theory of structural determination should explain why structural equations models have the properties they do. In particular, it should explain why they allow us to correctly evaluate causal counterfactual conditionals in this way, and it should explain why a correct system of structural equations is modular. It would be a benefit of an account if it could explain why only certain ways of setting the values of variables leave the downstream structural determination relations unaffected, as well as providing a principled way of distinguishing the ways of setting the values of variables which do from those which do not constitute interventions, in our technical sense. In §4, I will provide a theory which meets each of these explanatory demands. I will not attempt to account for probabilistic structural determination relations (of the sort that I believe are implicated in probabilistic causation). Nor will I be concerned with backwards structural determination relations, in which the future state of the world structurally determines the past state of the world. That's not because I think that there aren't,

[7] See Cartwright (2009).

[8] Here and throughout, I'm using 'actual' as an indexical like 'here', and *not* as a rigid designator for the actual world.

or couldn't be, probabilistic or backwards structural determination relations. Considering these issues here would simply muddy already murky waters. Another task for another day.

3 The Causal Counterfactual Understanding

Let's say that a structural equation $V := \phi_V(\mathbf{PA}(V))$ is *descriptively adequate* at world w just in case $V_w = \phi_V(\mathbf{PA}(V)_w)$. Just as mere descriptive adequacy is not sufficient for a universal generalization to be a law of nature, mere descriptive adequacy is not enough for a structural equation to be correct. There must additionally be some kind of genuine *determination* of V by $\mathbf{PA}(V)$. So an account of structural equations models must say something about what it takes, beyond mere descriptive adequacy, for a structural equation to be correct.

One of the more popular ways of understanding structural equations models appeals to causal counterfactual conditionals. Hitchcock articulates this view in his 2001 (p. 283–84):

> A system of structural equations is an elegant means for representing a whole family of counterfactuals...The correctness of a set of structural equations, and of the corresponding graph, depends upon the truth of these counterfactuals.

On this account, what it is for an isolated structural equation $V := \phi_V(\mathbf{PA}(V))$ to be correct is just for it to be the case that, for any subvector $\mathbf{X} \subseteq \mathbf{PA}(V)$, were $\mathbf{X}$ to take on the values $\mathbf{x}$, V would take on the value $\phi_V(\mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}})$,

$$(V) \qquad \forall \mathbf{X} \subseteq \mathbf{PA}(V) \quad \forall \mathbf{x} \ \ (\mathbf{X} = \mathbf{x} \ \ \boxdot\!\!\rightarrow \ \ V = \phi_V(\mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}))$$

where $\mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}$ is the assignment given to V's structural parents by $\mathbf{X} = \mathbf{x}$ (if $\mathbf{X} = \mathbf{x}$ doesn't assign any value to one of V's structural parents, then $\mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}$ gives that parent its actual value). (V) tells us that V's value counterfactually depends upon those of its parents, in the way specified by ϕ_V.

More generally, we can say that what it is for a structural equation ϕ_V, in a causal model $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{E})$, to be correct at a world w, is just for it to be the case that, for every subvector $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$, and any assignment of values $\mathbf{x}$ to $\mathbf{X}$, were $\mathbf{X}$ to take on those values, V would take on the value $\phi_V(\mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}})$

$$(\mathcal{V}1) \quad \forall V \in \mathcal{V} \quad \forall \mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V) \quad \forall \mathbf{x} \quad (\mathbf{X} = \mathbf{x} \ \ \boxdot\!\!\rightarrow \ \ V = \phi_V(\mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}))$$

where $\mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}$ assigns V's parents the values determined by $(\mathcal{U} - \mathbf{X})_w, \mathbf{x}$, and $\mathcal{E}$ –

$\bigcup_i(\phi_{X_i})$, for each endogenous $X_i \in \mathbf{X}$. That is, $\mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}$ assigns $\mathbf{PA}(V)$ the values determined by the mutilated model $\mathcal{M}_{\mathbf{X}=\mathbf{x}}$, with the actual assignment of values to the exogenous variables. ($\mathcal{V}1$) tells us that B's value counterfactually depends upon the values of its parents, and *only* the values of its parents, in the way specified by ϕ_V.

The causal counterfactuals in (V) and ($\mathcal{V}1$) are to be evaluated in the standard way (see Stalnaker 1968, Lewis 1973). To evaluate a counterfactual $\mathrm{A} \boxdot\!\!\rightarrow \mathrm{C}$, we consider some privileged set of possibilities determined by A and the world of evaluation w, and check to see whether C is true in those possibilities. Exactly which possibilities we ought to check is a complicated and controversial matter. However, for the most part, we can sidestep these issues here. We need only endorse the following general framework: there is a *selection function*, f, which is a function from pairs of propositions, A, and worlds, w, to sets of worlds, $f(\mathrm{A}, w)$. Whenever a counterfactual conditional $\mathrm{A} \boxdot\!\!\rightarrow \mathrm{C}$ is true at a world w, what makes it the case that $\mathrm{A} \boxdot\!\!\rightarrow \mathrm{C}$ is true is that $f(\mathrm{A}, w) \models \mathrm{C}$.[9] Different accounts of the selection function will yield different truth conditions for counterfactual conditionals. However, for my purposes, it won't matter what we say about f, so long as we agree that it satisfies the following three properties.[10]

($f1$) $\quad f(\mathrm{A}, w) \models \mathrm{A}$

($f2$) $\quad$ if $\mathrm{A} \models \mathrm{B}$ then $f(\mathrm{B}, w) \cap \mathrm{A} \subseteq f(\mathrm{A}, w)$

($f3$) $\quad$ at w, there is a hypothetical *intervention* to set any $\mathbf{V} \subseteq \mathcal{U} \cup \mathcal{V}$ to any values $\mathbf{v}$ which yields a world $w_{\mathbf{V}=\mathbf{v}} \in f(\mathbf{V} = \mathbf{v}, w)$

Returning to ($\mathcal{V}1$): that condition imposes two constraints on a structural equation, ϕ_V, in a causal model $\mathcal{M}$. In the first place, it says that the value of ϕ_V's left-hand-

[9] From the standpoint of Lewis (1973)'s account of counterfactuals, it will appear that, by adopting this general framework, I am tolerating the so-called *limit assumption*—the assumption that, for any arbitrary antecedent A and world w, there is a set of *most similar* A-worlds (see Lewis 1973, Stalnaker 1980). Appearances are deceiving. The limit assumption is not needed for any of my arguments here. In Lewis's framework, for any case in which the limit assumption fails and $\mathrm{A} \boxdot\!\!\rightarrow \mathrm{C}$ is true, we can just define $f(\mathrm{A}, w)$ to be the largest sphere centered on w containing at least one A world and throughout which the matieral conditional $\mathrm{A} \supset \mathrm{C}$ is true. So long as $\mathrm{A} \boxdot\!\!\rightarrow \mathrm{C}$ is true, there will be some such sphere. If it is false, of course, there won't be such a sphere, so this won't do as an account of the truth conditions of these counterfactuals. However, I am not interested in providing truth conditions for these counterfactuals. Rather, I am interested in the question of whether the truth of a set of counterfactuals is sufficient to guarantee the correctness of a structural equations model. And this trick will tell us what we can infer from the truth of $\mathrm{A} \boxdot\!\!\rightarrow \mathrm{C}$ at a world w, on Lewis's account.

[10] ($f1$) corresponds to Lewis's 2nd condition on the selection function; and ($f2$) is a weakening of Lewis's 4th condition. See Lewis (1973, p. 58).

$$\mathcal{E}_4 = \begin{pmatrix} L := C \\ D := L \end{pmatrix} \qquad C \longrightarrow L \longrightarrow D$$

FIGURE 4: The system of structural equations $\mathcal{E}_4$.

side variable, V, is sensitive to the values of ϕ_V's right-hand-side variables, $\mathbf{PA}(V)$, and they are sensitive in precisely the way specified by ϕ_V. Changes in the values of those variables would lead to changes in the value of V, and they would lead to precisely the changes specified by ϕ_V. Additionally, it says that the value of V is *only* directly sensitive to the values of $\mathbf{PA}(V)$. Holding fixed those values, changes in the values of the other variables in the model would not lead to changes in the value of V.

Note that requiring a causal model to satisfy ($\mathcal{V}1$) is stronger than merely requiring it to satisfy (V), for each endogenous variable $V \in \mathcal{V}$. To illustrate: suppose that the right rifleman takes the day off, so that the causal model shown in figure 4 correctly describes the structural determination relations between the captain's giving the order (C), the left rifleman's firing (L), and the deserter's dying (D). Suppose that, at the actual world, the captain doesn't give the order. Given the method for evaluating causal counterfactuals introduced in §2, this model entails (1).

(1) $$C = 1 \boxdot\!\!\rightarrow D = 1$$

However, (1) does not follow from the truth of (V), for each of the structural equations in $\mathcal{E}_4$,

$$C = 0 \boxdot\!\!\rightarrow L = 0$$
$$C = 1 \boxdot\!\!\rightarrow L = 1$$
$$L = 0 \boxdot\!\!\rightarrow D = 0$$
$$L = 1 \boxdot\!\!\rightarrow D = 1$$

since the counterfactual conditional is not transitive. ($\mathcal{V}1$), by contrast, will require, *inter alia*, that both (2) and (3) be true in order for the system fo structural equations $\mathcal{E}_4$ to be correct.

(2) $$C = 1 \boxdot\!\!\rightarrow L = 1$$

(3) $$(C = 1 \wedge L = 1) \boxdot\!\!\rightarrow D = 1$$

$$\mathcal{E}_5 = (W := S \vee B)$$

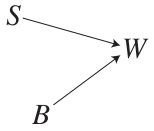

FIGURE 5: The system of structural equations $\mathcal{E}_5$.

And (2) and (3) *do* entail (1), given ($f1$) and ($f2$).

However, the truth of ($\mathcal{V}1$) is not sufficient for the correctness of a structural equations model. Take the familiar example of Suzy and Billy throwing their rocks at a window. Both Suzy and Billy have excellent aim, so if either of them throws their rock, then the window will shatter; and the window is sturdy enough that if neither of them throw their rock, then the window will not shatter. Suppose that Suzy actually throws and Billy doesn't, and that (4–11) are all true.

(4) $(B = 0 \wedge S = 0) \;\; \square\!\!\rightarrow \;\; W = 0$

(5) $(B = 0 \wedge S = 1) \;\; \square\!\!\rightarrow \;\; W = 1$

(6) $(B = 1 \wedge S = 0) \;\; \square\!\!\rightarrow \;\; W = 1$

(7) $(B = 1 \wedge S = 1) \;\; \square\!\!\rightarrow \;\; W = 1$

(8) $S = 1 \;\; \square\!\!\rightarrow \;\; W = 1$

(9) $S = 0 \;\; \square\!\!\rightarrow \;\; W = 0$

(10) $B = 1 \;\; \square\!\!\rightarrow \;\; W = 1$

(11) $B = 0 \;\; \square\!\!\rightarrow \;\; W = 1$

(where B, S, and W are binary variables with the natural interpretation). If the truth of ($\mathcal{V}1$) were sufficient for the correctness of a structural equations model, then the system of structural equations $\mathcal{E}_5$, shown in figure 5, would have to be correct.

However, $\mathcal{E}_5$ says more than the counterfactuals (4–11) do. When $B = 0$ and $S = 1$, $\mathcal{E}_5$ entails that were Billy to have thrown, Suzy (still) would have, $B = 1 \;\square\!\!\rightarrow\; S = 1$. But it is consistent with the truth of (4–11) that Suzy wouldn't have thrown her rock if Billy had thrown his. That is, it is consistent with (4–11) that the value of S is determined by the value of B. And if S is determined by B, then $\mathcal{E}_5$ would be incorrect in virtue of missing a necessary determination relation.

We can fix this problem by requiring that each exogenous variables is counterfactually independent of all the other variables in the model—*i.e.*, at a world w, for any exogenous variable $U \in \mathcal{U}$, and any assignment of values $\mathbf{x}$ to any subvector

$\mathbf{X} \subseteq (\mathcal{U} - U) \cup \mathcal{V}$, were $\mathbf{X}$ to take on those values, U would (still) take on its actual value, U_w.

$$(\mathcal{U}1) \qquad \forall U \in \mathcal{U} \quad \forall \mathbf{X} \subseteq (\mathcal{U} - U) \cup \mathcal{V} \quad \forall \mathbf{x} \quad (\mathbf{X} = \mathbf{x} \boxdot\!\!\rightarrow U = U_w)$$

This gives us the following account of the correctness of a causal model $\mathcal{M}$:

($\mathcal{M}1$) $\mathcal{M}$ is correct at w iff:
1. $\mathcal{V}$ satisfies ($\mathcal{V}1$) at w
2. $\mathcal{U}$ satisfies ($\mathcal{U}1$) at w

That is: $\mathcal{M}$ is correct iff 1) for every endogenous variable $V \in \mathcal{V}$, a) the value of V counterfactually depends upon the values of the variables in $\mathbf{PA}(V)$ in precisely the manner specified by ϕ_V, and b) holding fixed the value of $\mathbf{PA}(V)$, V *doesn't* counterfactually depend upon the values of the variables in $(\mathcal{U} \cup \mathcal{V}) - \mathbf{PA}(V)$; and 2) for every $U \in \mathcal{U}$, the value of U is counterfactually independent of the values of the other variables in the model.

3.1 Problems with Modularity

In order for a structural equations model to be correct, the equations in $\mathcal{E}$ must be *modular*—that is, that there be in-principle interventions to set the values of any subset of $\mathcal{U} \cup \mathcal{V}$ which leaves the structural equations of the non-intervened-upon endogenous variables intact. The problem is that modularity does not follow from ($\mathcal{M}1$) alone; nor can we formulate the requirement of modularity in terms of any finite number of counterfactuals.

Distinguish two kinds of modularity: weak and strong. According to weak modularity, when there is an intervention or interventions to set the values of variables in a correct causal model, the structural equations of the non-intervened-upon endogenous variables in the model will still be descriptively adequate. That is, when we perform hypothetical interventions on the values of the variables in $\mathcal{U} \cup \mathcal{V}$, taking us to the world w_i, then, for every non-intervened-upon endogenous variable V,

$$w_i \models V = \phi_V(\mathbf{PA}(V))$$

According to strong modularity, whenever there is an intervention or interventions to set the values of variables in a correct causal model, the structural equations of the non-intervened-upon endogenous variables in the model will still be *correct*. That is,

when we perform hypothetical interventions on the values of the variables in $\mathcal{U} \cup \mathcal{V}$, taking us to the world w_i, for every non-intervened-upon endogenous variable V,

$$w_i \models V := \phi_V(\mathbf{PA}(V))$$

On the counterfactual account, this means that it must at least be true that $w_i \models (\mathcal{V}1)$. That is, for every non-intervened-upon endogenous variable V,

(12) $\quad w_i \models \forall \mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V) \;\; \forall \mathbf{x} \;\; (\mathbf{X} = \mathbf{x} \;\; \square\!\!\rightarrow \;\; V = \phi_V(\mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}))$

$(\mathcal{M}1)$ does not guarantee strong modularity because it does not guarantee the truth of (12).

To see why, consider again the structural equations model shown in figure 4. For that model, $(\mathcal{M}1)$ requires that $C = 1 \;\square\!\!\rightarrow\; D = 1$: at all the closest worlds at which the captain gives the order, the deserter will die, $f(C = 1, @) \models D = 1$.[11] However, $(\mathcal{M}1)$ does not require that $f(C = 1, @) \models L = 0 \;\square\!\!\rightarrow\; D = 0$. It does not require that, at all the closest worlds at which the captain gives the order, whether the deserter dies counterfactually depends upon whether the left rifleman fires. And that means that $(\mathcal{M}1)$ does not require that, at the worlds at which a hypothetical intervention setting C to 1 occurs, D's value is still structurally determined by L's value. So $(\mathcal{M}1)$ will not guarantee that the equations in $\mathcal{E}$ are strongly modular.

Weak modularity is not modularity enough. Structural equations don't merely represent accidentally true patterns amongst variable values. They represent *determination* relations between variable values. To say that the structural equation $V := \phi(\mathbf{PA}(V))$ is unaffected by an intervention on another variable should be to say that the *determination* of V by $\mathbf{PA}(V)$ is unaffected.

Here's a thought about how to achieve strong modularity: we don't merely require that $(\mathcal{V}1)$ be satisfied. We additionally require that $\mathcal{V}$ meets the following condition.

$(\mathcal{V}2)$ $\quad \forall \mathbf{X} \subset \mathcal{U} \cup \mathcal{V} \quad \forall \mathbf{x} \;\; (\mathbf{X} = \mathbf{x} \;\square\!\!\rightarrow\; (\forall W \in \mathcal{V} - \mathbf{X}$

$\forall \mathbf{Y} \subseteq \mathcal{U} \cup (\mathcal{V} - W) \quad \forall \mathbf{y} \;\; (\mathbf{Y} = \mathbf{y} \;\square\!\!\rightarrow\; W = \phi_W(\mathbf{PA}(W)_{\mathbf{X}=\mathbf{x},\mathbf{Y}=\mathbf{y}}))))$

$(\mathcal{V}2)$ says that, if there were an intervention to set the values of any set of variables $\mathbf{X}$, then $(\mathcal{V}1)$ would still hold for all the non-intervened-upon variables. This solution is not satisfying. With this new account, we are told that what it is for a causal model to be correct is, *inter alia*, for *both* $(\mathcal{V}1)$ *and* $(\mathcal{V}2)$ to be satisfied. But while $(\mathcal{V}2)$ guarantees

[11] Throughout, I use '@' to denote the actual world.

that, at the world where the hypothetical intervention occurs, ($\mathcal{V}1$) will hold, we have as yet no guarantee that, at that world, ($\mathcal{V}2$) will be satisfied. But if what it is for a structural equation to be correct is for both ($\mathcal{V}1$) and ($\mathcal{V}2$) to hold, then this account fails to guarantee that the structural equation will still be correct post-intervention; that is, it fails to secure strong modularity.

It's actually a bit worse than that. ($\mathcal{M}1$) cannot even guarantee *weak* modularity. For it could turn out that, for three variables V_1, V_2, and V_3,

$$f(V_1 = v_1 \wedge V_2 = v_2, @) \models V_3 = \phi_{V_3}(\mathbf{PA}(V_3)) \tag{13}$$

even though

$$f(V_2 = v_2, f(V_1 = v_1, @)) \not\models V_3 = \phi_{V_3}(\mathbf{PA}(V_3)) \tag{14}$$[12]

While ($\mathcal{M}1$) guarantees (13), it is consistent with (14). But this means that ($\mathcal{M}1$) fails to guarantee that the equations in $\mathcal{E}$ will even be *descriptively adequate* after multiple sequential interventions. And the number of potential interventions is unbounded (we can always just set the value of X to x, then set it to $x' \neq x$, then set it back to x, then back to x', and so on and so forth, indefinitely). So there is no finite number of counterfactuals that is sufficient to guarantee that the equations in $\mathcal{E}$ are even weakly modular.

4 The Nomic Sufficiency Understanding

In this section, I will suggest that we can retain all of the virtues of the causal counterfactual understanding of structural equation models, without running into the problems with modularity raised in §3.1 above, by moving to an understanding of structural equations according to which what makes them correct is that they are descriptively adequate throughout an area of modal space meeting certain constraints. For instance, an isolated structural equation $V := \phi_V(\mathbf{PA}(V))$ is correct just in case, for every world w in some set of worlds $\mathfrak{F}_V$, $V_w = \phi_V(\mathbf{PA}(V)_w)$.

A useful orienting picture here is Mackie (1965)'s notion of a *causal field*. Mackie argues that causal claims must be evaluated relative to a set of alternate states of affairs within which the causes are parts of an occurrent minimally sufficient condition for the

[12] Above, I didn't define f for *sets* of worlds. Let's say that $f(\mathrm{B}, f(\mathrm{A}, w))$ is the union of $f(\mathrm{B}, w')$ for every $w' \in f(\mathrm{A}, w)$.

effect.[13] He calls this set of alternative states of affairs the *causal field*. This is roughly how I am thinking of the set of worlds $\mathfrak{F}_V$. Just as, on Mackie's account, the causes are the parts of an occurent minimally sufficient condition for the effect within the causal field, on the nomic sufficiency account, the values of a variable V's structural parents, $\mathbf{PA}(V)$, are minimally sufficient for the value of V within $\mathfrak{F}_V$.[14] In virtue of this resemblance, I will call the set of possibilities $\mathfrak{F}_V$ V's *causal field*.

Of course, this is far too rough. For any structural equation ϕ_V, it will be easy to find *some* set of worlds within which ϕ_V is descriptively adequate. A structural equation according to which my height structurally determines the size of the earth will be descriptively adequate throughout $\mathfrak{F}_V$ if I only include worlds in $\mathfrak{F}_V$ in which the earth's diameter is a constant multiple of my height. But my height does not determine the size of the earth. The question of which possibilities to consider when evaluating the determination of one variable by another is a complicated one, but it is one that is faced by the nomic sufficiency account and the counterfactual account both. The counterfactual account solves it by appeal to some suitable selection function f. And I see no reason why the nomic sufficiency account cannot similarly avail itself of this very selection function—whichever one we fancied for the counterfactual understanding of structural equations models—to characterize the worlds which must be included in $\mathfrak{F}_V$.

As a first step, if we're considering an isolated structural equation $V := \phi_V(\mathbf{PA}(V))$ at a world w, then we can require that, for every assignment $\mathbf{x}$ to any $\mathbf{X} \subseteq \mathbf{PA}(V)$, every world in $f(\mathbf{X} = \mathbf{x}, w)$ must be included in $\mathfrak{F}_V$. Similarly, if we're considering a structural equation $V := \phi_V(\mathbf{PA}(V))$ in a causal model $\mathcal{M}$ at a world w, then we can require that, for every assignment of values $\mathbf{x}$ to any $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$, all the worlds in $f(\mathbf{X} = \mathbf{x}, w)$ must be included in $\mathfrak{F}_V$.

$$(\mathfrak{F}1) \qquad \forall \mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V) \quad \forall \mathbf{x} \quad f(\mathbf{X} = \mathbf{x}, w) \subseteq \mathfrak{F}_V$$

Putting this together with the requirement that ϕ_V be descriptively adequate throughout $\mathfrak{F}_V$, we can say that a causal model $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{E})$ is correct at a world w only if

[13] The condition is *minimally* sufficient for the effect just in case no subset of the condition is also sufficient for the effect. The minimal sufficient condition is *occurrent* iff it actually obtained on the occasion in question.

[14] The values of $\mathbf{PA}(V)$ are *sufficient*, and not (or not necessarily) necessary, for the value of V because two different assignments of values to $\mathbf{PA}(V)$ could get mapped by ϕ_V to the very same value of V. $\mathbf{PA}(V)$ must be *minimally* sufficient for V's value because we require that ϕ_V be a non-constant function of each of its parents. See fn 5.

($\mathcal{V}3$).

($\mathcal{V}3$) $\quad \forall V \in \mathcal{V} \quad \exists \mathfrak{F}_V$ such that $\mathfrak{F}_V$ satisfies ($\mathfrak{F}1$) and $\mathfrak{F}_V \models V = \phi_V(\mathbf{PA}(V))$

If we stop here, then, in the presence of ($\mathcal{U}1$), we get an account which is equivalent to the counterfactual account's ($\mathcal{V}1$). That is, given ($\mathcal{U}1$), $\mathcal{M}$ satisfies ($\mathcal{V}1$) iff $\mathcal{M}$ satisfies ($\mathcal{V}3$). (Theorem 1, proved in the appendix, establishes the equivalence.)

Since this condition on the endogenous variables is equivalent to the counterfactual account's, if we stop here, we will run into the problems with modularity that we encountered in §3.1. However, we *needn't* stop here. We can additionally require that the condition imposed by ($\mathfrak{F}1$) holds, not only for the world of evaluation, but for every *other* world in $\mathfrak{F}_V$ as well.

($\mathfrak{F}2$) $\quad \forall w \in \mathfrak{F}_V \quad \forall \mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V) \quad \forall \mathbf{x} \quad f(\mathbf{X} = \mathbf{x}, w) \subseteq \mathfrak{F}_V$

This amounts to the requirement that the set $\mathfrak{F}_V$ is closed under counterfactual suppositions about the values of any of the variables in $\mathcal{U} \cup (\mathcal{V} - V)$. At any world $w \in \mathfrak{F}_V$, making counterfactual suppositions about the values of any of the the variables in $\mathcal{U} \cup (\mathcal{V} - V)$ will deliver a set of worlds *inside* of $\mathfrak{F}_V$.

Putting this together with the requirement that ϕ_V be descriptively adequate throughout $\mathfrak{F}_V$, we get an account according to which a causal model $\mathcal{M}$ is correct at a world w only if ($\mathcal{V}4$).

($\mathcal{V}4$) $\quad \forall V \in \mathcal{V} \quad \exists \mathfrak{F}_V \ni w$ such that $\mathfrak{F}_V$ satisfies ($\mathfrak{F}2$) and $\mathfrak{F}_V \models V = \phi_V(\mathbf{PA}(V))$

A structural equation belonging to a causal model satisfying ($\mathcal{V}4$) will continue to belong to a causal model satisfying ($\mathcal{V}4$) after any number of interventions to set the values of any of the other variables in the model. In the appendix, I prove the following theorem.

Theorem 2. *Given ($f3$), if ϕ_V belongs to a causal model satisfying ($\mathcal{V}4$) at a world w_0, then ϕ_V will continue to belong to a causal model satisfying ($\mathcal{V}4$) after any number of consecutive hypothetical interventions to set the values of any $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$.*

This means that the nomic sufficiency account is not subject to the objection I raised for the counterfactual account in the previous section—*viz.*, that it could not guarantee that a structural equation ϕ_V would continue to be correct after multiple sequential interventions to set the values of the variables other than V.

$$\mathcal{E}_6 = \begin{pmatrix} W := S \vee B \\ B := \overline{S} \end{pmatrix}$$

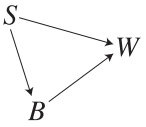

FIGURE 6: $\mathcal{E}_6$ eclipses $\mathcal{E}_5$.

In §3, we saw that a structural equations model says more than just that the endogenous variables $V \in \mathcal{V}$ are structurally determined by their parents, and are not structurally determined by any of the other variables in $\mathcal{U} \cup \mathcal{V}$. It additionally says that the exogenous variables aren't determined by any of the other variables in $\mathcal{U} \cup \mathcal{V}$. We can accomplish this within the nomic sufficiency account in the following way. Say that one model $\mathcal{M}' = (\mathcal{U}', \mathcal{V}', \mathcal{E}')$ *eclipses* another model $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{E})$, $\mathcal{M} \sqsubset \mathcal{M}'$, iff $\mathcal{M}'$ and $\mathcal{M}$ share all the same variables and $\mathcal{M}'$ contains strictly more structural determination relations between those variables. That is:

$\mathcal{M} \sqsubset \mathcal{M}'$ iff:
1. $\mathcal{U} \cup \mathcal{V} = \mathcal{U}' \cup \mathcal{V}'$
2. $\forall V \in \mathcal{U} \cup \mathcal{V}$, $\mathbf{PA}(V) \subseteq \mathbf{PA}'(V)$
3. $\exists V \in \mathcal{U} \cup V$, $\mathbf{PA}(V) \subsetneq \mathbf{PA}'(V)$

(Where '$\mathbf{PA}'(V)$' is a vector of V's structural parents in the model $\mathcal{M}'$.) Now, we can enrich our account of the correctness of causal models by requiring that a model not be eclipsed by any other model which satisfies ($\mathcal{V}4$).

($\mathcal{M}2$) $\mathcal{M}$ is correct at w iff:
1. $\mathcal{M}$ satisfies ($\mathcal{V}4$) at w
2. $\neg\exists\mathcal{M}'$ such that $\mathcal{M}'$ satisfies ($\mathcal{V}4$) at w and $\mathcal{M} \sqsubset \mathcal{M}'$

Returning to the example of Billy, Suzy, and the window (shown in figure 5): if Billy is eager to see the window shatter, and will throw his rock if (but only if) Suzy doesn't throw hers, then, given some assumptions about the selection function f, the system of structural equations shown in figure 6 will satisfy ($\mathcal{V}4$). ('$\overline{x}$' is the truth function $1 - x$.) And this causal model eclipses the model consisting of the sole structural equation $W := S \vee B$, shown in figure 5. So, according to ($\mathcal{M}2$), the causal model in figure 5 will not be correct, if this one is. So, if Billy's decision about whether or not to throw is determined by whether Suzy throws, then the model consisting of just the equation $W := S \vee B$ is not correct. That model tells us that whether Billy throws

isn't determined by whether Suzy throws, which is false.

4.1 Interventions and Modularity

The nomic sufficiency account of causal models affords an understanding of hypothetical interventions. On this understanding, a hypothetical intervention on an endogenous variable V is just a counterfactual supposition which takes one outside of the causal field $\mathfrak{F}_V$, while remaining inside the causal fields of all the other endogenous variables in the model. Consider, for instance, the causal model of the captain, the riflemen, and the deserter shown in figure 2. Suppose that this causal model is correct at w_0, and that $C_{w_0} = 0$—the captain doesn't give the order at w_0. Suppose that we wish to entertain a hypothetical intervention to set the value of L to 1. We know that this is to be modeled in the following way: we take the original system of structural equations $\mathcal{E}_2$ and replace it with $\mathcal{E}_{2,L=1}$, shown in figure 3. In the mutilated system of equations $\mathcal{E}_{2s,L=1}$, L does not merely take on the value of 1. Additionally, the value of L is not structurally determined by the value of C. Whether the left rifleman fires is not determined by whether the captain gives the order. This follows immediately from the correctness of the system of equations $\mathcal{E}_{2,L=1}$, given ($\mathcal{M}2$), since if L were still structurally determined by C, then $\mathcal{M}_{2,L=1}$ would be eclipsed by $\mathcal{M}_2$.

Since we've said that a structural equation ϕ_V is in force at a world w iff w lies inside of a causal field $\mathfrak{F}_V$ satisfying ($\mathfrak{F}2$), this means that a hypothetical intervention to set the value of L must take us to a world w_1 which lies *outside* of $\mathfrak{F}_L$ (since L is not structurally determined by C), but still *inside* of $\mathfrak{F}_R$ and $\mathfrak{F}_D$ (since D is still structurally determined by L and R, and R is still structurally determined by C), as shown in figure 7(a). This provides a semantic interpretation of what's going on when we model an intervention on L by removing Ls's structural equation and leaving the other structural equations in place.

It also provides an explanation of why only certain methods of setting the value of L to 1 count as *interventions*, and it provides a criterion for distinguishing those ways of setting the values of the variables which do from those which do not constitute interventions. For instance, if we were to get the left rifleman to fire by bribing the captain to give him the order, then this would not constitute an intervention on L, since it would leave us *inside* of the causal field $\mathfrak{F}_L$, as shown in figure 7(b). Similarly, suppose that the captain does not want to kill the deserter, but would welcome an opportunity to let the riflemen blow off some steam. Then, we might be able to get the left rifleman to fire by putting up a bullet-proof partition between the riflemen

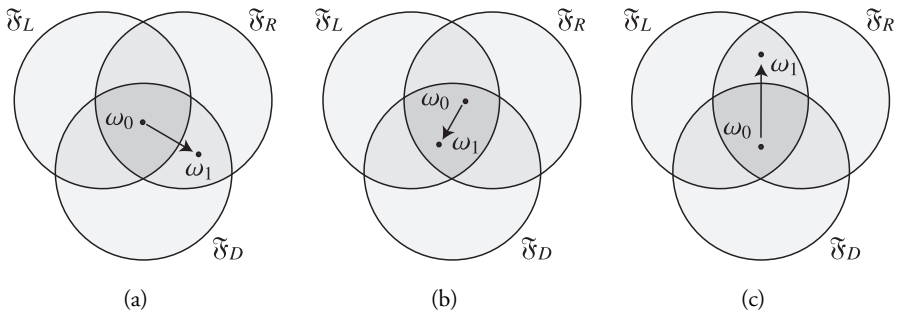

FIGURE 7: Interventions on the nomic sufficiency understanding

and the deserter. Then, the captain would give the order, and the left rifleman would fire. Even though this is an intervention which makes the left rifleman fire, it is not an intervention *on* on the value of L. Rather, since it leaves L and R's determination by C intact, but severs the determination of D by L and R, as shown in figure 7(c), it constitutes an intervention on D.

This generalizes. A causal model $\mathcal{M}$ will be correct throughout $\mathfrak{F}_{\mathcal{M}} \stackrel{\text{def}}{=} \bigcap_{V \in \mathcal{V}} \mathfrak{F}_V$. This is the area of modal space in which every endogenous variable's structural equation $V \in \mathcal{V}$ is in force—it is the area in which all of the causal fields of the endogenous variables overlap. Theorem 3, proved in the appendix, establishes that, given ($\mathcal{M}2$), this area of modal space will contain every assignment of values to $\mathcal{U} \cup \mathcal{V}$ which is consistent with the structural equations in $\mathcal{E}$, and no assignments of values to $\mathcal{U} \cup \mathcal{V}$ which is inconsistent with the structural equations in $\mathcal{E}$.

Theorem 3. *Given* ($f1$)*, if a causal model* $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{E})$ *is correct according to* ($\mathcal{M}2$)*, then* $\mathfrak{F}_{\mathcal{M}} \stackrel{\text{def}}{=} \bigcap_{V \in \mathcal{V}} \mathfrak{F}_V$ *contains all and only the allowed assignment of values to the variables in* $\mathcal{U} \cup \mathcal{V}$*, where an assignment is* allowed *just in case it is a solution to the equations in* $\mathcal{E}$*.*

In particular, this means that, for any assignment of values to the exogenous variables, there will be some area of modal space inside $\mathfrak{F}_{\mathcal{M}}$ where that assignment of values is realized. So there are in-principle hypothetical interventions to set the values of any of the exogenous variables without disrupting any of the structural determination relations in $\mathcal{E}$.

Moreover, it follows from ($\mathcal{M}2$) that, for any correct causal model $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{E})$, any $\mathbf{V} \subsetneq \mathcal{V}$, and any assignment of values to $\mathbf{V}$, there will always be an area of modal space which is *outside* the causal fields of all the members of $\mathbf{V}$ but still *inside* the causal

fields of all of the members of $\mathcal{V} - \mathbf{V}$, and which contains every possible assignment of values to $\mathbf{V} \cup \mathcal{U}$.

Theorem 4. *Given* ($\mathcal{M}2$) *and* ($f3$)*, for any* $\mathbf{V} \subseteq \mathcal{V}$*,*

$$\bigcap_{W \notin \mathbf{V}} \mathfrak{F}_W - \bigcup_{V \in \mathbf{V}} \mathfrak{F}_V$$

is non-empty and contains every assignment of values to the variables in $\mathbf{V} \cup \mathcal{U}$*.*

This means that, if a structural equations model is correct, according to ($\mathcal{M}2$), then there is an in-principle intervention to set any subset of the variables in $\mathcal{V}$ to any assignment of values which will leave the structural equations of the non-intervened-upon variables intact. This, together with theorems 2 and 3, guarantees that the structural equations in $\mathcal{E}$ are strongly modular.

4.2 Causal Counterfactual Dependence

($\mathcal{M}2$) allows us to provide an account of causal counterfactual conditionals in terms of relations of structural determination. On this account, a causal counterfactual $\mathrm{A} \boxdot\!\!\rightarrow \mathrm{C}$ is true at a world w iff there is a causal model $\mathcal{M}$, correct at w, such that, given the exogenous variable assignment $\mathcal{U}_w$, if $\mathcal{M}$ is minimally mutilated so as to make A true, then C is true in the resulting model.

($\boxdot\!\!\rightarrow_{\mathcal{M}}$) $\qquad \mathrm{A} \boxdot\!\!\rightarrow \mathrm{C} \iff \mathcal{M}_{\mathrm{A}}, \mathcal{U}_w \models \mathrm{C}$

With this kind of account, we could take structural determination to be more primitive than causal counterfactual dependence, and use the former to provide an account of the latter. That is to say: with this account, we need not define causal counterfactual dependence directly in terms of the selection function; rather, f can be used to provide truth conditions for $\mathcal{M}$, which can be used to provide truth conditions for $\boxdot\!\!\rightarrow$.

The counterfactual understanding, in contrast, retained an account of causal counterfactual conditionals according to which $\mathrm{A} \boxdot\!\!\rightarrow \mathrm{C}$ is true at the world of evaluation, w, iff all the worlds in $f(\mathrm{A}, w)$ are worlds at which C is true.

($\boxdot\!\!\rightarrow_f$) $\qquad \mathrm{A} \boxdot\!\!\rightarrow \mathrm{C} \iff f(\mathrm{A}, w) \models \mathrm{C}$

Depending upon our semantics for f, there may be cases in which ($\boxdot\!\!\rightarrow_{\mathcal{M}}$) and ($\boxdot\!\!\rightarrow_f$) diverge. Just to fix ideas: consider an account roughly like that of Lewis (1979) or

$$\mathcal{E}_8 = (W := B + B \cdot H)$$

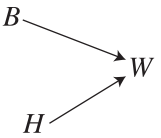

FIGURE 8: The system of structural equations $\mathcal{E}_8$

MAUDLIN (2007). On Maudlin's account, $f(\mathrm{A}, w)$ is the set of worlds that you get by performing a surgical alteration on w so as to make A true at the relevant time, and then time-evolving the resulting state of the world forward in time according to the fundamental laws of nature. While LEWIS (1979)'s account is slightly more complicated, it will achieve the same results as Maudlin's in the case I'll be considering.

Imagine that I've got a tychistically chancy coin—whether it lands heads is not determined by the previous microphysical state of the universe and the laws of nature; rather, given that it's flipped, the previous state of the universe and the laws of nature assign a probability of one half to the coin landing heads and a probability of one half to the coin landing tails. I'm going to flip the coin, and I offer you a bet on whether or not the coin lands heads. I'm an honest player, so if you take the bet and the coin lands heads, then you'll win some money. If you take the bet and the coin lands tails, then you'll lose some money. If you don't take the bet, then you'll neither win nor lose any money, independent of whether or not the coin lands heads. Let's stipulate that the chance that the coin lands heads is unaffected by whether you take the bet. In this scenario, it appears that the structural equations model shown in figure 8 is correct, where B is a binary variable that takes the value 1 if you accept the bet and 0 if you don't accept the bet, H is a binary variable which takes the value 1 if the coin lands heads and 0 if the coin lands tails, and W is a ternary variable which takes the value 0 if you neither win nor lose money, 1 if you lose money, and 2 if you win money. Let's say that, at the actual world, you refuse the bet and the coin lands heads.

Suppose that we adopt the Maudlin account of the selection function. Then, ($\mathcal{M}1$) and ($\Box\!\!\rightarrow_f$) will tell us that this structural equations model is not correct, since condition ($\mathcal{U}1$) will not be satisfied. ($\mathcal{U}1$), recall, required that, were some of the exogenous variables to have taken on different values, the other exogenous variables would have retained their actual values. However, $f(B = 1, @) \not\models H = 1$, since when we surgically alter the state of the world so as to make $B = 1$ true and time-evolve the resulting state of the world into the future according to the fundamental laws of nature, there are two possibilities: one in which the coin lands heads and one in which the coin lands tails. In contrast, the causal model in figure 8 will satisfy the second clause of ($\mathcal{M}2$)—

($\mathcal{M}2$, 2)—so long as there is no other structural equations model which satisfies ($\mathcal{M}2$, 1) according to which W is determined by B and H, and either B is determined by H or H is determined by B. Since $f(B = 1, @)$ contains both $H = 0$ worlds and $H = 1$ worlds, there will be no set of worlds containing $f(B = 1, @)$ which entails that H is any function of B. So H is not determined by B, according to ($\mathcal{M}2$, 1). Assuming that, under the counterfactual supposition that the coin lands tails, you still refused the bet, the structural equations model shown in figure 8 will satisfy ($\mathcal{M}2$, 2).

Of course, there's no reason that the counterfactual account can't replace ($\mathcal{U}1$) with the requirement that a causal model be uneclipsed. On an account like this, a model $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{E})$ will be correct iff $\mathcal{V}$ satisfies ($\mathcal{V}1$) and there is no other structural equations model $\mathcal{M}'$ which satisfies ($\mathcal{V}1$) and eclipses $\mathcal{M}$. Even this emended counterfactual account will fail to say that the system of equations in figure 8 is correct. For $f(B = 1, @) \not\models W = 2$, since there are some worlds in $f(B = 1, @)$ where the coin lands tails and you therefore lose the bet. So, ($\mathcal{V}1$), wedded with a Maudlin-esque account of the selection function, entails that the system of structural equations in figure 8 is incorrect.

Independent of its ability to vindicate the system of structural equations $\mathcal{E}_8$, the fact that this account of the selection function, together with ($\Box\!\!\rightarrow_f$), entails the falsity of $B = 1 \Box\!\!\rightarrow W = 2$ strikes me, as it has struck many,[15] as the wrong result. Whether the coin lands heads is entirely unaffected by whether you took the bet. Since the coin actually landed heads, if you had taken the bet, you *would* have won. Now, there are moves to be pulled here—we can alter our account of the selection function so that, if the coin actually lands heads, then only the worlds where the coin lands heads are included in $f(B = 1, @)$. Note, however, that the account consisting of ($\mathcal{M}2$) and ($\Box\!\!\rightarrow_{\mathcal{M}}$) need not avail itself of those maneuvers. Even with the bare Maudlin account of the selection function, that account entails that, were you to take the bet, you would have won. According to ($\mathcal{M}2$), the correctness of the structural equations model in figure 8 does not depend upon whether the worlds in $f(B = 1, @)$ are worlds in which the coin lands heads or tails, or whether they are worlds in which you win or lose. ($\mathcal{M}2$) only requires that, at all the worlds in $f(B = 1, @)$ at which the coin lands tails, you lose; and that, at all the worlds in $f(B = 1, @)$ at which the coin lands heads, you win. Assuming that similar remarks apply to all the other worlds in $\mathfrak{F}_W$, the structural determination relations shown in figure 8 will be in force. Then, those structural determination relations will entail, *via* ($\Box\!\!\rightarrow_{\mathcal{M}}$), that $B = 1 \Box\!\!\rightarrow W = 2$. So, according to ($\mathcal{M}2$) and ($\Box\!\!\rightarrow_{\mathcal{M}}$), it is possible for a counterfactual A $\Box\!\!\rightarrow$ C to be

[15] See Bennett (2003, ch. 15) and Kment (2006)

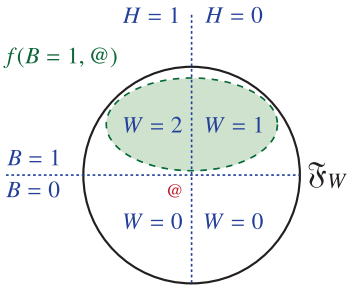

FIGURE 9: The relationship between the causal field $\mathfrak{F}_W$, the selection function f, and the counterfactual $B = 1 \boxrightarrow W = 2$.

true at a world w, even though $f(\mathrm{A}, w) \not\models \mathrm{C}$. (See figure 9.)

This is not easily mimicked by the counterfactual account, for that account is committed to both ($\boxrightarrow_{\mathcal{M}}$) and ($\boxrightarrow_f$). For instance, the counterfactual theorist might want to attempt to adopt the nomic sufficiency account's treatment of the coin toss case by emending ($\mathcal{V}1$) to read:

($\mathcal{V}5$) $\quad \forall V \in \mathcal{V} \quad \forall \mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V) \quad \forall \mathbf{x} \quad (\mathbf{X} = \mathbf{x} \;\boxrightarrow\; V = \phi_V(\mathbf{PA}(V)))$

($\mathcal{V}5$), unlike ($\mathcal{V}1$), does not require that, were $\mathbf{X}$ to take on the values $\mathbf{x}$, V would take on the value it is given in the mutilated model $\mathcal{M}_{\mathbf{X}=\mathbf{x}}$, with the actual assignment of values to the exogenous variables. It simply requires that, whatever values V's parent variables end up taking on when the values of $\mathbf{X}$ change, the value of V remains a function ϕ_V of those values. This would allow the counterfactual theorist to agree with the nomic sufficiency theorist that were you to have taken the bet, you would have won, $B = 1 \boxrightarrow W = 2$. However, since the counterfactual theorist is still committed to ($\boxrightarrow_f$), so long as they retain the simple Maudlin account of the selections function, they must also *deny* that were you to have taken the bet, you would have won, since $f(B = 1, @) \not\models W = 2$. And this is a straightforward contradiction.

The counterfactual theorist might want to respond to these kinds of considerations by denying ($\boxrightarrow_f$), and reformulating their account of the correctness conditions of causal models directly in terms of the selection function f, saying nothing of counterfactuals. That is, they could replace ($\mathcal{V}5$) with ($\mathcal{V}6$).

($\mathcal{V}6$) $\quad \forall V \in \mathcal{V} \quad \forall \mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V) \quad \forall \mathbf{x} \quad f(\mathbf{X} = \mathbf{x}, @) \models V = \phi_V(\mathbf{PA}(V))$

It follows from lemmas 1 and 2 (§5) that, in the presence of ($\mathcal{U}1$), ($\mathcal{V}6$) is equivalent

to ($\mathcal{V}1$). Of course, we just encountered reason for such a theorist to abandon ($\mathcal{U}1$)—namely that, together with the Maudlin-esque account of f, it is inconsistent with the correctness of $\mathcal{E}_8$. And once ($\mathcal{U}1$) has been replaced with a condition along the lines of ($\mathcal{M}2$, 2), ($\mathcal{V}6$) will no longer be equivalent to ($\mathcal{V}1$).[16] Given that it denies any direct connection between f and counterfactual conditionals, we might well wonder whether the resulting account deserves the name 'counterfactual' any longer, but put that question to the side. Whatever we call the resulting account, it is only able to avoid complicating its account of f by inching ever closer to the nomic sufficiency account. The only thing separating the two accounts at this point is the nomic sufficiency account's closure condition, ($\mathfrak{F}2$). This is the aspect of the account which solves the problems with modularity raised in §3.1 above. It appears that any counterfactual account built around ($\mathcal{V}6$) which was able to solve those problems with modularity would end up being equivalent to (or would entail) the nomic sufficiency account. For it appears that the only way to solve those problems is to impose a constraint on which worlds are reachable by repeated counterfactual supposition; for worlds that are so reachable, put them in the set $\mathfrak{F}_V$, and modularity will then guarantee that $\mathfrak{F}_V \models V = \phi_V(\mathbf{PA}(V))$. If that's right, then such an account would impose all the same constraints as the nomic sufficiency account; and counterfactual theorists would have mimicked nomic sufficiency theorists only by becoming nomic sufficiency theorists themselves.

4.3 A Remaining Worry

Above, I defined modularity as the thesis that any number of interventions on a set of variables $\mathbf{V}$ leaves the structural equations associated with every endogenous variable $V \notin \mathbf{V}$ unaffected. Theorems 2–4 guarantee that a correct system of structural equations will be modular in this sense. Note, however, that modularity does not guarantee that there will always be an intervention on a set of variables $\mathbf{V}$ such that, post-intervention, the variables in $\mathbf{V}$ are *no longer* determined by any of the variables in $\mathbf{PA}(\mathbf{V})$. For a schematic example, consider the causal fields shown in figure 10. In that diagram, the causal model $\mathcal{M}$ will be correct at the world $w_0 \in \mathfrak{F}_{\mathcal{M}} \stackrel{\text{def}}{=} \bigcap_{V \in \mathcal{V}} \mathfrak{F}_V$. Now, theorem 4 guarantees us that there is a set of worlds $f(\mathbf{V} = \mathbf{v}, w_0)$ which lie outside of $\bigcup_{V \in \mathbf{V}} \mathfrak{F}_V$, yet inside of $\bigcap_{U \notin \mathbf{V}} \mathfrak{F}_U$, and at which $\mathbf{V} = \mathbf{v}$, for any assignment $\mathbf{v}$. However, we have no guarantee that this set of worlds *doesn't* lie within some *other* causal field $\mathfrak{F}'_V$ such that $\mathfrak{F}'_V \models V = \phi'_V(\mathbf{PA}(V), \dots)$, for some $V \in \mathbf{V}$. And

[16] The case currently under discussion provides a counterexample to the equivalence.

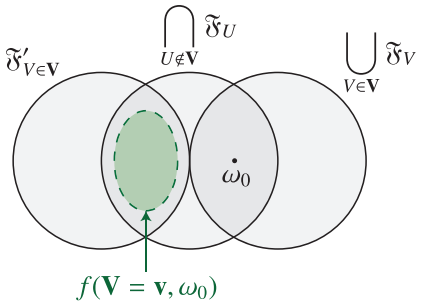

FIGURE 10: On the nomic sufficiency account, an intervened-upon variable may still be determined by its structural parents.

that means that, even though we have a guarantee that an *intervention* on a set of variables will sever the actual structural determination relations between $\mathbf{V}$ and $\mathbf{PA}(\mathbf{V})$, we *don't* have any guarantee that the intervention won't make it the case that some *other* structural determination relations link $\mathbf{PA}(\mathbf{V})$ to $\mathbf{V}$.

For a concrete example which might give rise to a case like this, consider the steam vent illustrated in figure 11. There, a switch, which may be placed to the left or to the right, will either divert the steam to the left or the right. (If the switch is left, the steam will go right, as shown in figure 11(a); if the switch is right, the steam will go left, as shown in figure 11(b).) There is a lid on the right steam vent. If the steam is directed up to the right vent, then the lid will heat up. Consider the variables S and L. S is 1 if the switch is to the left, and is 0 if the switch is to the right. L is 1 if the lid is hot and is 0 if the lid is not hot. When the system is as depicted in figure 11(a), the structural equation $L := S$ will be in force. Whether the lid is hot is determined by whether the switch is to the left or right. In figure 11(a), both S and L will be 1. Now, suppose that the lid is attached to a hinge, so that it can be pivoted to sit atop either the left or the right steam vent. There is then an intervention we may perform to set L to 0. That is, there is a method for making the lid not hot which will take us outside of the causal field $\mathfrak{F}_L$. We may simply pivot the lid on its hinge to put it atop the left steam vent, as in figure 11(c). Then, it will no longer be the case that $L = 1$, nor will it be the case that S determines L according to the equation $L := S$. However, even after this intervention has taken place, the value of S will determine the value of L. It will now do so according to the equation $L := \overline{S}$. If the switch is set to the left, then the lid will not be hot, and if the switch is set to the right, then the lid will be hot.

Given the account of interventions provided in §4.1 above, this will count as an intervention on the value of L. However, it would be inappropriate to model the

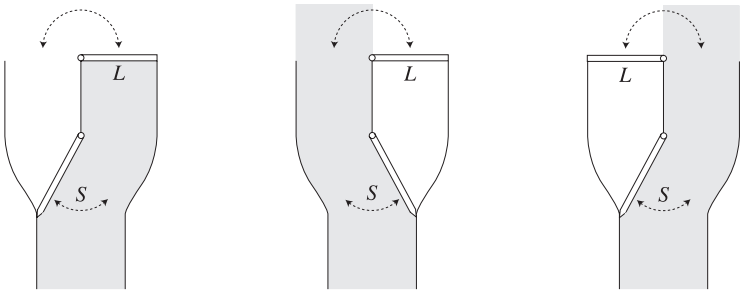

FIGURE 11: An example of how an intervened-upon variable may still be determined by its structural parents

result of this intervention by mutilating the model, removing L's structural equation, and replacing it with nothing. For, in order for the mutilated model to be correct, it must be uneclipsed by any correct structural equations model. And, in this case, the mutiliated model in which S does not structurally determine L would be eclipsed by the model containing the equation $L := \overline{S}$.

Cases such as these might also make trouble for ($\Box\!\!\rightarrow_{\mathcal{M}}$), as ($\Box\!\!\rightarrow_{\mathcal{M}}$) would predict that, if we are in the situation depicted in figure 11(a), then, were the lid to not be hot, then, if the switch were moved to the right, the lid would not be hot—$L = 0 \,\Box\!\!\rightarrow (S = 0 \,\Box\!\!\rightarrow L = 0)$. However, if we think that there are worlds in $f(L = 0, @)$ at which the lid has been pivoted on its hinge, then we might think that this counterfactual should be false.

It is unclear whether this ought to be regarded as a problem for the nomic sufficiency account. To the extent that one is inclined to think that $f(L = 0, @)$ includes worlds at which the lid has been pivoted on its hinge, it seems to me entirely correct to say that we ought not model an intervention on L which pivots the lid on its hinge by removing the structural determination relation between L and S, and it seems a mark in the nomic sufficiency account's favor that it says so. To the extent that one is inclined to think that $f(L = 0, @)$ contains worlds at which the lid is removed from its hinge, or perhaps worlds at which some kind of Lewisian miracle keeps the lid from getting hot even though the steam is being directed up towards it, it seems entirely correct to model this kind of intervention by removing the structural determination relation between S and L. Once there is a miracle to set L to 0, changes in the value of S will not affect the value of L, so long as God's hand is steady. If one is unhappy with

the possibility of L being determined by S post-intervention, then one may simply require that $f(L = 0, @)$ contain only worlds at which Lewisian miracles determine the value of L. Of course, nothing in the account guarantees that there will always be some possible Lewisian miracle which will constitute an intervention. However, if there are cases in which Lewisian miracle interventions are impossible, then I'm inclined to say just what I said above about the case in which $f(L = 0, @)$ contained worlds at which the lid was pivoted on its hinge: in such cases, it is incorrect to model the result of the intervention by mutilating the model, and it would be a mark against an account of structural determination if it said otherwise. So my settled judgment is that this is a feature, rather than a bug, of the nomic sufficiency account.

Some will disagree. They will have the following reaction to this case: the structural equations model containing just the equation $L := S$ ought not be deemed correct by an account of structural determination. For there is another, *better*, model which contains an additional variable describing the position of the lid. For instance, if we use the variable P, which is 1 if the lid is pivoted to the left and 0 if the lid is pivoted to the right, then the model $((S, P), (L), (L := S \veebar P))$ will be correct.[17,18] I agree that this new model is correct; however, it appears to me that, so long as the lid is pivoted to the right, the original model is correct, too. And I see no reason why these two models cannot both be correct together. There are a great many structural equations models to which we could, if we chose, add additional exogenous variables, but they are none the worse for that. A model in which whether the match lights is determined by whether it is struck is correct, even though we could add to it an additional variable for the presence of oxygen. We often wish to ignore certain background properties of the world which determine an outcome; and an account of structural equations models ought to permit this. However, if one is not persuaded by these considerations, and one wishes to rule out structural equations like $L := S$ when there are correct structural equations like $L := S \veebar P$ to replace them, then the nomic sufficiency account could be emended to achieve this in a variety of ways.[19] From my perspective, such emendations are unnecessary and ill-advised; but I would not be surprised to learn that others disagree.

[17] $x \veebar y$ is the exclusive 'or', which is 1 iff $x \neq y$.

[18] Thanks to an anonymous reviewer for pressing me to consider this objection.

[19] For one: we could simply emend the definition of *eclipsing* by removing condition (1).

5 In Summation

After developing the counterfactual understanding of structural determination, I argued that it faces difficulties in securing the modularity of structural determination relations. I advanced an alternate understanding of structural determination and I demonstrated that it guarantees that structural determination relations are modular. I showed that it provides a clear and straightforward way of thinking about hypothetical interventions, as well as a criterion for distinguishing hypothetical changes in the values of variables which constitute interventions from those that do not. By treating structural determination relations as more fundamental than causal counterfactuals, the resulting theory was able to yield a clean solution to a problem case for 'closest possible world' semantics for counterfactuals.

Proofs

Define the *rank* of a variable $V \in \mathcal{U} \cup \mathcal{V}$ recursively as follows:

$$\begin{aligned} rank(V) = 0 &\iff V \in \mathcal{U} \\ rank(V) = k+1 &\iff \max\{rank(P) \,:\, P \in \mathbf{PA}(V)\} = k \end{aligned}$$

Graphically, a variable's rank is the largest number of edges lying between that variable and an exogenous variable along a directed path. Let '$\mathbf{Rank}(i)$' denote the set of all variables of rank i, and let '$\mathbf{Rank}(i, j, \ldots, k)$' denote the union $\mathbf{Rank}(i) \cup \mathbf{Rank}(j) \cup \cdots \cup \mathbf{Rank}(k)$.

Lemma 1. *Given* ($\mathcal{V}1$), ($\mathcal{U}1$), *and* ($f1$), *for all* $V \in \mathcal{V}$, *all* $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$, *all* $\mathbf{x}$, *and all* $w' \in f(\mathbf{X} = \mathbf{x}, w)$, $\mathbf{PA}(V)_{w'} = \mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}$, *where* $\mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}$ *assigns the values to* $\mathbf{PA}(V)$ *determined by the structural equations in* $\mathcal{E} - \bigcup_i(\phi_{X_i})$, *for every endogenous* $X_i \in \mathbf{X}$, *and* $(\mathcal{U} - \mathbf{X})_w \cup \mathbf{x}$.

Proof. By induction on the rank of the variables in $\mathcal{V}$.

Base Case. *For all* $V \in \mathbf{Rank}(1)$, *all* $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$, *all* $\mathbf{x}$, *and all* $w' \in f(\mathbf{X} = \mathbf{x}, w)$, $\mathbf{PA}(V)_{w'} = \mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}$.

Proof. If $rank(V) = 1$, then every $P \in \mathbf{PA}(V)$ is exogenous. Without loss of generality, consider one $P \in \mathbf{PA}(V)$. If $P \in \mathbf{X}$, then $f(\mathbf{X} = \mathbf{x}, w) \models P = P_{\mathbf{X}=\mathbf{x}}$ (the value assigned to P by $\mathbf{x}$), by ($f1$). If $P \notin \mathbf{X}$, then $f(\mathbf{X} = \mathbf{x}, w) \models P = P_w$, by ($\mathcal{U}1$). In either case, P takes on the value assigned to it by $\mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}$. □

Inductive Step. *If for all* $V \in \mathbf{Rank}(1, 2, \ldots, k)$, *it is true that, for all* $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$, *all* $\mathbf{x}$, *and all* $w' \in f(\mathbf{X} = \mathbf{x}, w)$, $\mathbf{PA}(V)_{w'} = \mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}$, *then for all* $V \in \mathbf{Rank}(k+1)$, *it will be true that, for all* $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$, *all* $\mathbf{x}$, *and all* $w' \in f(\mathbf{X} = \mathbf{x}, w)$, $\mathbf{PA}(V)_{w'} = \mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}$.

Proof. Without loss of generality, consider one $V \in \mathbf{Rank}(k+1)$, one $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$, one $\mathbf{x}$, and one $P \in \mathbf{PA}(V)$. Either $P \in \mathbf{X}$ or $P \notin \mathbf{X}$. Suppose that $P \in \mathbf{X}$. Then, $f(\mathbf{X} = \mathbf{x}, w) \models P = P_{\mathbf{X}=\mathbf{x}}$, by ($f1$). If $P \notin \mathbf{X}$, then, since $rank(P) \leq k$, $\mathbf{PA}(P)_{w'} = \mathbf{PA}(P)_{\mathbf{X}=\mathbf{x}}$, for all $w' \in f(\mathbf{X} = \mathbf{x}, w)$, by the inductive hypothesis (since $P \notin \mathbf{X}, \mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - P)$). Then, ($\mathcal{V}1$) guarantees that $f(\mathbf{X} = \mathbf{x}, w) \models P = \phi_P(\mathbf{PA}(P)_{\mathbf{X}=\mathbf{x}})$. So, whether $P \in \mathbf{X}$ or $P \notin \mathbf{X}$, P takes on the value $P_{\mathbf{X}=\mathbf{x}}$ at every $w' \in f(\mathbf{X} = \mathbf{x}, w)$. Since P, V, $\mathbf{X}$, and $\mathbf{x}$ were arbitrary, for all $V \in \mathbf{Rank}(k+1)$, all $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$, and all $\mathbf{x}$, $\mathbf{PA}(V)_{w'} = \mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}$ for every $w' \in f(\mathbf{X} = \mathbf{x}, w)$. □

□

Lemma 2. *Given* ($\mathcal{V}3$), ($\mathcal{U}1$), *and* ($f1$) *for all* $V \in \mathcal{V}$, *all* $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$, *all* $\mathbf{x}$, *and all* $w' \in f(\mathbf{X} = \mathbf{x}, w)$, $\mathbf{PA}(V)_{w'} = \mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}$, *where* $\mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}$ *assigns the values to* $\mathbf{PA}(V)$ *determined by the assignment of values* $(\mathcal{U} - \mathbf{X})_w \cup \mathbf{x}$ *and the structural equations in* $\mathcal{E} - \bigcup_i(\phi_{X_i})$, *for every endogenous* $X_i \in \mathbf{X}$.

Proof. By induction on the rank of the variables in $\mathcal{V}$.

Base Case. *For all* $V \in \mathbf{Rank}(1)$, *all* $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$, *all* $\mathbf{x}$, *and all* $w' \in f(\mathbf{X} = \mathbf{x}, w)$, $\mathbf{PA}(V)_{w'} = \mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}$.

Proof. Consider, without loss of generality, a variable $V \in \mathbf{Rank}(1)$. Since V's rank is 1, every $P \in \mathbf{PA}(V)$ is exogenous. If $P \in \mathbf{X}$, then $f(\mathbf{X} = \mathbf{x}, w) \models P = P_{\mathbf{X}=\mathbf{x}}$ (the value assigned to X by $\mathbf{x}$), by ($f1$). If $P \notin \mathbf{X}$, then $f(\mathbf{X} = \mathbf{x}, w) \models P = P_w$, by ($\mathcal{U}1$). In either case, P takes on the value assigned to it by $\mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}$. □

Inductive Step. *If for all* $V \in \mathbf{Rank}(1, 2, \ldots, k)$, *it is true that, for all* $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$, *all* $\mathbf{x}$, *and all* $w' \in f(\mathbf{X} = \mathbf{x}, w)$, $\mathbf{PA}(V)_{w'} = \mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}$, *then for all* $V \in \mathbf{Rank}(k+1)$, *it will be true that, for all* $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$, *all* $\mathbf{x}$, *and all* $w' \in f(\mathbf{X} = \mathbf{x}, w)$, $\mathbf{PA}(V)_{w'} = \mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}$.

Proof. Without loss of generality, consider one $V \in \mathbf{Rank}(k+1)$, one $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$, one $\mathbf{x}$, and one $P \in \mathbf{PA}(V)$. Either $P \in \mathbf{X}$ or $P \notin \mathbf{X}$. Suppose that $P \in \mathbf{X}$. Then, $f(\mathbf{X} = \mathbf{x}, w) \models P = P_{\mathbf{X}=\mathbf{x}}$, by ($f1$). If $P \notin \mathbf{X}$, then, since $rank(P) \leq k$, $\mathbf{PA}(P)_{w'} = \mathbf{PA}(P)_{\mathbf{X}=\mathbf{x}}$, for all $w' \in f(\mathbf{X} = \mathbf{x}, w)$, by the inductive hypothesis (since $P \notin \mathbf{X}, \mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - P)$. Then, ($\mathcal{V}3$) and ($\mathfrak{F}1$) guarantee that $f(\mathbf{X} = \mathbf{x}, w) \subset \mathfrak{F}_P$ and $\mathfrak{F}_P \models P = \phi_P(\mathbf{PA}(P)_{\mathbf{X}=\mathbf{x}})$. So $f(\mathbf{X} = \mathbf{x}, w) \models P = \phi_P(\mathbf{PA}(P)_{\mathbf{X}=\mathbf{x}})$ So, whether $P \in \mathbf{X}$ or $P \notin \mathbf{X}$, P takes on the value $P_{\mathbf{X}=\mathbf{x}}$ at every $w' \in f(\mathbf{X} = \mathbf{x}, w)$. Since P, V, $\mathbf{X}$, and $\mathbf{x}$ were arbitrary, for all $V \in \mathbf{Rank}(k+1)$, all $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$, and all $\mathbf{x}$, $\mathbf{PA}(V)_{w'} = \mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}}$ for every $w' \in f(\mathbf{X} = \mathbf{x}, w)$. □

□

Theorem 1. *Given* ($f1$), *in a causal model* $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{E})$, *if* $\mathcal{U}$ *satisfies* ($\mathcal{U}1$), *then* $\mathcal{V}$ *satisfies* ($\mathcal{V}3$) *iff* $\mathcal{V}$ *satisfies* ($\mathcal{V}1$).

Proof. First assume that $\mathcal{V}$ satisfies ($\mathcal{V}1$). Then, we know that for all $V \in \mathcal{V}$, all $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$, and all assignments $\mathbf{x}$ to $\mathbf{X}$,

$$f(\mathbf{X} = \mathbf{x}, w) \models V = \phi_V(\mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}})$$

By lemma 1, it then follows that

$$\forall w' \in f(\mathbf{X} = \mathbf{x}, w), V_{w'} = \phi_V(\mathbf{PA}(V)_{w'})$$

So

$$f(\mathbf{X} = \mathbf{x}, w) \models V = \phi_V(\mathbf{PA}(V))$$

So, if for every V, every $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$, and every $\mathbf{x}$, we include every $w' \in f(\mathbf{X} = \mathbf{x}, w)$ in $\mathfrak{F}_V$, then we will have a set $\mathfrak{F}_V$ which satisfies ($\mathfrak{F}1$) and which entails that $V = \phi_V(\mathbf{PA}(V))$. So every $V \in \mathcal{V}$ will satisfy ($\mathcal{V}3$).

To establish the other direction, assume that $\mathcal{V}$ satisfies ($\mathcal{V}3$). Then, for every $V \in \mathcal{V}$, every $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$, and every $\mathbf{x}$, $f(\mathbf{X} = \mathbf{x}, w) \in \mathfrak{F}_V$ and $\mathfrak{F}_V \models V = \phi_V(\mathbf{PA}(V))$. By lemma 2, it then follows that, for all V, $\mathbf{X}$, and $\mathbf{x}$,

$$f(\mathbf{X} = \mathbf{x}, w) \models V = \phi_V(\mathbf{PA}(V)_{\mathbf{X}=\mathbf{x}})$$

So $\mathcal{V}$ must satisfy ($\mathcal{V}1$) as well. □

Theorem 2. *Given ($f3$), if ϕ_V belongs to a causal model satisfying ($\mathcal{V}4$) at a world w_0, then ϕ_V will continue to belong to a causal model satisfying ($\mathcal{V}4$) after any number of consecutive hypothetical interventions to set the values of any $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$.*

Proof. By induction on the number of interventions.

Inductive Step. *If ϕ_V belongs to a causal model satisfying ($\mathcal{V}4$) at world w_k after k interventions to set the values of any $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$, then ϕ_V will belong to a causal model satisfying ($\mathcal{V}4$) at the world w_{k+1} where there is a $k + 1^{st}$ intervention to set the values of any $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$.*

Proof. By the inductive hypothesis, ϕ_V belongs to a causal model satisfying ($\mathcal{V}4$) at w_k. This means that there must exist a set of worlds $\mathfrak{F}_V$ which satisfies ($\mathfrak{F}2$) and which contains w_k. By ($f3$), an intervention setting the value of some $\mathbf{X} \subseteq \mathcal{U} \cup (\mathcal{V} - V)$ to $\mathbf{x}$ must take us to a world $w_{k+1} \in f(\mathbf{X} = \mathbf{x}, w_k)$. Since $w_k \in \mathfrak{F}_V$, ($\mathfrak{F}2$) guarantees that $f(\mathbf{X} = \mathbf{x}, w_k) \subseteq \mathfrak{F}_V$, so $w_{k+1} \in \mathfrak{F}_V$ as well. And, by assumption, $\mathfrak{F}_V \models V = \phi_V(\mathbf{PA}(V))$. So there is a $\mathfrak{F}_V \ni w_{k+1}$ such that $\mathfrak{F}_V$ satisfies ($\mathfrak{F}2$) and $\mathfrak{F}_V \models V = \phi_V(\mathbf{PA}(V))$. As ϕ_V was arbitrary, the same holds for every $V' \notin \mathbf{X}$. So, ($\mathcal{V}4$) will hold at w_{k+1}. So ϕ_V will belong to a causal model satisfying ($\mathcal{V}4$) at w_{k+1}. □

Setting $k = 0$ in the proof of the inductive step establishes the base case. □

Theorem 3. *Given* ($f1$)*, if a causal model* $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{E})$ *is correct according to* ($\mathcal{M}2$)*, then* $\mathfrak{F}_{\mathcal{M}} \stackrel{\text{def}}{=} \bigcap_{V \in \mathcal{V}} \mathfrak{F}_V$ *contains all and only allowed assignment of values to the variables* $V \in \mathcal{U} \cup \mathcal{V}$*, where an assignment is* allowed *just in case it is a solution to the equations in* $\mathcal{E}$*.*

Proof. The proof proceeds by induction on the rank of the variables in $\mathcal{V}$.

Base Case. $\mathfrak{F}_{\mathcal{M}}$ *contains all and only allowed assignment of values to the variables in* $\mathbf{Rank}(0)$*.*

Proof. For every $V \in \mathcal{V}$, $\mathfrak{F}_V$ contains $f(\mathcal{U} = \mathbf{u}, w)$, for every assignment $\mathbf{u}$ to $\mathcal{U}$, and every $w \in \mathfrak{F}_V$. So $\mathfrak{F}_{\mathcal{M}}$ contains $f(\mathcal{U} = \mathbf{u}, w)$, for every assignment $\mathbf{u}$ to $\mathcal{U}$ and every $w \in \mathfrak{F}_{\mathcal{M}}$. If $U \in \mathbf{Rank}(0)$, then U is exogenous, $U \in \mathcal{U}$. Every assignment of values to the exogenous variables is allowed. So $\mathfrak{F}_{\mathcal{M}}$ contains all and only allowed assignments to the variables in $\mathbf{Rank}(0)$. □

Inductive Step. *If* $\mathfrak{F}_{\mathcal{M}}$ *contains all and only allowed assignment of values to the variables in* $\mathbf{Rank}(0, 1, \ldots, k)$*, then it contains all and only allowed assignment of values to the variables in* $\mathbf{Rank}(0, 1, \ldots, k, k+1)$*.*

Proof. Take an arbitrary $V \in \mathbf{Rank}(k+1)$. Since $\mathbf{PA}(V) \subseteq \mathbf{Rank}(0, 1, \ldots, k)$, the inductive hypothesis gets us that every and only the allowed assignment of values to $\mathbf{PA}(V)$ are realized in $\mathfrak{F}_{\mathcal{M}}$. And because $\mathcal{M}$ is correct, $V_w = \phi(\mathbf{PA}(V)_w)$, for every $w \in \mathfrak{F}_V$. Since $\mathfrak{F}_{\mathcal{M}} \subseteq \mathfrak{F}_V$, this means that $V_w = \phi(\mathbf{PA}(V)_w)$ for every $w \in \mathfrak{F}_{\mathcal{M}}$ as well. So $\mathfrak{F}_{\mathcal{M}}$ contains all and only the allowed values of V. Since V was arbitrary, the above holds for every $V \in \mathbf{Rank}(k+1)$. □

□

Theorem 4. *Given* ($\mathcal{M}2$) *and* ($f3$)*, for any* $\mathbf{V} \subsetneq \mathcal{V}$*,*

$$\bigcap_{W \notin \mathbf{V}} \mathfrak{F}_W - \bigcup_{V \in \mathbf{V}} \mathfrak{F}_V$$

is non-empty and contains every assignment of values to the variables in $\mathbf{V} \cup \mathcal{U}$*.*

Proof. Take an arbitrary $\mathbf{V} \subsetneq \mathcal{V}$, an arbitrary assignment of values $\mathbf{v}$ to $\mathbf{V}$, an arbitrary $W \notin \mathbf{V}$, and an arbitrary assignment $\mathbf{u}$ to $\mathcal{U}$. Then, $\mathfrak{F}_W$ contains worlds at which $\mathbf{V} \cup \mathcal{U}$ is set to $\mathbf{v} \cup \mathbf{u}$ by an intervention, by ($\mathfrak{F}2$) and ($f3$). These worlds are not in $\bigcup_{V \in \mathbf{V}} \mathfrak{F}_V$, by the definition of an *intervention*. Since W, $\mathbf{V}$, $\mathbf{v}$, $\mathbf{u}$, and V were arbitrary, for every $W \notin \mathbf{V}$, every $\mathbf{u}$, and every $V \in \mathbf{V}$, there are worlds in $\mathfrak{F}_W$ which are not in $\mathfrak{F}_V$ and at which the value of $V \cup \mathcal{U}$ is set to any value $\mathbf{v} \cup \mathbf{u}$. Thus, $\bigcap_{W \notin \mathbf{V}} \mathfrak{F}_W - \bigcup_{V \in \mathbf{V}} \mathfrak{F}_V$ is non-empty and contains every assignment of values to the variables in $\mathbf{V} \cup \mathcal{U}$. □

References

Baumgartner, Michael. 2013. "A Regularity Theoretic Approach to Actual Causation." *Erkenntnis*, vol. 78 (1): 85–109. [2]

Bennett, Jonathan. 2003. *A Philosophical Guide to Conditionals*. Clarendon Press, Oxford. [18]

Briggs, Rachael. 2012. "Interventionist Counterfactuals." *Philosophical Studies*, vol. 160: 139–166. [2]

Cartwright, Nancy. 2009. "How to Do Things with Causes." apa *Proceedings and Addresses*, vol. 83 (2). [5]

Glymour, Clark, David Danks, Bruce Glymour, Frederick Eberhardt, Joseph Ramsey, Richard Scheines, Peter Spirtes, Choh Man Teng & Jiji Zhang. 2010. "Actual Causation: A Stone Soup Essay." *Synthese*, vol. 175: 169–192. [2]

Glynn, Luke. 2013. "Of Miracles and Interventions." *Erkenntnis*, vol. 78 (1): 43–64. [2]

Hall, Ned. 2007. "Structural Equations and Causation." *Philosophical Studies*, vol. 132 (1): 109–136. [2]

Halpern, Joseph Y. 2008. "Defaults and Normality in Causal Structures." *Proceedings of the Eleventh International Conference on Principles of Knowledge Representation and Reasoning*, 198–208. [2]

Halpern, Joseph Y. & Christopher Hitchcock. 2010. "Actual Causation and the Art of Modeling." In *Heuristics, Probability and Causality: A Tribute to Judea Pearl*, Rina Dechter, Hechtor Geffner & Joseph Y. Halpern, editors, 383–406. College Publications. [2]

—. forthcoming. "Graded Causation and Defaults." *The British Journal for the Philosophy of Science*. [2]

Halpern, Joseph Y. & Judea Pearl. 2001. "Causes and Explanations: A Structural-Model Approach. Part 1: Causes." In *Proceedings of the Seventeeth Conference on Uncertainty in Artificial Intelligence*, John Breese & Daphne Koller, editors, 194–202. Morgan Kaufman, San Francisco. [2]

—. 2005a. "Causes and Explanations: A Structural-Model Approach. Part 1: Causes." *The British Journal for the Philosophy of Science*, vol. 56: 843–887. [2]

—. 2005b. "Causes and Explanations: A Structural-Model Approach. Part 2: Explanations." *The British Journal for the Philosophy of Science*, vol. 56: 889–911. [2]

Handfield, Toby, Charles R. Twardy, Kevin B. Korb & Graham Oppy. 2008. "The Metaphysics of Causal Models: Where's the Biff?" *Erkenntnis*, vol. 68: 149–168. [2]

Hiddleston, Eric. 2005. "A Causal Theory of Counterfactuals." *Noûs*, vol. 39 (4): 632–657. [2]

Hitchcock, Christopher. 2001. "The Intransitivity of Causation Revealed in Equations and Graphs." *The Journal of Philosophy*, vol. 98 (6): 273–299. [2], [7]

—. 2007. "Prevention, Preemption, and the Principle of Sufficient Reason." *Philosophical Review*, vol. 116 (4): 495–532. [2]

Hitchcock, Christopher & Joshua Knobe. 2009. "Cause and Norm." *Journal of Philosophy*, vol. 106 (11): 587–612. [2]

Kim, Jaegwon. 1973. "Causes and Counterfactuals." *Journal of Philosophy*, vol. 70 (17): 570–572. [2]

Kment, Boris. 2006. "Counterfactuals and Explanation." *Mind*, vol. 115: 261–309. [18]

Lewis, David K. 1973. *Counterfactuals*. Blackwell Publishers, Malden, MA. [2], [7], [8]

—. 1979. "Counterfactual Dependence and Time's Arrow." *Noûs*, vol. 13 (4): 455–476. [2], [17]

—. 1986. "Events." In *Philosophical Papers*, vol. II, 241–269. Oxford University Press, New York. [2]

Livengood, Jonathan. 2013. "Actual Causation in Simple Voting Scenarios." *Noûs*, vol. 47 (2): 316–345. [2]

Mackie, John L. 1965. "Causes and Conditions." *American Philosophical Quarterly*, vol. 2 (4): 245–55. [12]

Maudlin, Tim. 2007. "A Modest Proposal Concerning Laws, Counterfactuals, and Explanations." In *The Metaphysics within Physics*, 5–49. Oxford University Press, Oxford. [17]

Menzies, Peter. 2004. "Causal Models, Token Causation, and Processes." *Philosophy of Science*, vol. 71 (5): 820–832. [2]

—. 2007. "Causation in Context." In *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, Huw Price & Richard Corry, editors, chap. 8, 191–223. Clarendon Press, Oxford. [2]

—. 2008. "Counterfactual Theories of Causation (The Stanford Encyclopedia of Philosophy)." http://plato.stanford.edu/entries/causation-counterfactual/. [2]

Paul, L. A. & Ned Hall. 2013. *Causation: A User's Guide*. Oxford University Press, Oxford. [2]

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, second edn. [2], [4]

—. 2009. "Causal Inference in Statistics: An Overview." *Statistics Surveys*, vol. 3: 96–146. [2]

Shulz, Katrin. 2011. ""If you'd wiggled A, then B would've changed": Causality and counterfactual conditionals." *Synthese*, vol. 179: 239–251. [2]

Spirtes, Peter, Clark Glymour & Richard Scheines. 2000. *Causation, Prediction, and Search*. The MIT Press, Cambridge, MA, second edn. [2]

Stalnaker, Robert C. 1968. "A Theory of Conditionals." In *Studies in Logical Theory*, N. Rescher, editor, chap. 4, 98–112. Oxford University Press, Oxford. [2], [7]

—. 1980. "A Defense of Conditional Excluded Middle." In *Ifs*, W. L. Harper, R. Stalnaker & G. Pearce, editors, 87–104. D. Reidel, Dordrecht. [7]

Weslake, Brad. forthcoming. "A Partial Theory of Actual Causation." *The British Journal for the Philosophy of Science*. [2]

Woodward, James. 1999. "Causal Interpretation in Systems of Equations." *Synthese*, vol. 121: 199–247. [2]

—. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, Oxford. [2]

Woodward, James & Christopher Hitchcock. 2003a. "Explanatory Generalizations, Part I: A Counterfactual Account." *Noûs*, vol. 37 (1): 1–24. [2]

—. 2003b. "Explanatory Generalizations, Part II: Plumbing Explanatory Depth." *Noûs*, vol. 37 (2): 181–199. [2]