THE REVIEW OF SYMBOLIC LOGIC
Volume 6, Number 4, December 2013

STRUCTURAL EQUATIONS AND BEYOND

FRANZ HUBER
Department of Philosophy, University of Toronto

Abstract. Recent accounts of actual causation are stated in terms of extended causal models. These extended causal models contain two elements representing two seemingly distinct modalities. The first element are structural equations which represent the "(causal) laws" or mechanisms of the model, just as ordinary causal models do. The second element are ranking functions which represent normality or typicality. The aim of this paper is to show that these two modalities can be unified. I do so by formulating two constraints under which extended causal models with their two modalities can be subsumed under so called "counterfactual models" which contain just one modality. These two constraints will be formally precise versions of Lewis' (1979) familiar "system of weights or priorities" governing overall similarity between possible worlds.

§1. Introduction. Recent accounts of actual causation are stated in terms of extended causal models. These extended causal models contain two elements representing two seemingly distinct modalities. The first element are structural equations which represent the "(causal) laws" or mechanisms of the model, just as ordinary causal models do. The second element are ranking functions which represent normality or typicality. The aim of this paper is to show that these two modalities can be unified. I do so by formulating two constraints under which extended causal models with their two modalities can be subsumed under so called "counterfactual models" which contain just one modality. These two constraints will be formally precise versions of Lewis' (1979) familiar "system of weights or priorities" governing overall similarity between possible worlds.

Here is my strategy in a bit more detail. Elsewhere I have introduced counterfactual models which contain one element representing one modality: objective ranking functions representing counterfactuality. In a first step I will generalize extended causal models by relaxing certain restrictions. If anything, this makes my task more difficult. In a second step I interpret the ranking functions in these generalized extended causal models objectively as in counterfactual models. In a third step I formulate two constraints on these generalized and objectively interpreted extended causal models. The first constraint relates structural equations and ranking functions. It is reminiscent of Lewis' (1979, 472) two conditions that "[i]t is of the first importance to avoid big, widespread, diverse violations of law" and that "[i]t is of the third importance to avoid even small, localized, simple violations of law." I show that extended causal models satisfying this first constraint can be subsumed under counterfactual models. The second constraint relates ranking functions and actuality. It is reminiscent of Lewis' (1979, 472) condition that "[i]t is of the second importance to maximize the spatiotemporal region throughout which perfect match of particular fact prevails." I show that extended causal models that satisfy this second constraint in addition to the first constraint can be subsumed under counterfactual models in a conservative

Received: January 12, 2013.

© Association for Symbolic Logic, 2013
doi:10.1017/S175502031300018X

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

way. By that I mean that all counterfactual claims as well as all claims about lawhood, causality, and actuality are conserved. Therefore, given these two constraints, there is only one modality that is needed to model actual causation and causality in general. That one modality is counterfactuality, which unifies the two modalities of "(causal) laws" or mechanisms and of normality or typicality that figure in extended causal models. This unification is achieved by a formally precise version of Lewis' (1979, 472) "system of weights or priorities."

This result is primarily a result about counterfactuals. However, it may impact the theory of causality in the following way. On the new picture of extended causal models, actual causation is the wrong concept to focus on, because it is a hybrid that involves two seemingly distinct modalities. On this view the concept to focus on is the notion of a "(causal) law" or mechanism as represented by a structural equation. In combination with normality or typicality, as well as what is actually the case, "(causal) laws" or mechanisms somehow give rise to actual causation. On a more traditional picture the concept to focus on is that of actual causation, which is to be analyzed in terms of counterfactuals (Lewis, 1973a, 1986a, 2000). I do not want to take sides on the issue of which causal notion to focus on. The issue I want to take sides on is how to represent counterfactuals. The traditional picture has come under attack because it has the wrong theory of counterfactuals (Lewis, 1973b, 1979). The new picture of extended causal models receives incredulous stares because it has an incomplete theory of counterfactuals. It reaches for a second modality in order to compensate for this incompleteness. However, in contrast to the first modality of "(causal) laws" or mechanisms the second modality of normality or typicality seems to be partly subjective. This flies in the face of the seemingly objective nature of actual causation. Hence the incredulous stares. The present account corrects the theory of counterfactuals underlying the traditional picture. It completes the theory of counterfactuals underlying the new picture by unifying the two modalities of the latter. Therefore, the present account provides the framework in terms of which a counterfactual theory of causality should be formulated, if one wants to defend such a theory.[1]

§2. Structural equations and defaults. The most promising framework for analyzing causation seems to be the structural equations approach (Spirtes *et al.*, 2000; Pearl, 2009, chap. 7; see also Halpern & Pearl, 2005a, 2005b; Hitchcock, 2001, 2007). While structural equations are primarily used for the analysis of causation, they are of independent interest for studying the logic of counterfactuals (see Briggs, 2012; Halpern, 2013). I will touch upon some issues in this connection below, but first we have to get started. The following definition is due to Halpern (2008).

$\mathcal{M} = (\mathcal{S}, \mathcal{F})$ is a *causal model* if and only if $\mathcal{S}$ is a signature and $\mathcal{F} = \{F_1, \ldots, F_n\}$ represents a set of n modifiable structural equations. $\mathcal{S} = (\mathcal{U}, \mathcal{V}, R)$ is a *signature* if and only if $\mathcal{U}$ is a finite set of *exogenous* variables, $\mathcal{V} = \{V_1, \ldots, V_n\}$ is a set of n *endogenous* variables disjoint from $\mathcal{U}$, and $R\colon \mathcal{U} \cup \mathcal{V} \to \mathcal{R}$ assigns each variable X in $\mathcal{U} \cup \mathcal{V}$ its *range* $R(X) \subseteq \mathcal{R}$. $\mathcal{W} = \times_{X \in \mathcal{U} \cup \mathcal{V}} R(X)$ is the set of *possible worlds*.

$\mathcal{F} = \{F_1, \ldots, F_n\}$ represents a set of n *modifiable structural equations* if and only if each F_i is a function from $\mathcal{W}_i = \times_{X \in \mathcal{U} \cup \mathcal{V} \setminus \{V_i\}} R(X)$ into the range $R(V_i)$ of the endogenous variable V_i. A causal model $\mathcal{M} = (\mathcal{S}, \mathcal{F})$ is *acyclic* if and only if there is

[1] For a quite different way of relating ranking functions and structural equations via causation see Spohn (2010).

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

no cycle $V_{i1}, \ldots, V_{im}, V_{i1}$ in $\mathcal{V}$ such that the value of $F_{i(j+1)}$ depends on $R\left(V_{ij}\right)$ for $j = 1, \ldots, m-1$, and the value of F_{i1} depends on $R\left(V_{im}\right)$. Dependence is functional dependence: F_i depends on $R\left(V_j\right)$ just in case there are $\vec{w}_i$ and $\vec{w}_i{}'$ in $\mathcal{W}_i = \times_{X \in \mathcal{U} \cup \mathcal{V} \setminus \{V_i\}} R(X)$ that differ only in the value from $R\left(V_j\right)$ such that $F_i\left(\vec{w}_i\right) \neq F_i\left(\vec{w}_i{}'\right)$.

Let $Pa\left(V_i\right)$ be the set of exogenous or endogenous variables X in $\mathcal{U} \cup \mathcal{V}$ such that F_i depends on $R(X)$. The members of $Pa\left(V_i\right)$ are called the *parents* of the endogenous variable V_i. Let $An\left(V_i\right)$ be the ancestral, or transitive closure, of $Pa\left(V_i\right)$, which is defined inductively as follows. $Pa\left(V_i\right) \subseteq An\left(V_i\right)$; and if $V \in An\left(V_i\right)$, then $Pa(V) \subseteq An\left(V_i\right)$. The members of $An\left(V_i\right)$ are called the *ancestors* of the endogenous variable V_i. They are the parents of V_i, $Pa\left(V_i\right)$, and the parents of all parents (but excluding V_i itself, unless the model is cyclic).

A *context* is a specification of the values of all exogenous variables and so can be formalized as a vector $\vec{u}$ in $R(\mathcal{U}) = \times_{U \in \mathcal{U}} R(U)$. A basic fact about causal models is that every acyclic causal model has a unique solution for any context. An acyclic causal model can be represented by a directed acyclic graph whose nodes are the exogenous and endogenous variables in $\mathcal{U} \cup \mathcal{V}$ and whose arrows point into each endogenous variable V_i from all of the latter's parents in $Pa\left(V_i\right)$.

The signature provides the framework or language of the model. It contains more structure than a set of possible worlds because there is a distinction between exogenous and endogenous variables. What may be even more important is the way one understands these variables. I understand them as *singular variables* and briefly want to explain why.

Philosophers such as Woodward (2003), following the lead of Spirtes *et al.* (2000) and Pearl (2009), are mainly interested in causal relevance between properties rather than actual causation between events (or, more cautiously, the relata of actual causation; see Paul, 2000). That is, they understand the variables in the generic way they are understood in science, especially those areas of science that rely on statistical methods, as assigning values to a population of individuals from which one can draw samples. For instance, the population may be the set of people at a certain age and in a certain geographical region, and the generic variable may assign values to these individuals—say, value i is assigned to an individual in that population if i mg ibuprofen are administered to that individual. With this generic understanding of the variables it might indeed be possible to test counterfactual claims of what would happen under certain interventions by "carry[ing] out the interventions described in the[...] antecedents and then check[ing] to see whether certain correlations hold" (Woodward, 2003, 72–73). For instance, it might indeed be possible to test the causal relevance claim that the administration of ibuprofen causes relief of pain by carrying out the intervention of administering a certain number of mg ibuprofen to some select subgroup of the population and then checking if pain is relieved in the members of that group.[2]

However, we cannot use generic variables if we want to construct a set of possible worlds in the way we have done above. In order to understand the Cartesian product of all possible values of all variables as a set of possible worlds we have to understand the variables in a singular sense. Otherwise, the resulting possibilities are not exclusive. For instance, the

[2] It is not entirely clear to me how Woodward (2003) can distinguish between the test of a counterfactual conditional, the test of an indicative conditional, and the test of a claim about conditional probabilities. How to empirically test or confirm counterfactuals on the account presented in section §4 is explained in Huber (ms 1).

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

variable may assign value i to a possible world if i mg ibuprofen are administered to me at noon on July 1, 2014, in that possible world. By moving from generic variables to singular variables we may lose some connection to science, but we get closer to philosophy. The reason is that now we can understand better the counterfactual claims implicit in a causal claim. Here is how.

If we can interpret the Cartesian product of all possible values of all variables as a set of possible worlds, then we can rely on a well-developed theory of counterfactuals. According to that theory a counterfactual conditional of the from 'if A were the case, then C would be the case' is true at a world if C is true in all worlds of a certain subset of the A-worlds. This understanding of counterfactuals is not obviously available if we work with generic variables. The reason is that it is not obvious how to construct possible worlds out of generic variables. And even if one has succeeded in constructing possible worlds out of generic variables, it is not obvious how to understand counterfactuals in the sense of this theory while still be able to test them in the way envisaged by Woodward (2003) and sketched above.

Another reason why it is important to understand the variables of the causal model as singular variables is that the restriction to acyclic causal models, which will be important later on, is only plausible for singular variables. For generic variables acyclicity is clearly false. A related point is made by Kistler (forthcoming).

Pearl (2009, chap. 10), Hitchcock (2001), Woodward (2003, sect. 2.7) and Halpern & Pearl (2005a) have provided increasingly sophisticated definitions of actual causation in terms of acyclic causal models (the particular way these authors formalize causal models differs in detail). However, Hiddleston (2005) presents two acyclic causal models where the "intuitively correct" causal judgments differ, even though the two models are isomorphic (two examples illustrating this point will be presented in the next section). As Halpern (2008) puts it: "there must be more to causality than just the structural equations." I refer to this claim as the *insufficiency thesis*: structural equations representing the "(causal) laws" or mechanisms of a model are insufficient for causality.

In order to solve this problem, Hall (2007) and Hitchcock (2007) distinguish between normal or *default* values and abnormal or *deviant* values of a variable. In Halpern (2008) and Halpern & Hitchcock (2010), these defaults are modeled in terms of ranking functions (Spohn, 1988). The latter are defined as follow. A function $\varrho : \mathcal{W} \to \mathbb{N}$ is a *ranking function* if and only if ϱ assigns rank 0 to at least one possible world w in $\mathcal{W}$. Usually ranking functions are interpreted epistemically as grades of disbelief, and then their defining clause is a consistency constraint saying that one should not disbelieve every possible world. A ranking function ϱ on the set of possible worlds $\mathcal{W}$ is extended to a function $\varrho^+ : \wp(\mathcal{W}) \to \mathbb{N} \cup \{\infty\}$ on the powerset of (the propositions over) $\mathcal{W}$, $\wp(\mathcal{W})$, by setting $\varrho^+(A) = \min\{\varrho(w) : w \in A \subseteq \mathcal{W}\}$ and $\varrho^+(\emptyset) = \infty$. I will abuse notation and write 'ϱ' instead of 'ϱ^+'.

$\mathcal{M} = (\mathcal{S}, \mathcal{F}, \varrho)$ is an *extended* (acyclic) causal model if and only if $(\mathcal{S}, \mathcal{F})$ is a(n) (acyclic) causal model and ϱ is a ranking function on $\mathcal{W}$. As suggested—unintentionally, but nevertheless appropriately—by Halpern (2008, sect. 4), the ranking function ϱ should be indexed to the set of contexts, because what is normal may vary from context to context. Thus, extended (acyclic) causal models really are of the form $\mathcal{M} = \left(\mathcal{S}, \mathcal{F}, (\varrho_{\vec{u}})_{\vec{u} \in R(\mathcal{U})}\right)$, where $R(\mathcal{U}) = \times_{U \in \mathcal{U}} R(U)$ is the set of all contexts or specifications of the values of all exogenous variables.

The definition of actual causation then runs as follows (Halpern & Hitchcock, 2010: sect. 3). $X_1 = x_1 \wedge \ldots \wedge X_k = x_k$, or simply: $\vec{X} = \vec{x}$, is an *actual cause* of ϕ in the

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

extended acyclic causal model $\mathcal{M} = \left(\mathcal{S}, \mathcal{F}, (\varrho_{\vec{u}})_{\vec{u} \in R(\mathcal{U})}\right)$ in context $\vec{u}$ if and only if:

1. $\vec{X} = \vec{x}$ and ϕ are true in $\mathcal{M}$ in $\vec{u}$.
2. There is a partition $\left\{\vec{Z}, \vec{W}\right\}$ of the endogenous variables $\mathcal{V}$ with $\vec{X} \subseteq \vec{Z}$, and there are vectors of values $\vec{x'}$ and $\vec{w}$ of $\vec{X}$ and $\vec{W}$, respectively, with $\varrho_{\vec{u}}\left(\vec{X} = \vec{x'} \wedge \vec{W} = \vec{w}\right) \leq \varrho_{\vec{u}}\left(w_{\vec{u}}\right)$ such that: if $\vec{Z} = \vec{z^*}$ is true in $\mathcal{M}$ in $\vec{u}$, then
   (a) $\vec{X} = \vec{x'} \wedge \vec{W} = \vec{w} \; \square\!\!\rightarrow_{SE} \neg\phi$ is true in $\mathcal{M}$ in $\vec{u}$; and
   (b) for all $\vec{W^-} \subseteq \vec{W}$ and all $\vec{Z^-} \subseteq \vec{Z}$: $\vec{X} = \vec{x} \wedge \vec{W^-} = \vec{w} \wedge \vec{Z^-} = \vec{z^*} \; \square\!\!\rightarrow_{SE} \phi$ is true in $\mathcal{M}$ in $\vec{u}$.
3. There is no proper subset $\vec{X^-}$ of $\vec{X}$ such that 1. and 2. hold for $\vec{X^-}$.

In order to understand this definition, we need to know the truth conditions for counterfactuals of the form $\vec{X} = \vec{x} \; \square\!\!\rightarrow_{SE} \phi$ in an extended acyclic causal model $\mathcal{M}$ in a context $\vec{u}$. It is these counterfactuals that are my main target. In what follows I ignore the use/mention distinction whenever possible so that the notation does not become even more cumbersome.

For an endogenous variable X in $\mathcal{V}$ and a value x in $R(X)$, $X = x$ is an atomic sentence. An atomic sentence $X = x$ is true in $\mathcal{M}$ in $\vec{u}$ just in case all solutions to the equations represented by $\mathcal{F}$ assign value x to the endogenous variable X when the exogenous variables are set to $\vec{u}$. Since we are restricting the discussion to extended acyclic causal models which have a unique solution in any given context, this means that $X = x$ is true in $\mathcal{M}$ in $\vec{u}$ if and only if x is the value of X in the unique solution to all equations in $\mathcal{M}$ in $\vec{u}$. The truth conditions for negations and conjunctions are given in the usual way.

A counterfactual $X_1 = x_1 \wedge \ldots \wedge X_k = x_k \; \square\!\!\rightarrow_{SE} \phi$, or simply: $\vec{X} = \vec{x} \; \square\!\!\rightarrow_{SE} \phi$, is true in $\mathcal{M}$ in $\vec{u}$ just in case ϕ is true in $\mathcal{M}_{\vec{X}=\vec{x}} = \left(\mathcal{S}_{\vec{X}}, \mathcal{F}^{\vec{X}=\vec{x}}\right)$ (but the same $\vec{u}$). The latter model results from $\mathcal{M}$ by replacing the equations for X_i by the equations $X_i = x_i$, $i = 1, \ldots, k$. Formally, this means two things (i–ii).

(i) The signature $\mathcal{S}$ is reduced to $\mathcal{S}_{\vec{X}} = \left(\mathcal{U}, \mathcal{V} \setminus \{X_1, \ldots, X_k\}, \mathcal{R}\,|_{\mathcal{U} \cup \mathcal{V} \setminus \{X_1, \ldots, X_k\}}\right)$, where $\mathcal{R}\,|_{\mathcal{U} \cup \mathcal{V} \setminus \{X_1, \ldots, X_k\}}$ is $\mathcal{R}$ with its domain restricted from $\mathcal{U} \cup \mathcal{V}$ to $\mathcal{U} \cup \mathcal{V} \setminus \{X_1, \ldots, X_k\}$.

(ii) $\mathcal{F}$ is reduced to $\mathcal{F}^{\vec{X}=\vec{x}}$ which results from $\mathcal{F}$ by deleting the functions F_{X_i} representing the equations for X_i and by changing the remaining functions F_Y in $\mathcal{F} \setminus \left\{F_{X_1}, \ldots, F_{X_k}\right\}$ as follows. First, restrict the domain of each F_Y from $\times_{X \in \mathcal{U} \cup \mathcal{V} \setminus \{Y\}} R(X)$ to $\times_{X \in \mathcal{U} \cup \mathcal{V} \setminus \{Y, X_1, \ldots, X_k\}} R(X)$. Second, replace F_Y by $F_Y^{\vec{X}=\vec{x}}$ which results from F_Y by setting $X_1, \ldots, X_k$ to $x_1, \ldots, x_k$, respectively.

While this definition is fairly complicated, the idea behind it is quite simple. In evaluating the counterfactual $\vec{X} = \vec{x} \; \square\!\!\rightarrow_{SE} \phi$ in model $\mathcal{M}$ in context $\vec{u}$, first validate the antecedent by deleting the equations for the endogenous variables $\vec{X}$ and setting their values to $\vec{x}$. In a second step set the exogenous variables to $\vec{u}$ and let the remaining equations determine the values of the remaining endogenous variables. In a third step check if the resulting solution yields the right value for ϕ.

The equations represent the "(causal) laws" or mechanisms of the model. It is important to stress the relativity to the model and that laws, as understood here, may fail to meet many of the traditional criteria for lawfulness (Woodward, 2003, chap. 6). The laws of the model can represent the workings of your fridge, the economics of the food market in the country I live in, the laws of gravitation of some planetary system, or Schrödinger's equation.

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

Several features of the formal language from above are worth being pointed out. First, all sentences are built up from endogenous variables. Second, Structural-Equations-counterfactuals or SE-counterfactuals cannot be iterated (embeddings can be defined, though, as shown by Halpern, 2013). Third, the antecedents of SE-counterfactuals are restricted to nonempty conjunctions of atomic sentences, although the consequents of SE-counterfactuals can be arbitrary Boolean combinations of atomic sentences.

As Halpern and Hitchcock (2010) note, the introduction of defaults makes the notion of actual causation doubly "subjective" (Halpern & Hitchcock, 2010, 384) or relative: judgments of actual causation depend on the choice of the exogenous and endogenous variables and on the choice of the default values for these variables. Let us look at their *FIRE example*.

Endogenous variable L takes on the value 1 if there is lightning, and 0 otherwise. Endogenous variable M takes on the value 1 if there is an arsonist dropping a lit match, and 0 otherwise. Endogenous variable F takes on the value 1 if there is a forest fire, and 0 otherwise. Furthermore exogenous variable (U_L, U_M) determines the values of L and M. The functions $F_L : ((i, j), m, f) \mapsto i$, $F_M : ((i, j), l, f) \mapsto j$, and $F_F : ((i, j)), l, m) \mapsto \max\{l, m\}$ describe the following equations:

- (U_L, U_M)
- $L = U_L$
- $M = U_M$
- $F = L \vee M$

According to Halpern and Hitchcock (2010), in the context where $U_L = 1$ and $U_M = 1$ so that there is lightning ($L = 1$) and there is an arsonist dropping a lit match ($M = 1$) and there is a forest fire ($F = 1$), the arsonist's dropping a lit match ($M = 1$) is an actual cause of the forest fire ($F = 1$). This is so, because:

1. $M = 1$ and $F = 1$ are true in $\mathcal{M}$ in $(u_L, u_M) = (1, 1)$.
2. For the partition $\{\{M, F\}, \{L\}\}$ and the values 0 and 0 of M and L we have $\varrho_{(1,1)}(M = 0 \wedge L = 0) \leq \varrho_{(1,1)}\left(w_{(1,1)}\right)$ and: $(M, F) = (1, 1)$ is true in $\mathcal{M}$ in $(1, 1)$ and
   (a) $M = 0 \wedge L = 0 \,\square\!\!\rightarrow_{SE} F \neq 1$ is true in $\mathcal{M}$ in $(1, 1)$, and so are
   (b) $M = 1 \,\square\!\!\rightarrow_{SE} F = 1$, $M = 1 \wedge F = 1 \,\square\!\!\rightarrow_{SE} F = 1$, $M = 1 \wedge L = 0 \,\square\!\!\rightarrow_{SE} F = 1$, $M = 1 \wedge L = 0 \wedge F = 1 \,\square\!\!\rightarrow_{SE} F = 1$.
3. There is no proper subset of $\{M\}$ such that 1. and 2. hold.

The relevant inequality for the ranking function $\varrho_{(1,1)}$ says that the most typical world where there is no lightning and no arsonist dropping a lit match is at least as typical as the actual world where there are lightning and an arsonist dropping a lit match and a forest fire. This equation holds (in the context where $U_L = 1$ and $U_M = 1$) for the following reason. It is more typical that there is no lightning ($L = 0$) than that there is lightning ($L = 1$). It is more typical that there is no arsonist dropping a lit match ($M = 0$) than that there is an arsonist dropping a lit match ($M = 1$). It is more typical that there is no forest fire ($F = 0$) than that there is a forest fire ($F = 1$).

In addition to this the structural equations seem to put a constraint on the ordering of normality or typicality. Even though it is more typical that there is no forest fire than that there is a forest fire, it is more typical that there is lightning and a forest fire than that there is lightning and no forest fire. Similarly, even though it is more typical that there is no forest fire than that there is a forest fire, it is more typical that there are an arsonist dropping a lit

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

match and a forest fire than that there is an arsonist dropping a lit match and no forest fire. Finally, even though it is more typical that there is no forest fire than that there is a forest fire, it is much more typical that there are lightning and an arsonist dropping a lit match and a forest fire than that there are lightning and an arsonist dropping a lit match, but there is no forest fire. And this is so no matter which context we are in.

More generally, the structural equations seem to put the following constraint on the ordering of normality or typicality. It seems that worlds which violate an equation are less typical than worlds that obey all equations (the latter are called "legal" in Glymour *et al.*, 2010). And it seems that worlds violating certain equations and then some are less typical than worlds violating only certain equations.

It is easy to see, though, that this constraint does not hold for equations such as $L = U_L$ and $M = U_M$, if only because we do not know what U_L and U_M stand for. However, it would be wrong to take this as a reason to reject the constraint that the structural equations seem to put on the ordering of normality or typicality. The fact that the constraint does not hold for equations such as $L = U_L$ and $M = U_M$ should rather be taken as a reason to reject the above model.

Let me explain. The only reason Halpern and Hitchcock (2010) include the "dummy variables" U_L and U_M and the "dummy equations" $U = U_L$ and $U = U_M$ is that they want to say that $L = 1$ and $M = 1$ are actual causes of $F = 1$, but cannot do so unless both L and M are endogenous variables. Besides that these variables and equations do no work and could be dropped if the artificial restriction were not in place that only endogenous variables can be causally efficacious. If that restriction were not in place, L and M would be the exogenous variables, and $F = L \vee M$ the only equation. Indeed, this is the model one would use in the framework of Hitchcock (2007).

§3. Generalizing causal models. *FIRE example, version 2:*

Let *exogenous* variable L take on the value 1 if there is lightning, and 0 otherwise. Let *exogenous* variable M take on the value 1 if there is an arsonist dropping a lit match, and 0 otherwise. Let endogenous variable F take on the value 1 if there is a forest fire, and 0 otherwise. The function $F_F : (l, m) \mapsto \max\{l, m\}$ describes the following equation:

- L
- M
- $F = L \vee M$

In this model it is true that worlds that violate an equation are less typical than worlds that obey all equations. My first proposal therefore is to relax the restriction in the (extended acyclic) causal models of Halpern (2008) and Halpern and Hitchcock (2010) and define an atomic sentence to be of the form $X = x$ for an exogenous or endogenous variable X in $\mathcal{U} \cup \mathcal{V}$ and a value x in $R(X)$. Then we do not have to include arbitrary exogenous variables to render L and M endogenous and thus be able to state counterfactual and causal claims with them.

For this to make sense we have to define the truth conditions for sentences in a slightly different way. An atomic sentence $X = x$ is true in $\mathcal{M}$ in $\vec{u}$ just in case all solutions to the equations represented by $\mathcal{F}$ when the exogenous variables are set to $\vec{u}$ assign value x to the exogenous or endogenous variable X. Since we keep restricting the discussion to acyclic models which have a unique solution in any context, this means that $X = x$ is true in $\mathcal{M}$ in $\vec{u}$ if and only if x is the value of X in the unique solution to all equations in $\mathcal{M}$ in $\vec{u}$. The truth conditions for negations and conjunctions are again given in the usual way.

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

A counterfactual $X_1 = x_1 \wedge \ldots \wedge X_k = x_k \; \square\!\!\rightarrow_{SE} \phi$, or simply: $\vec{X} = \vec{x} \; \square\!\!\rightarrow_{SE} \phi$, is true in $\mathcal{M}$ in $\vec{u}$ just in case ϕ is true in $\mathcal{M}_{\vec{X}=\vec{x}} = \left(\mathcal{S}_{\vec{X}}, \mathcal{F}^{\vec{X}=\vec{x}}\right)$ in $\vec{u}_{\vec{X}=\vec{x}}$. The latter model and context result from $\mathcal{M}$ and $\vec{u}$ by replacing the equations for X_i by the equations $X_i = x_i$, $i = 1, \ldots, k$. Formally, this means two things (i–ii). (i) The signature $\mathcal{S}$ is reduced to $\mathcal{S}_{\vec{X}} = \left(\mathcal{U}, \mathcal{V} \setminus \{X_1, \ldots, X_k\}, \mathcal{R}\,|_{\mathcal{U} \cup (\mathcal{V} \setminus \{X_1, \ldots, X_k\})}\right)$, where $\mathcal{R}\,|_{\mathcal{U} \cup (\mathcal{V} \setminus \{X_1, \ldots, X_k\})}$ is $\mathcal{R}$ with its domain restricted from the original $\mathcal{U} \cup \mathcal{V}$ to those variables $\mathcal{U} \cup (\mathcal{V} \setminus \{X_1, \ldots, X_k\})$ that remain after deleting the endogenous variables among $\{X_1, \ldots, X_k\}$.

(ii) $\mathcal{F}$ is reduced to $\mathcal{F}^{\vec{X}=\vec{x}}$ which results from $\mathcal{F}$ by deleting the functions F_{X_i} representing the equations for the endogenous X_i and by changing the remaining functions F_Y in $\mathcal{F} \setminus \left\{F_{X_1}, \ldots, F_{X_k}\right\}$ as follows. First, restrict the domain of each F_Y from $\times_{X \in \mathcal{U} \cup \mathcal{V} \setminus \{Y\}} R(X)$ to $\times_{X \in \mathcal{U} \cup (\mathcal{V} \setminus \{Y, X_1, \ldots, X_k\})} R(X)$. Second, replace F_Y by $F_Y^{\vec{X}=\vec{x}}$ which results from F_Y by setting $X_1, \ldots, X_k$ to $x_1, \ldots, x_k$, respectively.

The new context $\vec{u}_{\vec{X}=\vec{x}}$ results from the original context $\vec{u}$ as follows. First, set the values of the exogenous variables among $\{X_1, \ldots, X_k\}$ to $x_1, \ldots, x_k$, respectively. Second, leave the values of the other exogenous variables in $\mathcal{U} \setminus \{X_1, \ldots, X_k\}$ as they are in $\vec{u}$.

The definition of actual causation has to be changed slightly: in clause (2) we consider a partition of all variables, exogenous or endogenous, $\mathcal{U} \cup \mathcal{V}$ rather than a partition of the endogenous variables $\mathcal{V}$ only.

The *SURVIVAL example* (Halpern & Hitchcock, 2010, 400) explains why we need ranking functions in addition to the structural equations. Let exogenous variable A take on the value 1 if Assassin does not put in poison, and 0 otherwise. Let exogenous variable B take on the value 1 if Bodyguard puts in antidote, and 0 otherwise. Let endogenous variable S take on the value 1 if Victim survives, and 0 otherwise. The function $F_S : (a, b) \mapsto \max\{a, b\}$ describes the following equation:

- A
- B
- $S = A \vee B$

The structural equation for the *SURVIVAL example* is isomorphic to that for the *FIRE example, version 2*. However, people have different intuitions about the correct causal judgment for these two examples. In the *FIRE example, version 2* people say that the arsonist's dropping a lit match is an actual cause of the forest fire if there are lightning and an arsonist dropping a lit match (and a forest fire). In the *SURVIVAL example* people do not say that Bodyguard's putting in antidote is an actual cause of Victim's survival, if Bodyguard puts in antidote and Assassin does not put in poison (and Victim survives). This difference in people's intuitions about the correct causal judgment is explained by appeal to normality or typicality. While the structural equation for the *SURVIVAL example* is isomorphic to that for the *FIRE example, version 2*, the ordering of normality or typicality for the former differs from that of the latter in the following way.

It is more typical that Assassin does not put in poison ($A = 1$) than that Assassin puts in poison ($A = 0$). It is more typical that Bodyguard does not put in antidote ($B = 0$) than that Bodyguard puts in antidote ($B = 1$). It is more typical that Victim survives ($S = 1$) than that Victim does not survive ($S = 0$). In addition to this the structural equation seems to put a constraint on the ordering of normality or typicality. Even though it is more typical that Victim survives than that Victim does not survive, it is more typical that Assassin puts in poison and Bodyguard does not put in antidote and Victim does not survive than that Assassin puts in poison and Bodyguard does not put in antidote and Victim survives.

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

This helps us see why Bodyguard's putting in antidote is no actual cause of Victim's survival, if Bodyguard puts in antidote and Assassin does not put in poison and Victim survives, $A = 1$, $B = 1$, and $S = 1$.

1. $B = 1$ and $S = 1$ are true in $\mathcal{M}$ in $(a, b) = (1, 1)$; but
2. for the partition $\{\{B, S\}, \{A\}\}$ (and any other partition) there are no values b and a of B and A with $\varrho_{(1,1)}(B = b \wedge A = a) \leq \varrho_{(1,1)}\left(w_{(1,1)}\right)$ and: $(B, S) = (1, 1)$ is true in $\mathcal{M}$ in $(1, 1)$ and
   (a) $B = b \wedge A = a \; \square\!\!\rightarrow_{SE} S \neq 1$ is true in $\mathcal{M}$ in $(1, 1)$, and so are
   (b) $B = 1 \; \square\!\!\rightarrow_{SE} S = 1$, $B = 1 \wedge S = 1 \; \square\!\!\rightarrow_{SE} S = 1$, $B = 1 \wedge A = a \; \square\!\!\rightarrow_{SE} S = 1$, $B = 0 \wedge A = a \wedge S = 1 \; \square\!\!\rightarrow_{SE} S = 1$; and
3. there is no proper subset of $\{B\}$ such that 1. and 2. hold.

The reason is that the values b and a of B and A needed for $B = b \wedge A = a \; \square\!\!\rightarrow S \neq 1$ to come out true in $\mathcal{M}$ in $(1, 1)$ are 0 and 0. However, any world in which Bodyguard does not put in antidote and Assassin puts in poison, that is, where $B = 0 \wedge A = 0$ is true, is less typical than the actual world $w_{(1,1)}$ where Bodyguard puts in antidote and Assassin does not put in poison—or so Halpern and Hitchcock (2010, sect. 5) claim.

In fact, however, this is not true for the ranking function used by Halpern and Hitchcock (2010). Their ranking function assigns rank 1 to both the world that would be needed where Bodyguard does not put in antidote and Assassin puts in poison, as well as to the actual world where Bodyguard puts in antidote but Assassin does not put in poison. What is true, though, is that the world that would be needed where Bodyguard does not put in antidote and Assassin puts in poison is less typical than *the most typical world* where Assassin does not put in poison, viz. the world where Bodyguard does not put in antidote and Assassin does not put in poison.

We therefore have to slightly adjust the definition of actual causation (in the spirit of Hitchcock, 2007, who also refers to the actual value of $\vec{W}$ rather than the actual world) as follows: in condition (2), $\varrho_{\vec{u}}\left(\vec{X} = \vec{x'} \wedge \vec{W} = \vec{w}\right) \leq \varrho_{\vec{u}}\left(\vec{W} = \vec{w}\vec{u}\right)$, where $\vec{w}_{\vec{u}}$ is the actual value of $\vec{W}$ in model $\mathcal{M}$ in context $\vec{u}$.

For the sake of completeness I state the slightly revised definition of actual causation in extended acyclic causal models: $X_1 = x_1 \wedge \ldots \wedge X_k = x_k$, or simply: $\vec{X} = \vec{x}$, is an *actual cause* of ϕ in the extended acyclic causal model $\mathcal{M} = \left(\mathcal{S}, \mathcal{F}, (\varrho_{\vec{u}})_{\vec{u} \in R(\mathcal{U})}\right)$ in context $\vec{u}$ if and only if:

1. $\vec{X} = \vec{x}$ and ϕ are true in $\mathcal{M}$ in $\vec{u}$.
2. There is a partition $\left\{\vec{Z}, \vec{W}\right\}$ of all variables, exogenous or endogenous, $\mathcal{U} \cup \mathcal{V}$ with $\vec{X} \subseteq \vec{Z}$, and there are vectors of values $\vec{x'}$ and $\vec{w}$ of $\vec{X}$ and $\vec{W}$, respectively, with $\varrho_{\vec{u}}\left(\vec{X} = \vec{x'} \wedge \vec{W} = \vec{w}\right) \leq \varrho_{\vec{u}}\left(\vec{W} = \vec{w}_{\vec{u}}\right)$ such that: if $\vec{Z} = \vec{z^*}$ is true in $\mathcal{M}$ in $\vec{u}$, then
   (a) $\vec{X} = \vec{x'} \wedge \vec{W} = \vec{w} \; \square\!\!\rightarrow_{SE} \neg\phi$ is true in $\mathcal{M}$ in $\vec{u}$; and
   (b) for all $\vec{W}^- \subseteq \vec{W}$ and all $\vec{Z}^- \subseteq \vec{Z}$: $\vec{X} = \vec{x} \wedge \vec{W}^- = \vec{w} \wedge \vec{Z}^- = \vec{z^*} \; \square\!\!\rightarrow_{SE} \phi$ is true in $\mathcal{M}$ in $\vec{u}$.
3. There is no proper subset $\vec{X}^-$ of $\vec{X}$ such that 1. and 2. hold for $\vec{X}^-$.

This completes the first step of my argument as it was outlined in section §1. In a second step I now want to step back from Halpern and Hitchcock's (2010) interpretation of the ranking functions $\varrho_{\vec{u}}$. Instead of interpreting them solely in terms of normality or typicality,

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

I propose to interpret them as that notion—let us call it *(counterfactual) distance*—that gives truth conditions to counterfactuals. In the way I propose to interpret them, ranking functions represent a modality, the modality of counterfactuality, that is as objective as counterfactuals are. Therefore, I refer to them as *objective* ranking functions.

Counterfactual distance figures as a primitive on my account. It is the same notion that Stalnaker (1968) and Lewis (1973b, 1979) interpret in terms of overall similarity between possible worlds. While I do not think that overall similarity is an adequate interpretation of counterfactual distance (else I would not treat the latter as primitive), it may be helpful to the reader to think of objective ranking functions as formalizing overall similarity.

This formalization in terms of objective ranking functions differs slightly[3] from Stalnaker's (1968) formalization in terms of selection functions and from Lewis' (1973b) formalization in terms of a system of spheres. However, these slight differences do not affect the logic of counterfactuals in any way that is relevant for present purposes.[4]

Interim report: I have taken Halpern's (2008) notion of an extended (acyclic) causal model in terms of which Halpern and Hitchcock (2010) define actual causation. First I have slightly generalized these models by indexing the ranking functions in them to the contexts rather than assuming one fixed ranking function for all contexts. Then I have further generalized these models in the spirit of Hitchcock (2007) by dropping the restriction that only endogenous variables can be causally efficacious. Finally, after fixing a small bug in the definition of actual causation I have reinterpreted the ranking functions in these generalized extended (acyclic) causal models objectively as that notion which gives truth conditions to counterfactuals. This completes the first and second step of my argument as it was outlined in the Introduction. In the next three sections I will carry out the third step.

§4. Laws and counterfactuality. As stressed by Collins *et al.* (2004, 2ff) the logical properties of the counterfactual conditional do not suffice for a counterfactual theory of causation, if only because they do not exclude backtracking counterfactuals. This is why Lewis (1979) imposes four constraints on the similarity relation that is governing the logic of counterfactuals on his account, in addition to its defining features that fix the logical properties of the counterfactual conditional via the system **VC**.

I will impose two constraints as well.[5] The first constraint concerns the relation between structural equations and ranking functions and is a strong-dominance version of Lewis' (1979, 472) conditions that "[i]t is of the first importance to avoid big, widespread, diverse violations of law" and that "[i]t is of the third importance to avoid even small, localized, simple violations of law", except that it is relative to the causal model (see Menzies, 2004).

We start with some terminology relative to an extended acyclic causal model $\mathcal{M} = \big(\mathcal{S}, \mathcal{F}, (\varrho_{\vec{u}})_{\vec{u}\in R(\mathcal{U})}\big)$. Say that a world $w = (\vec{u}, v_1, \ldots, v_n)$ *violates the equation* for the

[3] The difference is that the limit assumption, which is rejected by Lewis (1973b), holds on the formalization in terms of objective ranking functions.

[4] One reason why I think that similarity is not an adequate interpretation of counterfactual distance is that the axiom(s) of strong centering (and weak centering) come out as (analytic) truths on this interpretation. I think that neither strong centering nor weak centering holds for counterfactuals. For criticism of similarity see Hàjek (ms). For criticism of weak centering and strong centering see Leitgeb (2012a, 2012b) and Menzies (2004, sect. 6).

[5] In stressing that it is an art to come up with an appropriate model for a given scenario or case Hitchcock (2007) states various constraints on appropriate models. His constraints concern the relation between the model and the case to be modeled. In contrast to these the constraints I impose are inherent to the model and independent of the case to be modeled.

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

endogenous variable V_i if and only if $v_i \neq F_i(\vec{u}, v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n)$. Let $\mathcal{V}^*(w) \subseteq \mathcal{V}$ be the set of endogenous variables V_i such that w violates the equation for V_i. Next say that a world w *weakly Halpern-dominates* a world w' if and only if for each endogenous variable $X \in \mathcal{V}^*(w) \setminus \mathcal{V}^*(w')$ there is an endogenous variable $X' \in \mathcal{V}^*(w') \setminus \mathcal{V}^*(w)$ such that $X' \in An(X)$. Finally, say that a world w *strongly Halpern-dominates* a world w' if and only if w weakly Halpern-dominates w', but w' does not weakly Halpern-dominate w (and so $\mathcal{V}^*(w') \setminus \mathcal{V}^*(w)$ is not empty).

Now we are in a position to formulate our first constraint. The idea is that worlds that violate certain equations and then some are (counterfactually) more distant than worlds that violate only certain equations. However, since a violation of the equation for an endogenous variable early on in the causal hierarchy affects everything causally downstream of that variable, a violation early on is worse—infinitely worse—than a violation later on. If we adopt the terminology of Lewis (1979), a violation of an equation early on in the causal hierarchy amounts to an infinitely bigger miracle than a violation of an equation later on. This is why the first constraint has to be stated in terms of ancestors.[6]

An extended acyclic causal model $\mathcal{M} = (\mathcal{S}, \mathcal{F}, (\varrho_{\vec{u}})_{\vec{u} \in R(\mathcal{U})})$ *respects the equations* if and only if the following holds for all worlds w and w' in $\mathcal{W}$: if w strongly Halpern-dominates w', then it holds for all contexts $\vec{u}$ in $R(\mathcal{U})$: $\varrho_{\vec{u}}(w) < \varrho_{\vec{u}}(w')$.[7]

The idea behind respect for the equations is quite simple. First associate with each world the set of endogenous variables whose equation the world violates. Then, when comparing two given worlds for (counterfactual) distance, ignore those endogenous variables whose equations are violated by both worlds. Finally check whether, among the remaining endogenous variables, for each endogenous variable whose equation is violated by the first world there is an endogenous variable that is causally upstream and whose equation is violated by the second world. In addition, check whether the converse is not true. In other words, check if any violation in the first world is compensated for by a violation in the second world that is worse, because it is further up in the causal hierarchy. In addition check if the converse is not true. If so, then the first world is (counterfactually) less distant, or closer, to any world than the second world. If the second world violates all the equations that are violated by the first world and then some we have the special case where, after ignoring the common violations, no violations in the first world are left.

We are approaching the summit of this paper. My aim is to show that by objectively interpreting the ranking functions in them, causal models respecting the equations can be subsumed under so called "counterfactual models" because the ranking functions thus interpreted yield all structural equations. In fact, counterfactual models give us more than causal models, because they define truth conditions for counterfactuals with arbitrary antecedents, something that is hard to come by in the structural equations approach (Briggs, 2012; Halpern, 2008, sect. 5). Furthermore, in counterfactual models counterfactuals may

[6] Woodward (2003, 141) can be read as endorsing our first constraint when he points to the following "important general difference between Lewis's scheme and the manipulationist picture. On the manipulationist account [...] "[l]ate" miracles, even numerous, are automatically preferred to "early" miracles, even if single. By contrast, in Lewis's theory, whether we [...] insert many late miracles [...] or whether instead we [insert some early miracle] [...] depends on whether [the effects] have many causes or just one. This sort of sensitivity leads to the insertion of miracles in what, intuitively, is the wrong place."

[7] The formulation of respect for the equations has undergone several changes. The present one is due to Joseph Y. Halpern, for whose many most helpful comments and suggestions I am very grateful.

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

not only be embedded but can also be iterated. Finally, the sentences in the formal language for counterfactual models are built up from exogenous and endogenous variables.

Here is the definition. $\mathcal{M}^* = \left(\mathcal{S}, (\varrho_w)_{w \in \mathcal{W}}\right)$ is a *counterfactual model* if and only if $\mathcal{S} = (\mathcal{U}, \mathcal{V}, R)$ is a signature and, for each world w in $\mathcal{W}$, $\varrho_w : \mathcal{W} \to \mathbb{N}$ is a ranking function on $\mathcal{W}$. Rather than indexing the ranking functions to the context $\vec{u}$ or the "legal" world $w_{\vec{u}}$ determined by that context, ranking functions are now indexed to the set of all possible worlds. The reason is that truth is a relation between sentences and possible worlds, and not between sentences and contexts (or between sentences and "legal" worlds). This makes it necessary to be explicit about the exogenous variables. From now on $\mathcal{U}$ is the set of m exogenous variables $\{U_1, \ldots, U_m\}$.

An atomic sentence $X_i = x$, $i = 1, \ldots, m + n$, is true in $\mathcal{M}^*$ in world $w \in \mathcal{W}$ if and only if $w \in \{(u_1, \ldots, u_m, v_1, \ldots, v_m) = (x_1, \ldots, x_{m+n}) \in \mathcal{W} : x_i = x\}$. Negations and conjunctions are defined as usual, and where ϕ and ψ are arbitrary sentences, the counterfactual $\phi \,\Box\!\!\rightarrow\, \psi$ is true in the counterfactual model $\mathcal{M}^*$ in the world w just in case all ϱ_w-minimal ϕ-worlds are ψ-worlds. The system **V** is sound and complete with respect to this semantics (Huber, ms 2).

In a causal model the structural equations are given and then used to define truth conditions for a limited set of counterfactual conditionals. In a counterfactual model the counterfactual conditionals are given via the ranking functions ϱ_w. Therefore, we have to say what it means for a structural equation represented by some function F to hold in a counterfactual model. For this we first restrict the functions F to those from $\mathcal{W}_i = \times_{X \in \mathcal{U} \cup \mathcal{V} \setminus \{V_i\}} R(X)$ into $R(V_i)$, for some endogenous variable V_i from $\mathcal{V}$. Call such a function *eligible* for V_i.

A function $F : \mathcal{W}_i \to R(V_i)$, which is eligible for V_i, *holds* in a counterfactual model $\mathcal{M}^*$ just in case, for every world w in $\mathcal{W}$, the following counterfactuals are all true in $\mathcal{M}^*$ in w: $\mathcal{U} \cup \vec{\mathcal{V}} \setminus \{V_i\} = \vec{w}_i \,\Box\!\!\rightarrow\, V_i = F_i(\vec{w}_i)$, where $\vec{w}_i$ is in $\mathcal{W}_i$. For an eligible function F to hold in a counterfactual model the above counterfactuals must be true in every world in that model. In contrast to counterfactuals in general, whose truth value is world-dependent, the structural equations hold world-independently. In this sense they are necessarily true. Therefore, talk of "(causal) *laws*" is appropriate.

My thesis is that the one modality of counterfactuality suffices for actual causation and causality in general. We have seen why to subscribe to the insufficiency thesis according to which "there must be more to causality than just the structural equations." We should not infer from the insufficiency thesis that we need a second modality. What we should infer from the insufficiency thesis is that the limited set of counterfactuals we get from the structural equations is not enough to represent the one relevant modality of counterfactuality.

To put it bluntly: structural equations are insufficient and unnecessary for causality. They are insufficient because they do not give us all counterfactuals, and because they do not give us all correct causal claims. They are unnecessary because we get them for free once we have moved beyond them, on to objective ranking functions. This is the content of the following theorem, which completes the first part of the third step of my argument as it was outlined in section §1. The second part of the third step follows in the next chapter.

THEOREM 4.1. *For each extended acyclic causal model* $\mathcal{M} = \left(\mathcal{S}, \mathcal{F}, (\varrho_{\vec{u}})_{\vec{u} \in R(\mathcal{U})}\right)$ *which respects the equations there is a counterfactual model* $\mathcal{M}^* = \left(\mathcal{S}, (\varrho_w)_{w \in \mathcal{W}}\right)$ *such that:*

SE F_i *holds in* $\mathcal{M}$ *iff* F_i *holds in* $\mathcal{M}^*$

D For all $\vec{u} \in R(\mathcal{U})$ *and all* $w \in \mathcal{W}$: $\varrho_{\vec{u}}(w) = \varrho_{w_{\vec{u}}}(w)$, *where* $w_{\vec{u}}$ *is the unique solution to all equations in* $\mathcal{M}$ *in* $\vec{u}$.

Proof. Appendix 7.1. □

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

§5. Counterfactuality and actuality. Let us look at the counterfactual models for our two examples if we use the ranking functions from Halpern and Hitchcock (2010) and evaluate counterfactuals in terms of them rather than the structural equations.

In the *SURVIVAL example* it is false in the actual context where Assassin does not put in poison and Bodyguard puts in antidote (and Victim survives) that Victim would not survive if Bodyguard did not put in antidote, $B = 0 \boxdot\!\!\rightarrow S \neq 1$. The reason is that one of the (counterfactually) least distant, or closest, worlds where Bodyguard does not put in antidote, viz. the world where Bodyguard does not put in antidote, Assassin does not put in poison, and Victim survives, is a world where Victim survives.

In the *FIRE example, version 2* it is true in the actual context where there are lightning and an arsonist dropping a lit match (and a forest fire) that there would be no forest fire if there were no arsonist dropping a lit match, $M = 0 \boxdot\!\!\rightarrow F \neq 1$. The reason is that all the (counterfactually) least distant, or closest, worlds where there is no arsonist dropping a lit match, viz. the world where there is no arsonist dropping a lit match, no lightning, and no forest fire, are also worlds where there is no forest fire.

This means that Theorem 4.1 is not enough. For it is *not* true that there would be no lightning if there were no arsonist dropping a lit match. On the contrary, even if there were no arsonist dropping a lit match, there would still be lightning, and hence there would still be a forest fire. This is also how the counterfactual $M = 0 \boxdot\!\!\rightarrow_{SE} F \neq 1$ is evaluated according to the structural models approach of Halpern and Hitchcock (2010).

This highlights the fact that the counterfactuals defined in terms of the structural equations of a causal model and the counterfactuals defined in terms of a counterfactual model may differ even if all and only the structural equations of the causal model hold in the counterfactual model. So far the only counterfactuals the two approaches agree on are those with maximally specific antecedents: $\mathcal{U} \cup \vec{\mathcal{V}} \setminus \{V_i\} = \vec{w}_i \boxdot\!\!\rightarrow_{(SE)} V_i = F_i(\vec{w}_i)$, where $\vec{w}_i$ is in $\mathcal{W}_i$. These are the necessarily true "(causal) *laws*" that are true in all worlds or contexts.

Defeat is not the appropriate reaction to this mismatch, though. What the mismatch shows is that we cannot define a counterfactual $\phi \boxdot\!\!\rightarrow \psi$ to be true in a world w in a model $\mathcal{M}$ if and only if all ϱ_w-minimal antecedent worlds are consequent worlds *and interpret* ϱ_w *solely in terms of normality or typicality*. For that means that $\phi \boxdot\!\!\rightarrow \psi$ is true if ϕ-worlds normally are ψ-worlds. And that is not right. More specifically, that is too weak.

The *LIGHTNING example* due to Christopher R. Hitchcock (personal correspondence) helps us see what is still missing to get the counterfactuals right. Let exogenous variable L take on the value 1 if there is lightning, and 0 otherwise. Let endogenous variable F take on the value 1 if there is a forest fire, and 0 otherwise. The function $F_F : l \mapsto f$ describes the following equation:

- L
- $F = L$

The equation says that there would be a forest fire if there were lightning. In the context where there is lightning, $L = 1$, we want to say that (even) if there were no forest fire there would (still) be lightning, $F = 0 \boxdot\!\!\rightarrow L = 1$. That is, we do not want our counterfactuals to backtrack. However, the world where there is lightning and no forest fire violates the equation, whereas the world where there is no lightning and no forest fire does not. Therefore, if all we require is respect for the equations we get the wrong result that, in the context where there is lightning, there would be no lightning if there were no forest fire, $F = 0 \boxdot\!\!\rightarrow L = 0$. In order to get the right result that there would (still)

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

be lightning, (even) if there were no forest fire, we additionally need to hold fixed what is actually true in the context of evaluation.

When we formulate the antecedent of a counterfactual we keep fixed as much of the actual context as is consistent with the antecedent. In the *LIGHTNING example* we keep fixed that there is lightning. The same is true of the *FIRE example, version 2*, where we also keep fixed that there is lightning. That is why it is true that if there were no arsonist dropping a lit match there would still be lightning, and hence there would still be a forest fire.[8]

Consequently, the second constraint concerns the relation between ranking functions and actuality. It is a strong-dominance version of Lewis' (1979, 472) condition that "[i]t is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails", except that it is relative to the causal model (again, see Menzies, 2004).

As before we start with some terminology relative to an extended acyclic causal model $\mathcal{M} = \left(\mathcal{S}, \mathcal{F}, (\varrho_{\vec{u}})_{\vec{u} \in R(\mathcal{U})}\right)$. Say that a world $w = (u_1, \ldots, u_m, \vec{v})$ *differs* from a world $w^+ = \left(u_1^+, \ldots, u_m^+, \vec{v}^+\right)$ in the value for the exogenous variable U_i if and only if $u_i \neq u_i^+$. Let $\mathcal{U}^*_{w^+}(w)$ be the set of exogenous variables for whose value w differs from w^+. Next say that a world w *weakly dominates* a world w' in terms of *focus on a world* w^+ if and only if $\mathcal{U}^*_{w^+}(w) \subseteq \mathcal{U}^*_{w^+}\left(w'\right)$. Finally say that a world w *strongly dominates* a world w' in terms of *focus on a world* w^+ if and only if w weakly dominates w' in terms of focus on w^+, but w' does not weakly dominate w in terms of focus on w^+.

Now we are in a position to formulate our second constraint. The idea is that worlds that differ from the actual world in the values of certain exogenous variables and then some are (counterfactually) more distant from the actual world than worlds that differ from the actual world only in the values for certain exogenous variables. In contrast to the global constraint of respect for the equations focus on actuality is a local constraint. This is so because what is actual varies from context to context. And that is why we now quantify over contexts at the beginning of the relevant clause.

An extended acyclic causal model $\mathcal{M} = \left(\mathcal{S}, \mathcal{F}, (\varrho_{\vec{u}})_{\vec{u} \in R(\mathcal{U})}\right)$ is *focused on actuality* if and only if the following holds for all contexts $\vec{u}$ in $R(\mathcal{U})$ and all worlds w and w' in $\mathcal{W}$: if w strongly dominates w' in terms of focus on the world $w_{\vec{u}}$, then: $\varrho_{\vec{u}}(w) < \varrho_{\vec{u}}\left(w'\right)$.

However, we cannot simply demand of an extended acyclic causal model that it satisfy focus on actuality in addition to respect for the equations. Focus on actuality is more important than respect for the equations, as the above example shows. For this reason, as well as to make sure that the two constraints do not conflict with each other, respect for the equations has to be restricted to worlds which agree on the values for the exogenous

[8] Note that we cannot hold fixed everything that is consistent with the antecedent. Consider the counterfactual 'If there were no lightning or no arsonist dropping a lit match, there would still be a forest fire.' This counterfactual has no truth-value on the structural models approach, even in its generalized form, because the antecedent is a disjunction. On our counterfactual models account this counterfactual does have a truth-value. Its antecedent is consistent with there being lightning. Its antecedent is also consistent with there being an arsonist dropping a lit match. However, its antecedent is not consistent with there jointly being lightning as well as an arsonist dropping a lit match. Thus we cannot hold fixed everything that is consistent with the antecedent.

Nor can we hold fixed only what is common to all antecedent-worlds. For then we would only consider worlds where there is neither lightning nor an arsonist dropping a lit match. The worlds we want to consider are such that either there is lightning but no arsonist dropping a lit match, or else there is no lightning but an arsonist dropping a lit match. For it is those worlds that hold fixed as much of the actual context as is consistent with the antecedent.

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

variables in $\mathcal{U}$. This means that we have a system of priorities rather than a system of weights (cf. Lewis, 1979, 472; Kroedel & Huber, forthcoming). Its content is that extended acyclic causal models have to be focused on actuality and subsequently respect the equations in the following sense. (Note that I have omitted this point in outlining my argument in the Introduction.)

An extended acyclic causal model $\mathcal{M} = \left(\mathcal{S}, \mathcal{F}, (\varrho_{\vec{u}})_{\vec{u} \in R(\mathcal{U})}\right)$ is *focused on actuality and subsequently respects the equations* if and only if $\mathcal{M}$ is focused on actuality and the following holds for all worlds w and w' in $\mathcal{W}$ *that agree on the values of the exogenous variables* $\mathcal{U}$: if w strongly Halpern-dominates w', then it holds for all contexts $\vec{u}$ in $R(\mathcal{U})$: $\varrho_{\vec{u}}(w) < \varrho_{\vec{u}}\left(w'\right)$.

For extended acyclic causal models which are focused on actuality and subsequently respect the equations the mismatch between the truth values of counterfactuals in the structural models approach and in the counterfactual models account disappears. This is the content of the following theorem.

THEOREM 5.1. *For each extended acyclic causal model* $\mathcal{M} = \left(\mathcal{S}, \mathcal{F}, (\varrho_{\vec{u}})_{\vec{u} \in R(\mathcal{U})}\right)$ *which is focused on actuality and subsequently respects the equations there is a counterfactual model* $\mathcal{M}^* = \left(\mathcal{S}, (\varrho_w)_{w \in \mathcal{W}}\right)$ *such that:*

C For all statements ϕ *in the language of the generalized version of Halpern and Hitchcock (2010) and all contexts* $\vec{u} \in R(\mathcal{U})$*:*
ϕ is true in $\mathcal{M}$ in $\vec{u}$ according to the structural equations approach iff ϕ is true in $\mathcal{M}^*$ in $w_{\vec{u}}$ according to the counterfactual models account.

Proof. Appendix 7.2. □

This almost completes the third step of my argument as it was outlined in section §1. There is one more twist to the story that will be topic of the next section when we put things together. However, before doing so I want to present a slightly different formulation of focus on actuality and subsequent respect for the equations that may be more accessible.

Respect for the equations is a global constraint on the endogenous variables and the structural equations governing them. Focus on actuality is a local constraint on the exogenous variables and their values in a given context. The distinction between exogenous and endogenous variables is relative to the model, and an exogenous variable may become endogenous if one refines a model by including further variables. Therefore, one may sometimes want to think of the exogenous variables as potentially endogenous, governed by structural equations that are temporarily set to a constant value for practical purposes, say, for the model to be simple.

From this point of view, it is natural to adopt the following terminology relative to an extended acyclic causal model $\mathcal{M} = \left(\mathcal{S}, \mathcal{F}, (\varrho_{\vec{u}})_{\vec{u} \in R(\mathcal{U})}\right)$. Say that the equation for an exogenous variable U_j in context $\vec{u} = (u_1, \ldots, u_m)$ is represented by the constant function $F_{u_j} : \left(u_1 \ldots, u_{j-1}, u_{j+1}, \ldots, u_m, \vec{v}\right) \mapsto u_j$ from $\times_{X \in \mathcal{U} \cup \mathcal{V} \setminus \{U_j\}} R(X)$ into $R\left(U_j\right)$. Next say that a world $w = (u_1, \ldots, u_m, \vec{v})$ *violates the equation for the exogenous variable* U_j *in context* $\vec{u}^+ = \left(u_1^+, \ldots, u_m^+\right)$ if and only if $u_j \neq F_{u_j^+}\left(u_1, \ldots, u_{j-1}, u_{j+1}, \ldots, u_m, \vec{v}\right) = u_j^+$. Let $\mathcal{X}_{\vec{u}}^*(w) \subseteq \mathcal{U} \cup \mathcal{V}$ be the set of exogenous or endogenous variables X such that w violates the equation for X (in context $\vec{u}$). Finally say that an extended acyclic causal model $\mathcal{M} = \left(\mathcal{S}, \mathcal{F}, (\varrho_{\vec{u}})_{\vec{u} \in R(\mathcal{U})}\right)$ is *respectful* if and only if the following holds for all contexts $\vec{u}$ and all worlds w and w' in $\mathcal{W}$: if for each exogenous or endogenous variable $X \in \mathcal{X}_{\vec{u}}^*(w) \setminus \mathcal{V}_{\vec{u}}^*\left(w'\right)$ there is an exogenous or endogenous variable $X' \in \mathcal{V}_{\vec{u}}^*\left(w'\right) \setminus \mathcal{V}_{\vec{u}}^*(w)$ such that $X' \in An(X)$, but the converse does not hold, then: $\varrho_{\vec{u}}(w) < \varrho_{\vec{u}}\left(w'\right)$.

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

Respectfulness is a mixed constraint on the exogenous and endogenous variables of a model, the values of the former in a given context, and the structural equations governing the latter in all contexts. It unifies the prioritized combination of focus on actuality and subsequent respect for the equations and allows us to state the following (strictly weaker) corollary of Theorem 5.1.

THEOREM 5.2. *For each extended acyclic causal model* $\mathcal{M} = \left(\mathcal{S}, \mathcal{F}, (\varrho_{\vec{u}})_{\vec{u} \in R(\mathcal{U})}\right)$ *which is respectful there is a counterfactual model* $\mathcal{M}^* = \left(\mathcal{S}, (\varrho_w)_{w \in \mathcal{W}}\right)$ *such that:*

C For all statements ϕ *in the language of the generalized version of Halpern and Hitchcock (2010) and all contexts* $\vec{u} \in R(\mathcal{U})$*:*
ϕ is true in $\mathcal{M}$ in $\vec{u}$ according to the structural equations approach iff ϕ is true in $\mathcal{M}^*$ in $w_{\vec{u}}$ according to the counterfactual models account.

§6. Beyond structural equations. It is time to put things together. Typicality and actuality can come apart. Actuality matters for counterfactuality. So, one might think, even counterfactual models are insufficient for causality. However, consider Spohn's (2006) account of causation. He starts out with a ranking function ϱ over a set of possible worlds which is generated by a set of singular variables in the same way as ours. Spohn interprets the ranking function ϱ subjectively in terms of grades of disbelief. He defines actual causation in terms of the conditional ranking function $\varrho(\cdot \mid H_w)$, where H_w is the complete history of the actual world w up to right before the effect, but excluding the cause (a temporal ordering relation over the variables allows Spohn to give a precise definition of this clause). So the seemingly objective nature of actual causation in this purely subjective account is partially captured by conditionalizing on what is actually the case.

This paves the way for the final move, suggested by Wolfgang Spohn (personal correspondence). Let us follow Halpern and Hitchcock (2010) in interpreting the *unconditional* ranking functions in terms of typicality. Furthermore, suppose our extended acyclic causal model $\mathcal{M} = \left(\mathcal{S}, \mathcal{F}, (\varrho_{\vec{u}})_{\vec{u} \in R(\mathcal{U})}\right)$ respects the equations. Typicality and actuality come apart in context $\vec{u}$ only if the unconditional ranking function $\varrho_{\vec{u}}$ and the conditional ranking function $\varrho_{\vec{u}}\left(\cdot \mid \vec{\mathcal{U}} = \vec{u}\right)$ differ for the rank assigned to some proposition $U_i = u_i$, for some exogenous variable U_i and some value u_i in $R(U_i)$. But nothing forces us to use the unconditional ranking function $\varrho_{\vec{u}}$ in evaluating counterfactuals in $\vec{u}$. We are free to use the conditional ranking function $\varrho_{\vec{u}}\left(\cdot \mid \vec{\mathcal{U}} = \vec{u}\right)$ to evaluate counterfactuals in $\vec{u}$.

Here is a restricted, but hopefully more comprehensible version of the main result detailed below. Suppose the model $\mathcal{M}$ with its family of unconditional ranking functions $(\varrho_{\vec{u}})_{\vec{u} \in R(\mathcal{U})}$ respects the equations. This implies that the model $\mathcal{M}_{\vec{\mathcal{U}}}$ with the family of *conditional* ranking functions $\left(\varrho_{\vec{u}}\left(\cdot \mid \vec{\mathcal{U}} = \vec{u}\right)\right)_{\vec{u} \in R(\mathcal{U})}$ is focused on actuality and subsequently respects the equations, provided we momentarily exclude the exogenous variables from the sentences of our language (this assumption will be dropped below). The conditional ranking functions give us the counterfactuals in the various contexts (or worlds, if we do not exclude the exogenous variables from the sentences of our language). If two scenarios or cases agree on the conditional ranking functions and the counterfactuals they represent, as is the case for the *FIRE example, version 2* and of the *SURVIVAL example*, they may still differ in the unconditional ranking functions they arise from and the defaults these latter represent. We do not need to introduce a second element in our model.

In a nutshell causality and counterfactuality interact in the following way. Typicality is represented by the unconditional or "prior" ranking functions. Counterfactuality includes

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

typicality, but goes beyond it by respecting the equations and, in the context of causality, by being focused on actuality (and subsequently respecting the equations). In the context of causality, counterfactuality is represented by the conditional or "posterior" ranking functions that arise from the unconditional ranking functions by conditionalizing on what is actually the case. Both unconditional as well as conditional ranking functions are to respect the equations. In addition to this the latter, but not the former, are to be focused on actuality. As a consequence, the latter, but not the former, do not represent typicality anymore, if, as may happen, typicality and actuality come apart. As Lewis might put it, typicality "is of little or no importance" (Lewis, 1979, 472).

Even though conditionalizing on what is actually the case may erase the traces of typicality, we can still refer back to the unconditional roots. This is exactly what we do if we adopt Halpern and Hitchcock's (2010) definition of actual causation. In the relevant clause (2) we use the unconditional ranking function to determine the default values of the variables, whereas we use the conditional ranking function to determine the truth values of the counterfactuals. Halpern and Hitchcock (2010) use two different formalisms, viz. structural equations and ranking functions, to represent the "(causal) laws" and typicality, respectively. I use just one formalism, viz. objective ranking functions, that, due to its conditional nature, is sufficiently rich to capture both of these dimensions of counterfactuality.[9]

Things are more complicated if we allow for exogenous variables in the sentences of our language. Then the following more general move has to be made. Take $\mathcal{M}$ with its family of unconditional ranking functions $(\varrho_{\vec{u}})_{\vec{u}\in R(\mathcal{U})}$. Instead of *strictly* conditionalizing every $\varrho_{\vec{u}}$ on $\vec{\mathcal{U}} = \vec{u}$ to obtain the model $\mathcal{M}_{\vec{\mathcal{U}}}$ with its family of conditional ranking functions $\left(\varrho_{\vec{u}}\left(\cdot \mid \vec{\mathcal{U}} = \vec{u}\right)\right)_{\vec{u}\in R(\mathcal{U})}$, merely *Shenoy* conditionalize every $\varrho_{\vec{u}}$ on $\vec{\mathcal{U}} = \vec{u}$ by an appropriately chosen number max to obtain the model $\mathcal{M}^{max}_{\vec{\mathcal{U}}}$ with its family of "Shenoy shifted" ranking functions $\left(\varrho_{\vec{u}}\left(\cdot \uparrow \vec{\mathcal{U}} = \vec{u}\right)\right)_{\vec{u}\in R(\mathcal{U})}$.

Shenoy conditionalization is defined as follows. If $\varrho : \wp(\mathcal{W}) \to \mathbb{N} \cup \{\infty\}$ is the unconditional ranking function on the powerset over $\mathcal{W}$, $\wp(\mathcal{W})$, then the result of Shenoy conditionalizing ϱ on the proposition A from $\wp(\mathcal{W})$ by rank $k \in N \cup \{\infty\}$, $\varrho_{A\uparrow k}$, is defined as follows. For each B from $\wp(\mathcal{W})$,

$$\varrho_{A\uparrow k}(B) = \min\left\{\varrho(B \cap A) + 0 - min, \varrho\left(B \cap \overline{A}\right) + k - min\right\},$$

where $min = \min\left\{0 + \varrho(A), k + \varrho\left(\overline{A}\right)\right\}$ is a normalization parameter which depends on the ranking function ϱ which is to be updated, the partition $\left\{A, \overline{A}\right\}$, and the input parameters $\{0, k\}$ by which the elements $A, \overline{A}$ of the partition are shifted. The effect of normalizing by min is that at least one possible world is assigned rank 0 rather than rank min. Strictly conditionalizing ϱ on A results in the same ranking function as Shenoy conditionalizing ϱ on A by ∞ so that $\mathcal{M}_{\vec{\mathcal{U}}} = \mathcal{M}^{\infty}_{\vec{\mathcal{U}}}$. Shenoy conditionalization was introduced by Shenoy (1991). It is the ranktheoretic counterpart to probability theory's Field conditionalization (Field, 1978).

The family of Shenoy shifted ranking functions in terms of which we evaluate counterfactuals results from the original family of unconditional ranking functions by a series of m Shenoy shifts, one for each exogenous variable U_j. We start with $\varrho_{\vec{u}} =: \varrho_0$ from the

[9] It should be noted that this story cannot be told on an account of counterfactuals such as Lewis' (1973b) or Stalnaker's (1968) because these accounts lack the operation of conditionalisation: there are no such things as a conditional sphere of similarity or conditional selection functions.

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

family $(\varrho_{\vec{u}})_{\vec{u}\in R(\mathcal{U})}$, which, following Halpern and Hitchcock (2010), we interpret in terms of typicality. What we need to do is to Shenoy conditionalize on what is actually the case in context $\vec{u} = (u_1, \ldots, u_m)$. We do this by first Shenoy conditionalizing ϱ_0 on $U_1 = u_1$ by $max = \max\{\varrho_{\vec{u}}(w) : w \in \mathcal{W}\} + 1$, which is sufficiently large but finite. This has two effects. First, all worlds that differ from the actual world $w_{\vec{u}}$ in the value for the exogenous variable U_1 are shifted upwards by $max - min_1$, where min_1 depends, among others, on ϱ_0. Second, all worlds that agree with the actual world $w_{\vec{u}}$ on the value for the exogenous variable U_1 are shifted downwards by min_1 (and so at least one of those latter worlds is assigned rank 0). The result is $\varrho_{0,U_1=u_1\uparrow max} =: \varrho_1$.

We continue by Shenoy conditionalizing ϱ_1 on $U_2 = u_2$ by max to obtain $\varrho_{1,U_2=u_2\uparrow max} =: \varrho_2$ and so on until we finally arrive at $\varrho_{m-1,U_m=u_m\uparrow max} = \varrho_m =: \varrho_{\vec{u}}(\cdot \uparrow \vec{\mathcal{U}} = \vec{u})$. ϱ_m differs from the original ϱ_0 in that worlds that differ from the actual world $w_{\vec{u}}$ in the value for exactly k exogenous variables have been shifted upwards or further away by $k \cdot max$, modulo normalization.

The first thing this means is that the model with the Shenoy shifted ranking functions ϱ_ms instead of the unconditional ranking functions $\varrho_{\vec{u}}$s is focused on actuality. By the choice of max every world that differs from the actual world in the value of some exogenous variable now has a higher rank than any world that agrees with the actual world in the value of all exogenous variables. More generally, every world that dominates another world in terms of focus on the actual world is assigned a smaller rank than the dominated world.[10]

The second thing this mean is that, in the Shenoy shifted ranking functions ϱ_m, the relative position of two worlds that agree on the values for the exogenous variables is the same as it is in the unconditional ranking functions $\varrho_{\vec{u}}$ (the two worlds are always shifted together). Therefore, the model with the Shenoy shifted ranking functions ϱ_m instead of the unconditional ranking functions $\varrho_{\vec{u}}$ still respects the equations for those worlds that agree on the values for the exogenous variables.

Therefore, the model $\mathcal{M}^{max}_{\vec{\mathcal{U}}}$, call it the *appropriate Shenoy shift* of $\mathcal{M}$ on $\vec{\mathcal{U}}$, with its family of Shenoy shifted ranking functions $\left(\varrho_{\vec{u}}(\cdot \uparrow \vec{\mathcal{U}} = \vec{u})\right)_{\vec{u}\in R(\mathcal{U})}$ is focused on actuality and subsequently respects the equations, if the extended acyclic causal model $\mathcal{M}$ with its family of unconditional ranking functions $(\varrho_{\vec{u}})_{\vec{u}\in R(\mathcal{U})}$ respects the equations.

In the same way, we can form the appropriate Shenoy shift $\mathcal{M}^{*max}_{\vec{\mathcal{U}}}$ of a counterfactual model $\mathcal{M}^*$. As the proofs of Theorems 4.1 and 5.1 make clear, $\mathcal{M}^{*max}_{\vec{\mathcal{U}}}$ constructed in this way is one of the counterfactual models $\mathcal{M}^{max*}_{\vec{\mathcal{U}}}$ that exist for each extended acyclic causal model $\mathcal{M}^{max}_{\vec{\mathcal{U}}}$ which is focused on actuality and subsequently respects the equations. Thus we arrive at

THEOREM 6.1. *For each extended acyclic causal model* $\mathcal{M} = \left(\mathcal{S}, \mathcal{F}, (\varrho_{\vec{u}})_{\vec{u}\in R(\mathcal{U})}\right)$ *which respects the equations and its appropriate Shenoy shift* $\mathcal{M}^{max}_{\vec{\mathcal{U}}}$ *which is focused on actuality and subsequently respects the equations there is a counterfactual model* $\mathcal{M}^* = \left(\mathcal{S}, (\varrho_w)_{w\in\mathcal{W}}\right)$ *and its appropriate Shenoy shift* $\mathcal{M}^{*max}_{\vec{\mathcal{U}}}$ *such that:*

SE F_i *holds in* $\mathcal{M}$ *iff* F_i *holds in* $\mathcal{M}^{max}_{\vec{\mathcal{U}}}$ *iff* F_i *holds in* $\mathcal{M}^*$ *iff* F_i *holds in* $\mathcal{M}^{*max}_{\vec{\mathcal{U}}}$

[10] Shenoy conditionalizing just once on the conjunction $\vec{\mathcal{U}} = \vec{u}$ by max does not guarantee that the resulting model is focused on actuality, because in that case all that matters is whether a world differs from the actual world in the value for at least one or no exogenous variable.

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

D *For all* $\vec{u} \in R(\mathcal{U})$ *and all* $w \in \mathcal{W}$: $\varrho_{\vec{u}}(w) = \varrho_{w_{\vec{u}}}(w)$, *where* $w_{\vec{u}}$ *is the unique solution to all the equations of* $\mathcal{M}$ *in* $\vec{u}$.

C *For all statements* ϕ *in the language of the generalized version of Halpern and Hitchcock (2010) and all contexts* $\vec{u} \in R(\mathcal{U})$:
ϕ *is true in* $\mathcal{M}$ *in* $\vec{u}$ *according to the structural equations approach iff*
ϕ *is true in* $\mathcal{M}^{max}_{\vec{\mathcal{U}}}$ *in* $\vec{u}$ *according to the structural equations approach iff*
ϕ *is true in* $\mathcal{M}^{*max}_{\vec{\mathcal{U}}}$ *in* $w_{\vec{u}}$ *according to the counterfactual models account.*

Theorem 6.1 shows that we can do everything with objective ranking functions that we can do with structural equations together with normality or typicality, and more. It does not show that we can do everything. The reason I am belaboring the obvious is that it may well be that someone comes up with examples which are modeled by isomorphic counterfactual models, and of which it is claimed that the "intuitively correct" causal judgments differ (see, however, Glymour *et al.*, 2010).

In the same way, one may come up with examples which are modeled by extended acyclic causal models in which, "intuitively", respect for the equations does not hold. The following one due to Christopher R. Hitchcock (personal correspondence) might be a case in point. I think it is not, because counterfactuality trumps typicality in the sense that the most typical $A \wedge C$-worlds are more typical than the most typical $A \wedge \neg C$-worlds if $A \boxdot\!\!\rightarrow C$ is true.

Here is Hitchcock's *VICTIM example*. Let exogenous variable A take on the value 1 if Assassin shoots, and 0 otherwise. Let endogenous variable B take on the value 1 if Backup shoots, and 0 otherwise. Let endogenous variable V take on the value 1 if Victim dies, and 0 otherwise. The functions $F_B : a \mapsto 1 - a$ and $F_V : (a, b) \mapsto \max\{a, b\}$ describe the following equations:

- A
- $B = 1 - A$
- $V = A \vee B$

In every context, it is less typical for Assassin as well as Backup to shoot than not to shoot, and for Victim to die than not to die.

The first equation implies that Backup would shoot if Assassin did not shoot. Respect for the equations forces us to say that the world where Assassin does not shoot, Backup does not shoot, and Victim does not die is less typical than the world where Assassin does not shoot, Backup shoots, and Victim does not die. The reason is that the latter world strongly Halpern-dominates the former world: the latter world violates the equation for V (an no other equation), the former world violates the equation for B (and no other equation), and $B \in An(V)$, but $V \notin An(B)$. For a similar reason we have to say that the world where Assassin does not shoot and Backup does not shoot and Victim dies is less typical than the world where Assassin does not shoot, Backup shoots, and Victim dies.

Therefore, we *must* say that it is more typical that Assassin does not shoot and Backup shoots than that Assassin does not shoot and Backup does not shoot. I think this is correct because it conforms with the counterfactual that Backup would shoot if Assassin did not shoot.

Another example is the *PEN example* mentioned in Halpern and Hitchcock (forthcoming). Let endogenous variable PS take on the value 1 if Professor Smith takes a pen, and 0 otherwise. Let endogenous variable CP take on the value 1 if the department chair institutes a policy forbidding faculty members from taking pens, and 0 otherwise. Let

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

exogenous variable PO take on the value 1 if a problem occurs, and 0 otherwise. The function $F : c \mapsto c$ describes the following equation:

- CP
- PS
- $PO = PS$

It is more typical for Professor Smith to not take a pen than to take a pen. In the context where the department chair institutes a policy forbidding faculty members from taking pens, $CP = 1$, and where Professor Smith takes a pen, $PS = 1$, it is true that Professor Smith would (still) take a pen (even) if the department chair instituted a policy forbidding faculty members from taking pens, $CP = 1 \boxdot\!\!\rightarrow PS = 1$. So far so good.

Here is the important point. Halpern and Hitchcock (forthcoming) claim that it is more "typical" that the department chair institutes a policy forbidding faculty members from taking pens and Professor Smith does not take a pen than that the department chair institutes a policy forbidding faculty members from taking pens and Professor Smith takes a pen. The reason is that Professor Smith violates a norm when he takes a pen in the context where the department chair institutes a policy forbidding faculty members from taking pens. This norm, or rather its violation, is claimed to have an impact on what is typical in that context.

However, what Halpern and Hitchcock (forthcoming) call "typicality" involves a *deontic* modality. The *PEN example* contains the conditional obligation that Professor Smith should not take a pen given that the department chair institutes a policy forbidding faculty members from taking pens, $Ought\,(PS = 0 \mid CP = 1)$. And while I hold the view that typicality or normality respects for the equations, I do not hold the view that deontic modalities do. Quite the opposite is the case. Given that the department chair institutes a policy forbidding faculty members from taking pens, Professor Smith should not, but (still) would, take a pen. This, I submit, implies that is *less* typical that the department chair institutes a policy forbidding faculty members from taking pens and Professor Smith does not take a pen than that the department chair institutes a policy forbidding faculty members from taking pens and Professor Smith takes a pen.

§7. Appendix

7.1. Proof of Theorem 1

THEOREM 7.1. *For each extended acyclic causal model* $\mathcal{M} = \left(\mathcal{S}, \mathcal{F}, (\varrho_{\vec{u}})_{\vec{u} \in R(\mathcal{U})}\right)$ *which respects the equations there is a counterfactual model* $\mathcal{M}^* = \left(\mathcal{S}, (\varrho_w)_{w \in \mathcal{W}}\right)$ *such that:*

SE F_i *holds in* $\mathcal{M}$ *iff* F_i *holds in* $\mathcal{M}^*$

D For all $\vec{u} \in R(\mathcal{U})$ *and all* $w \in \mathcal{W}$: $\varrho_{\vec{u}}(w) = \varrho_{w_{\vec{u}}}(w)$, *where* $w_{\vec{u}}$ *is the unique solution to all equations in* $\mathcal{M}$ *in* $\vec{u}$.

Proof. Let $\mathcal{M} = \left(\mathcal{S}, \mathcal{F}, (\varrho_{\vec{u}})_{\vec{u} \in R(\mathcal{U})}\right)$ be an extended acyclic causal model which respects the equations. I will construct a counterfactual model $\mathcal{M}^* = \left(\mathcal{S}, (\varrho_w)_{w \in \mathcal{W}}\right)$ with the appropriate features. Take $\mathcal{S}$ from $\mathcal{M}$.

For each context $\vec{u} \in R(\mathcal{U})$ the equations in $\mathcal{F}$ determine a unique "legal" world $w_{\vec{u}} \in \mathcal{W}$. $\mathcal{W}_0 = \{w_{\vec{u}} \in \mathcal{W} : \vec{u} \in R(\mathcal{U})\}$ is the set of all "legal" worlds, that is, the set of all worlds that satisfy all equations. For $w_{\vec{u}} \in \mathcal{W}_0$ we define $\varrho_{w_{\vec{u}}}(w) = \varrho_{\vec{u}}(w)$ for all $w \in \mathcal{W}$. For the "illegal" worlds $w \in \mathcal{W} \setminus \mathcal{W}_0$ which violate at least one equation we let the ranking

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

functions ϱ_w copy an arbitrary ranking function $\varrho_{w_{\vec{u}}}$, $w_{\vec{u}} \in \mathcal{W}_0$. The counterfactual model $\mathcal{M}^*$ constructed in this way satisfies D. It remains to be shown that it also satisfies SE.

Let F_i represent the equation for V_i, $i = 1, \ldots, n$. Obviously F_i is eligible for V_i. We have to show that F_i holds in $\mathcal{M}^*$. This means we have to show for every world $w \in \mathcal{W}$ that the following counterfactuals are all true in $\mathcal{M}^*$ in w: $\vec{\mathcal{U} \cup \mathcal{V} \setminus \{V_i\}} = \vec{w}_i \boxdot\!\!\rightarrow V_i = F_i(\vec{w}_i)$, where $\vec{w}_i \in \mathcal{W}_i = \times_{X \in \vec{\mathcal{U} \cup \mathcal{V} \setminus \{V_i\}}} R(X)$. Since the ϱ_ws for the "illegal" worlds $w \in \mathcal{W} \setminus \mathcal{W}_0$ copy some $\varrho_{w_{\vec{u}}}$, for a "legal" world $w_{\vec{u}} \in \mathcal{W}_0$, it suffices to show that this holds for every "legal" world $w_{\vec{u}} \in \mathcal{W}_0$.

Each antecedent of the form $\vec{\mathcal{U} \cup \mathcal{V} \setminus \{V_i\}} = \vec{w}_i$, for $\vec{w}_i \in \mathcal{W}_i$, is true in the set of worlds $\{(\vec{w}_i, v_i) : v_i \in R(V_i)\}$. There is exactly one $v_i^* \in R(V_i)$, viz. the value F_i assigns to $\vec{w}_i$, such that $(\vec{w}_i, v_i^*)$ does not violate the equation for V_i. For all other $v_i \in R(V_i)$ the resulting world $(\vec{w}_i, v_i)$ violates the equation for the endogenous variable V_i. Hence $V_i \in \mathcal{V}^*(\vec{w}_i, v_i) \setminus \mathcal{V}^*(\vec{w}_i, v_i^*)$ for all $v_i \neq v_i^*$. Furthermore, $(\vec{w}_i, v_i^*)$ and $(\vec{w}_i, v_i)$ agree on the values of all variables other than V_i.

Suppose $X \in \mathcal{V}^*(\vec{w}_i, v_i^*) \setminus \mathcal{V}^*(\vec{w}_i, v_i)$ for an arbitrary $v_i \neq v_i^*$. Since $(\vec{w}_i, v_i^*)$ and $(\vec{w}_i, v_i)$ agree on the value of X, and since, by assumption, $(\vec{w}_i, v_i)$ does not violate the equation for X, there must be an exogenous or endogenous variable Y such that $Y \in An(X)$ and $(\vec{w}_i, v_i^*)$ and $(\vec{w}_i, v_i)$ do not agree on the value of Y. Since $(\vec{w}_i, v_i^*)$ and $(\vec{w}_i, v_i)$ agree on the values of all variables other than V_i, this variable Y must be V_i. That is, if $X \in \mathcal{V}^*(\vec{w}_i, v_i^*) \setminus \mathcal{V}^*(\vec{w}_i, v_i)$, then $V_i \in An(X)$. Since $V_i \in \mathcal{V}^*(\vec{w}_i, v_i) \setminus \mathcal{V}^*(\vec{w}_i, v_i^*)$ for all $v_i \neq v_i^*$, this means that $(\vec{w}_i, v_i^*)$ weakly Halpern-dominates $(\vec{w}_i, v_i)$. Since, in acyclic causal models, $X \notin An(V_i)$ if $V_i \in An(X)$, and since $V_i \in \mathcal{V}^*(\vec{w}_i, v_i) \setminus \mathcal{V}^*(\vec{w}_i, v_i^*)$, $(\vec{w}_i, v_i)$ does not weakly Halpern-dominate $(\vec{w}_i, v_i^*)$.

Respect for the equations implies that $\varrho_{\vec{u}}((\vec{w}_i, v_i^*)) < \varrho_{\vec{u}}((\vec{w}_i, v_i))$ for all $v_i \neq v_i^*$. Since $V_i = F_i(\vec{w}_i)$ is true in $(\vec{w}_i, v_i^*)$ it follows that all $\varrho_{\vec{u}}$-minimal, that is, all $\varrho_{w_{\vec{u}}}$-minimal, antecedent worlds are consequent worlds. And this is so for all contexts $\vec{u} \in R(\mathcal{U})$, that is, all "legal" worlds $w_{\vec{u}} \in \mathcal{W}_0$.

The if-direction follows from the fact that, for each endogenous variable V_i, at most one eligible function holds in a given counterfactual model $\mathcal{M}^*$. For two such functions F and F' differ only if there is a $\vec{w}_i$ such that $F(\vec{w}_i) \neq F'(\vec{w}_i)$. In that case the two counterfactuals $\vec{\mathcal{U} \cup \mathcal{V} \setminus \{V_i\}} = \vec{w}_i \boxdot\!\!\rightarrow V_i = F(\vec{w}_i)$ and $\vec{\mathcal{U} \cup \mathcal{V} \setminus \{V_i\}} = \vec{w}_i \boxdot\!\!\rightarrow V_i = F'(\vec{w}_i)$ have inconsistent consequents, and so cannot be jointly true at any world w. □

7.2. Proof of Theorem 2

THEOREM 7.2. *For each extended acyclic causal model $\mathcal{M} = (\mathcal{S}, \mathcal{F}, (\varrho_{\vec{u}})_{\vec{u} \in R(\mathcal{U})})$ which is focused on actuality and subsequently respects the equations there is a counterfactual model $\mathcal{M}^* = (\mathcal{S}, (\varrho_w)_{w \in \mathcal{W}})$ such that:*

C For all statements ϕ in the language of the generalized version of Halpern and Hitchcock (2010) and all contexts $\vec{u} \in R(\mathcal{U})$:
ϕ is true in $\mathcal{M}$ in $\vec{u}$ according to the structural equations approach iff ϕ is true in $\mathcal{M}^*$ in $w_{\vec{u}}$ according to the counterfactual models account.

Proof. Let $\mathcal{M} = (\mathcal{S}, \mathcal{F}, (\varrho_{\vec{u}})_{\vec{u} \in R(\mathcal{U})})$ be an extended acyclic causal model which is focused on actuality and subsequently respects the equations. Construct $\mathcal{M}^*$ as in the proof of theorem 1.

Suppose ϕ is an atomic sentence of the form $X_i = x$ for some exogenous or endogenous variable X_i. If ϕ is true in $\mathcal{M}$ in context $\vec{u}$ this means that x is the value of X_i in the

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

unique solution $w_{\vec{u}}$ to all the equations in $\mathcal{F}$. But then $w_{\vec{u}} \in \{(u_1, \ldots, u_m, v_1, \ldots, v_n) = (x_1, \ldots, x_{m+n}) : x_i = x\}$. Conversely, if ϕ is not true in $\mathcal{M}$ in context $\vec{u}$ this means that x is not the value of X_i in the unique solution $w_{\vec{u}}$ to all the equations in $\mathcal{F}$, in which case $w_{\vec{u}} \notin \{(x_1, \ldots, x_{m+n}) : x_i = x\}$.

Now suppose ϕ is Boolean. Since negations and conjunctions are defined in the same way in the structural equations approach and the counterfactual models account ϕ is true in $\mathcal{M}$ in context $\vec{u}$ iff ϕ is true in $\mathcal{M}^*$ in "legal" world $w_{\vec{u}}$.

Finally, suppose ϕ is of the form $X_1 = x_1 \wedge \ldots \wedge X_k = x_k \boxdot\!\!\rightarrow \psi$, for short: $\vec{X} = \vec{x} \boxdot\!\!\rightarrow \psi$, where ψ is Boolean. Then ϕ is true in $\mathcal{M}$ in $\vec{u}$ according to the structural equations account just in case ψ is true in that model $\mathcal{M}_{\vec{X}=\vec{x}} = \left(\mathcal{S}_{\vec{X}}, \mathcal{F}^{\vec{X}=\vec{x}}\right)$ and that context $\vec{u}_{\vec{X}=\vec{x}}$ that result from $\mathcal{M}$ and $\vec{u}$ by replacing the equations for X_i by the equations $X_i = x_i$, $i = 1, \ldots, k$. On the other hand, ϕ is true in $\mathcal{M}^*$ in $w_{\vec{u}}$ just in case all $\varrho_{w_{\vec{u}}}$-minimal $\vec{X} = \vec{x}$-worlds are ψ-worlds.

It suffices to consider the case where ψ is an atomic sentence of the form $Z_i = z$. In this case ψ is true in the first sense just in case z is the value of Z_i in the unique solution $w_{\vec{u}_{\vec{X}=\vec{x}}}^{\vec{X}=\vec{x}} =: w^*$ to all equations represented by $\mathcal{F}^{\vec{X}=\vec{x}}$ in context $\vec{u}_{\vec{X}=\vec{x}}$.

We need to show that w^* is the one and only $\varrho_{w_{\vec{u}}}$-minimal $\vec{X} = \vec{x}$-world. w^* is an $\vec{X} = \vec{x}$-world and differs from any other $\vec{X} = \vec{x}$-world w' at most in the values assigned to $\mathcal{U} \cup \mathcal{V} \setminus \{X_1, \ldots, X_k\}$. w^* agrees with $w_{\vec{u}}$ in the values for the exogenous variables $\mathcal{U} \setminus \{X_1, \ldots, X_k\}$. Therefore, if an $\vec{X} = \vec{x}$-world w' differs from w^* in the value of some exogenous variable U, w' differs also from $w_{\vec{u}}$ in the value of U. This means that w^* dominates any such world w' in terms of focus on $w_{\vec{u}}$. Focus on actuality implies that any such world w' has a higher rank in $w_{\vec{u}}$ and so is not among the $\varrho_{\vec{u}}$-minimal $\vec{X} = \vec{x}$-worlds.

This leaves only $\vec{X} = \vec{x}$-worlds which differ from w^* in at most the values for the endogenous variables $\mathcal{V} \setminus \{X_1, \ldots, X_k\}$. Let w' be such a world and suppose $X \in \mathcal{V}^*(w^*) \setminus \mathcal{V}^*(w')$. Since w^* satisfies the equations for all endogenous variables $\mathcal{V} \setminus \{X_1, \ldots, X_k\}$, it must be that $X \in \{X_1, \ldots, X_k\}$. Since w' and w^* agree on the values of $X_1, \ldots, X_k$, and since, by assumption, w' satisfies the equation for X, there must be an exogenous or endogenous variable Y such that $Y \in An(X)$ and w' and w^* differ in the value for Y. The latter implies that Y is endogenous, but not not among $X_1, \ldots, X_k$, and therefore w^* does not violate the equation for Y. If w' violates the equation for Y, we are done. So suppose w' does not violate the equation for Y.

w^* and w' agree on the values of $\mathcal{U}$ as well as $X_1, \ldots, X_k$, w^* satisfies the equations for $\mathcal{V} \setminus \{X_1, \ldots, X_k\}$, and $Y \in \mathcal{V} \setminus \{X_1, \ldots, X_k\}$. Hence, if w' satisfies the equation for Y, there must be an exogenous or endogenous variable Z such that $Z \in An(Y) \subseteq An(X)$ and w' and w^* differ in the value of Z. As before it follows that Z is endogenous, but not among $X_1, \ldots, X_k$, and that w^* satisfies the equation for Z. If w' violates the equation for Z, we are done. If not, there must be another endogenous variable $Z' \in An(Z) \subseteq An(Y) \subseteq An(X)$ with the same properties. Since there are only finitely many variables, and since the model is acyclic, we finally arrive at an endogenous variable $Z^* \in An(X)$ such that w' violates the equation for Z^*, but w^* does not. Hence w^* weakly Halpern-dominates w'.

Note that $\mathcal{V}^*(w') \setminus \mathcal{V}^*(w^*)$ is not empty, if w' differs from w^*. For suppose it is. Then all variables whose equation are violated by w' are also violated by w^*. Since w^* does not violate the equations for $\mathcal{V} \setminus \{X_1, \ldots, X_k\}$, and since w' and w^* agree on the values of $\mathcal{U}$ as well as $X_1, \ldots, X_k$, w' and w^* agree on the values for all variables, and thus are identical.

Since, in acyclic models, $X \notin An(Z^*)$ if $Z^* \in An(X)$, and since $Z^* \in \mathcal{V}^*(w') \setminus \mathcal{V}^*(w^*)$ for at least one endogenous variable Z^*, w' does not weakly Halpern-dominate

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

w^*. Focus on actuality and subsequent respect for the equations implies that any such world w' has a higher rank in $w_{\vec{u}}$ and so is not among the $\varrho_{\vec{u}}$-minimal $\vec{X} = \vec{x}$-worlds. □

§8. Acknowledgments. I am grateful to Rachael Briggs, Thomas Kroedel, Jim Joyce, and, especially, Joe Halpern, Chris Hitchcock, and Wolfgang Spohn for helpful comments on earlier versions of this paper. Part of my research was supported by the German Research Foundation through its Emmy Noether program.

BIBLIOGRAPHY

Briggs, R. (2012). Interventionist counterfactuals. *Philosophical Studies*, **160**, 139–166.

Collins, J., Hall, N., & Paul, L. A. (2004). Counterfactuals and causation: History, problems, and prospects. In Collins, J., Hall, N., and Paul, L. A., editors. *Causation and Counterfactuals*. Cambridge, MA: MIT Press, pp. 1–57.

Field, H. (1978). A note on Jeffrey conditionalization. *Philosophy of Science*, **45**, 361–367.

Glymour, C., Danks, D., Glymour, B., Eberhardt, F., Ramsey, J., Scheines, R., Spirtes, P., Teng, C. M., & Zhang, J. (2010). Actual causation: A stone soup essay. *Synthese*, **175**, 169–192.

Hájek, A. (ms). *Most Counterfactuals Are False*.

Hall, N. (2007). Structural equations and causation. *Philosophical Studies*, **132**, 109–136.

Halpern, J. Y. (2008). Defaults and normality in causal structures. *Proceedings of the Eleventh International Conference on Principles of Knowledge Representation and Reasoning (KR 2008)*, pp. 198–208.

Halpern, J. Y. (2013). From causal models to counterfactual structures. *The Review of Symbolic Logic*, **6**, 305–322.

Halpern, J. Y., & Hitchcock, C. R. (2010). Actual causation and the art of modelling. In Dechter, R., Geffner, H., and Halpern, J., editors. *Heuristics, Probability, and Causality*. London, UK: College Publications, pp. 383–406.

Halpern, J. Y., & Hitchcock, C. R. (forthcoming). Compact representations of extended causal models. *Cognitive Science*.

Halpern, J. Y., & Pearl, J. (2005a). Causes and explanations: A structural-model approach. Part I: Causes. *British Journal for the Philosophy of Science*, **56**, 843–887.

Halpern, J. Y., & Pearl, J. (2005b). Causes and explanations: A structural-model approach. Part II: Explanations. *British Journal for the Philosophy of Science*, **56**, 889–911.

Hiddleston, E. (2005). Causal powers. *British Journal for the Philosophy of Science*, **56**, 27–59.

Hitchcock, C. R. (2001). The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy*, **XCVIII**, 273–299.

Hitchcock, C. R. (2007). Prevention, preemption, and the principle of sufficient reason. *Philosophical Review*, **116**, 495–532.

Huber, F. (ms 1). *What Should I Believe About What Would Have Been the Case?* Unpublished manuscript.

Huber, F. (ms 2). *New Foundations for Counterfactuals*. Unpublished manuscript.

Kistler, M. (forthcoming). The interventionist account of causation and non-causal association laws. *Erkenntnis*.

Kroedel, T., & Huber, F. (forthcoming). Counterfactual dependence and arrow. *Noûs*.

Leitgeb, H. (2012a). A probabilistic semantics for counterfactuals. Part A. *Review of Symbolic Logic*, **5**, 26–84.

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press

Leitgeb, H. (2012b). A probabilistic semantics for counterfactuals. Part B. *Review of Symbolic Logic*, **5**, 85–121.
Lewis, D. K. (1973a). Causation. *Journal of Philosophy*, **70**, 556–567.
Lewis, D. K. (1973b). *Counterfactuals*. Cambridge, MA: Harvard University Press.
Lewis, D. K. (1979). Counterfactual dependence and time's arrow. *Noûs* **13**, 455–476.
Lewis, D. K. (1986). Postscripts to "Causation". In Lewis, D., editor. *Philosophical Papers II*. Oxford, UK: Oxford University Press, pp. 172–213.
Lewis, D. K. (2000). Causation as influence. *Journal of Philosophy*, **97**, 182–197.
Menzies, P. (2004). Difference-making in context. In Collins, J., Hall, N., and Paul, L. A., editors. *Causation and counterfactuals*. Cambridge, MA: MIT Press, pp. 139–180.
Paul, L. A. (2000). Aspect causation. *Journal of Philosophy*, **XCVII**, 235–256.
Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (second edition). Cambridge: Cambridge University Press.
Shenoy, P. P. (1991). On Spohn's rule for revision of beliefs. *International Journal of Approximate Reasoning*, **5**, 149–181.
Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, Prediction, and Search* (second edition). Cambridge, MA: MIT Press.
Spohn, W. (1988). Ordinal conditional functions: A dynamic theory of epistemic states. In Harper, W. L., and Skyrms, B., editors. *Causation in Decision, Belief Change, and Statistics II*. Dordrecht, The Netherlands: Kluwer, pp. 105–134.
Spohn, W. (2006). Causation: An alternative. *British Journal for the Philosophy of Science*, **57**, 93–119.
Spohn, W. (2010). The structural model and the ranking theoretic approach to causation: A comparison. In Dechter, R., Geffner, H., and Halpern, J., editors. *Heuristics, Probability, and Causality*. London, UK: College Publications, pp. 507–522.
Stalnaker, R. C. (1968). A theory of conditionals. In Rescher, N., editor. *Studies in Logical Theory*. American Philosophical Quaterly. Monograph Series, Vol. 2. Oxford, UK: Blackwell, pp. 98–112.
Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford, UK: Oxford University Press.

DEPARTMENT OF PHILOSOPHY
UNIVERSITY OF TORONTO
JACKMAN HUMANITIES BUILDING
170 ST. GEORGE ST.TORONTO
CANADA, ON M5R 2M8
E-mail: franz.huber@utoronto.ca

https://doi.org/10.1017/S175502031300018X Published online by Cambridge University Press