Nondeterministic Causal Models

Anonymous submission

Abstract

We generalize acyclic deterministic structural equation models to the nondeterministic case and argue that it offers an improved semantics for counterfactuals. The standard, deterministic, semantics developed by Halpern (and based on the initial proposal of Galles & Pearl) assumes that for each assignment of values to parent variables there is a unique assignment to their child variable, and it assumes that the actual world (an assignment of values to all variables of a model) specifies a unique counterfactual world for each intervention. Both assumptions are unrealistic, and therefore we drop both of them in our proposal. We do so by allowing *multi-valued* functions in the structural equations. In addition, we adjust the semantics so that the solutions to the equations that obtained in the actual world are preserved in any counterfactual world. We provide a sound and complete axiomatization of the resulting logic and compare it to the standard one by Halpern and to more recent proposals that are closer to ours. Finally, we extend our models to the probabilistic case and show that they open up the way to identifying counterfactuals even in Causal Bayesian Networks.

1 Introduction

Deterministic Structural Equation Models – DSEMs from now on – represent the causal relations between a set of endogenous variables by specifying an equation for each endogenous variable that determines the variable's value as a function of the values of some other variables, both endogenous ($\mathcal{V}$) and exogenous ($\mathcal{U}$) (Pearl 2009). Exogenous variables represent unobserved sources of variation whose existence has to be assumed in order to obtain *deterministic* equations.

DSEMs serve as the mathematical and conceptual foundation for Pearl's causal modelling approach. As they lack probabilities, probabilistic causal models are built on top of DSEMs in two stages. First, we get probabilistic DSEMs – PDSEMs – by adding a probability distribution over the exogenous variables $\mathcal{U}$. This induces a joint distribution over $(\mathcal{U}, \mathcal{V})$. Second, assuming that there are no cyclic causal relations, that exogenous variables are independent, and that no two endogenous variables share an exogenous parent (i.e., that the model is *Markovian*), we get *Causal Bayesian Networks* by marginalizing out the exogenous variables and directly considering the marginal distribution over $\mathcal{V}$ and its Markov factorization (Pearl 2009; Bongers et al. 2021).

We here present *Nondeterministic Structural Equation Models* – NSEMs – as a more general and improved foundation for causal models. The reason for doing so is that the heavy reliance on exogenous variables in DSEMs is unnecessarily restrictive. The generalization from DSEMs to NSEMs is given by dropping the assumption that there must exist exogenous variables such that the value of each endogenous variable can be uniquely determined by its causal parents. Note that dropping this assumption is not the same as dropping the assumption that the world is itself fundamentally deterministic. It is one thing to assume that there must exist some properties of the world such that the value of a variable is uniquely determined, it is quite another to assume that these properties can be neatly compartmentalized into sets of variables whose values determine the other value according to a stable functional relationship. In many situations the latter may not hold, and such situations are currently beyond the scope of SEMs. (And in even more situations, we may want to – or have to due to practical constraints – represent the world *as if* it does not hold.)

Furthermore, the use of exogenous variables also involves a commitment to an overly strong semantics for causal counterfactual statements, for it implies that there is always a *unique* solution to the model for any counterfactual query, given the actual values of all variables. That assumption is even stronger than the previous one. For example, consider a patient who may or may not receive treatment for some condition, and when she does she may or may not recover, depending on whether the treatment is effective. We observe that in the actual world, she does not receive treatment, and she does not recover. Representing this world using DSEMs requires specifying *actual* values of exogenous variables such that if she had been treated, she would have either certainly recovered, or she would certainly not have. Using NSEMs there is no such requirement, and thus we can simply consider both counterfactual worlds as possible.

The benefit of dropping this counterfactual *uniqueness property* should not be underestimated. The property has been severely criticized in the general philosophical literature on counterfactuals ever since the seminal work of (Lewis 1973) first did so. More relevant for the present purposes, it is one of the driving forces behind (Dawid 2000)'s influential criticism of the counterfactual semantics offered by both DSEMs and the related *Potential Outcomes* framework (Rubin 1974), and this criticism lies at the heart of the recent back-and-forth over the correct methodology for personalised treatment decisions, such as in the patient example above (Dawid and Senn 2023; Mueller and Pearl 2023). By offering a semantics for counterfactuals that avoids this property, NSEMs provide room for both sides of this hotly contested – and literally vital – dispute to meet in the middle.

We proceed as follows. We define NSEMs (Sec. 2), present a formal language and corresponding semantics for causal formulas that hold in NSEMs (Sec. 3), compare our logic to the standard one for DSEMs (Sec. 4), offer a sound and complete axiomatization and compare to other recent proposals that allow nondeterminism (Sec. 5), and offer a preliminary investigation into the probabilistic generalization of NSEMs, sketching how it allows for computing counterfactuals even in Causal Bayesian Networks (Sec. 6).

2 Nondeterministic Structural Equation Models

We take the definition of deterministic causal models by (Halpern 2016) and generalize it to the nondeterministic case by using multi-valued functions [1]. As a first step, we need to define a signature as the variables out of which a causal model is built up.

Definition 2.1: A signature $\mathcal{S}$ is a tuple $(\mathcal{U}, \mathcal{V}, \mathcal{R})$, where $\mathcal{U}$ is a set of *exogenous* variables, $\mathcal{V}$ is a set of *endogenous* variables, and $\mathcal{R}$ a function that associates with every variable $Y \in \mathcal{U} \cup \mathcal{V}$ a nonempty set $\mathcal{R}(Y)$ of possible values for Y (i.e., the set of values over which Y *ranges*). If $\vec{X} = (X_1, \ldots, X_n)$, $\mathcal{R}(\vec{X})$ denotes the crossproduct $\mathcal{R}(X_1) \times \cdots \times \mathcal{R}(X_n)$. ∎

A causal model expresses the causal relations between the endogenous variables of a signature. In addition to using multi-valued functions, we depart from Halpern by explicitly including the causal graph as an element of the causal model.

Definition 2.2 : A *causal model* (or a *Nondeterministic Structural Equation Model – NSEM*) M is a triple $(\mathcal{S}, \mathcal{F}, \mathcal{G})$, where $\mathcal{S}$ is a signature, $\mathcal{G}$ is a directed graph such that there is one node for each variable in $\mathcal{S}$, and $\mathcal{F}$ defines a function that associates with each endogenous variable X a *structural equation* F_X giving the possible values of X in terms of the values of some of the other endogenous and exogenous variables. A structural equation F_X takes on the form $X = f_X(\vec{Pa}_X)$, where $\vec{Pa}_X \subseteq (\mathcal{U} \cup \mathcal{V} - \{X\})$ are the *parents* of X as they appear in $\mathcal{G}$, and $f_X : \mathcal{R}(\vec{Pa}_X) \rightarrow \mathcal{P}(\mathcal{R}(X))$. (Here $\mathcal{P}(\mathcal{R}(X))$ is the *powerset* of $\mathcal{R}(X)$: the set that contains as its elements all subsets of $\mathcal{R}(X)$.)∎

[1](Halpern 2000) already suggested this generalization, but never implemented it. He did recently offer an even further generalization in order to allow for an infinite number of variables with infinite ranges, by doing away with equations altogether (Peters and Halpern 2021; Halpern and Peters 2022). Recently (Barbero 2024) and (Wysocki 2023) have likewise offered generalizations to the nondeterministic case. We compare our approach to these ones in Section 5.

We here restrict attention to the case in which $\mathcal{G}$ is acyclic (so that $\mathcal{G}$ is a DAG – a Directed Acyclic Graph). If for each f_X the co-domain does not contain the empty set, we say that a causal model is *total*. We here restrict ourselves to total causal models. This amounts to the assumption that for all possible settings $\vec{pa}_X$ of the parents, there exists at least one solution x for the child.

There are no equations for exogenous variables $\mathcal{U}$, as these are taken to represent the background conditions that are simply given. We call $\vec{u} \in \mathcal{R}(\mathcal{U})$ a *context*, $\vec{v} \in \mathcal{R}(\mathcal{V})$ a *state*, and $(\vec{u}, \vec{v}) \in \mathcal{R}(\mathcal{U} \cup \mathcal{V})$ is a *world*. In *deterministic* causal models all the functions f_X are standard as opposed to multi-valued, and thus each equation has a unique solution x for each choice of values $\vec{pa}_X$. In nondeterministic models, a *solution* of the equation $X = f_X(\vec{Pa}_X)$ is a tuple $(x, \vec{pa}_X)$ such that $x \in f_X(\vec{pa}_X)$. (The term "equation" is thus somewhat strange, but we stick with it given the common reference to causal models as "structural equation models". Furthermore, an equivalent characterization can be given in terms of a literal equation as well.)[2]. A solution of M is a world $(\vec{u}, \vec{v})$ that is a solution of all equations in $\mathcal{F}$.

In acyclic models the solutions of M given a context $\vec{u}$ can be determined recursively: determine the solutions of each equation in the partial order given by $\mathcal{G}$, and pass these on to the next equation. In general, deterministic causal models (DSEMs) need not have solutions for each context, nor do any such solutions have to be unique, but unicity and existence are guaranteed for acyclic deterministic models. When we move to nondeterministic causal models, unicity and existence are no longer guaranteed even for acyclic models, but existence is recovered in total models.

3 The Causal Language

Given a signature $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R})$, an *atomic formula* is a formula of the form $X = x$, for $X \in \mathcal{V}$ and $x \in \mathcal{R}(X)$. A *basic formula (over $\mathcal{S}$)* φ is a Boolean combination of atomic formulas. An *intervention* has the form $\vec{Y} \leftarrow \vec{y}$, where $Y_1, \ldots, Y_k$ are distinct variables in $\mathcal{V}$, and $y_i \in \mathcal{R}(Y_i)$ for each $1 \leq i \leq k$. A *basic causal formula* has the form $[Y_1 \leftarrow y_1, \ldots, Y_k \leftarrow y_k]\varphi$, where φ is a basic formula and $Y_1, \ldots, Y_k$ are distinct variables in $\mathcal{V}$. Such a formula is abbreviated as $[\vec{Y} \leftarrow \vec{y}]\varphi$. The special case where $k = 0$ is abbreviated as $[]\varphi$. Finally, a *causal formula* is a Boolean combination of basic causal formulas. The language $\mathcal{L}(\mathcal{S})$ that we consider consists of all causal formulas.

A causal formula ψ is true or false in a causal model, given a world. We write $(M, \vec{u}, \vec{v}) \models \psi$ if the causal formula ψ is true in causal model M given world $(\vec{u}, \vec{v})$. We call a model-world pair $(M, \vec{u}, \vec{v})$ a *causal setting*.

[2]Concretely, it turns out that the approach of (Bongers et al. 2021) can be used to express our nondeterministic "equations" as literal equations. Although they do not explicitly consider nondeterministic models, their models do allow "self-cycles", meaning they allow for X to depend on X. Note that $X = f_X(\vec{Pa}_X)$ is equivalent to $X = X + (1 - 1_{f_X}(X, \vec{Pa}_X))$, where 1_{f_X} is the indicator function that returns 1 iff $X \in f_X(\vec{Pa}_X)$. This correspondence opens up the possibility of integrating both approaches, but that is something we defer to future work.

We first define the $\models$ relation for basic formulas. $(M, \vec{u}, \vec{v}) \models X = x$ if x is the restriction of $\vec{v}$ to X. We extend $\models$ to basic formulas φ in the standard way. Note that the truth of basic formulas is determined solely by the state $\vec{v}$, and thus we often also write $\vec{v} \models \varphi$.

In order to define the $\models$ relation for causal formulas, we introduce two operations on a causal model, the *actualized refinement* that is the result of integrating the actual behavior of the equations as observed in a world $(\vec{u}, \vec{v})$ into the equations of a model M, and the *intervened* model that is the result of performing an intervention on the equations of a model M.

Definition 3.1: Given causal models M' and M over identical signatures $\mathcal{S}$ and with identical graphs $\mathcal{G}$, we say that *M' is a refinement of M* if for all $X \in \mathcal{V}$ and all $\vec{pa}_X \in \mathcal{R}(\vec{Pa}_X)$: $f_V^{M'}(\vec{pa}_X) \subseteq f_V^M(\vec{pa}_X)$. ∎

Since we are restricting ourselves to total causal models, the only refinement of a deterministic model is itself.[3] Hence we say that a deterministic model is *maximally refined*. Concretely, since the way that the equations determine the outcome in a deterministic model is identical across each world $(\vec{u}, \vec{v})$, there is no need to additionally consider how the equations behave in a specific world. This is no longer true for nondeterministic models, as there the actual values that obtained in a world inform us about how the equations behaved for those values. When evaluating formulas in a world, we need to take this information into account, and to do so requires refining the equations so that they incorporate this actual behavior.

Definition 3.2: Given a solution $(\vec{u}, \vec{v})$ of a model $M = (\mathcal{S}, \mathcal{F}, \mathcal{G})$, we define the *actualized refinement* $M^{(\vec{u},\vec{v})}$ as the refinement of M in which $\mathcal{F}$ is replaced by $\mathcal{F}^{(\vec{u},\vec{v})}$, as follows: for each variable $X \in \mathcal{V}$, its function f_X is replaced by $f_X^{(\vec{pa}_X, x)}$ that behaves identically to f_X for all inputs except for $\vec{pa}_X$. Instead, $f_X^{(\vec{pa}_X, x)}(\vec{pa}_X) = x$, where x and $\vec{pa}_X$ are the respective restrictions of $(\vec{u}, \vec{v})$ to X and $\vec{Pa}_X$. ∎

Setting the value of some variables $\vec{Y}$ to $\vec{y}$ in a causal model $M = (\mathcal{S}, \mathcal{F}, \mathcal{G})$ results in a new causal model, denoted $M_{\vec{Y} \leftarrow \vec{y}}$, which is identical to M, except that $\mathcal{F}$ is replaced by $\mathcal{F}^{\vec{Y} \leftarrow \vec{y}}$: for each variable $X \notin \vec{Y}$, $F_X^{\vec{Y} \leftarrow \vec{y}} = F_X$ (i.e., the equation for X is unchanged), while for each Y' in $\vec{Y}$, the equation $F_{Y'}$ for Y' is replaced by $Y' = y'$ (where y' is the value in $\vec{y}$ corresponding to Y'). Similarly, $\mathcal{G}$ is replaced with $\mathcal{G}^{\vec{Y} \leftarrow \vec{y}}$.

With these operations in place, we can define the $\models$ relation for basic causal formulas, relative to settings $(M, \vec{u}, \vec{v})$ such that $(\vec{u}, \vec{v})$ is a solution of M. $(M, \vec{u}, \vec{v}) \models [\vec{Y} \leftarrow \vec{y}]\varphi$ iff $\vec{v}' \models \varphi$ for all states $\vec{v}'$ such that $(\vec{u}, \vec{v}')$ is a solution of $(M^{(\vec{u},\vec{v})})_{\vec{Y} \leftarrow \vec{y}}$. We inductively extend the semantics to causal formulas in the standard way, that is, $(M, \vec{u}, \vec{v}) \models [\vec{Y} \leftarrow \vec{y}]\varphi_1 \wedge [\vec{Z} \leftarrow \vec{z}]\varphi_2$ iff $(M, \vec{u}, \vec{v}) \models [\vec{Y} \leftarrow \vec{y}]\varphi_1$ and $(M, \vec{u}, \vec{v}) \models [\vec{Z} \leftarrow \vec{z}]\varphi_2$, and similarly for $\neg$ and $\vee$.

[3] Concretely, if we drop totality, then a refinement allows for $f_X(\vec{pa}_X) = \emptyset$, which does not result in a deterministic model.

We define $\langle \vec{Y} \leftarrow \vec{y} \rangle \varphi$ as an abbreviation of $\neg[\vec{Y} \leftarrow \vec{y}]\neg\varphi$. So $(M, \vec{u}, \vec{v}) \models \langle \vec{Y} \leftarrow \vec{y} \rangle true$ iff there is some world $(\vec{u}, \vec{v}')$ that is a solution of $(M^{(\vec{u},\vec{v})})_{\vec{Y} \leftarrow \vec{y}}$. We then write $(M, \vec{u}, \vec{v}) \models \langle \vec{Y} \leftarrow \vec{y} \rangle \mathcal{V} = \vec{v}'$.

We can also evaluate formulas with respect to just a partial causal setting $(M, \vec{u})$, or even with respect to a model M by itself. For basic causal formulas, we define that $(M, \vec{u}) \models [\vec{Y} \leftarrow \vec{y}]\varphi$ iff $(M, \vec{u}, \vec{v}) \models [\vec{Y} \leftarrow \vec{y}]\varphi$ holds for all states $\vec{v}$ such that $(\vec{u}, \vec{v})$ is a solution of M. In a similar fasion, we define that $M \models [\vec{Y} \leftarrow \vec{y}]\varphi$ iff $(M, \vec{u}, \vec{v}) \models [\vec{Y} \leftarrow \vec{y}]\varphi$ holds for all solutions $(\vec{u}, \vec{v})$ of M. We again inductively extend to causal formulas in the standard way: $(M, \vec{u}) \models [\vec{Y} \leftarrow \vec{y}]\varphi_1 \wedge [\vec{Z} \leftarrow \vec{z}]\varphi_2$ iff $(M, \vec{u}) \models [\vec{Y} \leftarrow \vec{y}]\varphi_1$ and $(M, \vec{u}) \models [\vec{Z} \leftarrow \vec{z}]\varphi_2$, and similarly for $\neg$ and $\vee$. Likewise for $M \models \psi$.

We have now defined three different semantics for $\mathcal{L}(\mathcal{S})$: the first with respect to full causal settings $(M, \vec{u}, \vec{v})$, the second with respect to partial causal settings $(M, \vec{u})$, and the third with respect to M. The first evaluates causal formulas relative to a single world $(\vec{u}, \vec{v})$, and thus we will call the resulting logic the *single world counterfactual logic*, or **swc** logic for short. It formalizes counterfactual statements relative to a specific *actual world*. The second only requires specifying a single context $\vec{u}$, and thus it considers the entire set of worlds $(\vec{u}, \vec{v})$ that extend $\vec{u}$ together, and thus we call the resulting logic the *single context counterfactual logic*, or **scc** logic for short. Its formulas are also counterfactual, in the sense that they can express antecedents which run counter to the facts of a set of worlds extending a context. Hence both these logics occur on rung 3 of (Pearl 2009)'s causal hierarchy. Lastly, the third semantics evaluates formulas without assuming any factual knowledge that restricts the set of possible worlds determined by M, and thus its formulas are entirely forward-looking and interventionist. Hence these semantics are for expressions on rung 2 of Pearl's causal hierarchy, and therefore we call the resulting logic the *interventionist logic*. In the remainder of this paper our attention goes to the two counterfactual logics, but we flag the interventionist logic as a worthwhile study for future investigation.

4 Comparison to the Standard Semantics

To recap, the standard logic for causal models from (Halpern 2000, 2016) is a logic for DSEMs instead of NSEMs, and its semantics are defined identically to ours except that it does not use the actualized refinement. Given that for any DSEM the actualized refinement $M^{(\vec{u},\vec{v})}$ will simply be M anyway, the two semantics are in fact entirely equivalent for DSEMs.

Concretely, the standard semantics defines the $\models$ relation for basic causal formulas as: $(M, \vec{u}, \vec{v}) \models [\vec{Y} \leftarrow \vec{y}]\varphi$ iff $\vec{v}' \models \varphi$ for all states $\vec{v}'$ such that $(\vec{u}, \vec{v}')$ is a solution of $M_{\vec{Y} \leftarrow \vec{y}}$. Note that, importantly, this semantics does not depend on the actual state $\vec{v}$, and therefore it can be (and usually is) written and interpreted as a semantics for partial causal settings $(M, \vec{u})$. As a result, the distinction between the counterfactual logics **swc** and **scc** collapses in the case of DSEMs, and thus there is only a single counterfactual logic.

If the DSEMs are acyclic, then for each $\vec{u}$ there is a unique $\vec{v}$ such that $(\vec{u}, \vec{v})$ is a solution of M, and thus the counterfactual logic is with respect to a single actual world, as is the case for our **swc**. Also, for each $\vec{u}$, there is a unique $\vec{v}$ such that $(\vec{u}, \vec{v})$ is a solution of $M_{\vec{Y} \leftarrow \vec{y}}$ for any $\vec{Y} \leftarrow \vec{y}$, and hence the standard semantics for acyclic DSEMs satisfies the controversial uniqueness property we mentioned in Section 1. If the DSEMs are cyclic, then for each context there may be multiple solutions, or none. Thus, in this case the counterfactual logic is with respect to the set of worlds that extend a context $\vec{u}$, as is the case for our **scc** (except that for **scc** a solution is guaranteed to exist, given totality). As we discuss in Section 5, these semantics still satisfy a property that is conceptually very similar to the uniqueness property.

Importantly, as the following Theorem shows, when we move from DSEMs to NSEMs, the actualized refinement *only matters for* **swc**, and thus for both the **scc** logic and the interventionist logic our semantics are simply generalizations of the standard semantics to the nondeterministic case.

Theorem 4.1: *Given a nondeterministic causal model M, we have that for all $\vec{Y} \subseteq \mathcal{V}$, for all $\vec{y} \in \mathcal{R}(\vec{Y})$, and for all basic formulas φ:*

- *$M \models [\vec{Y} \leftarrow \vec{y}]\varphi$ iff $\vec{v} \models \varphi$ for all solutions $(\vec{u}, \vec{v})$ of $M_{\vec{Y} \leftarrow \vec{y}}$.*
- *For all contexts $\vec{u}$: $(M, \vec{u}) \models [\vec{Y} \leftarrow \vec{y}]\varphi$ iff $\vec{v} \models \varphi$ for all states $\vec{v}$ such that $(\vec{u}, \vec{v})$ is a solution of $M_{\vec{Y} \leftarrow \vec{y}}$.*

Proof: All proofs are to be found in the Supplementary Material. ∎

The intuition behind Theorem 4.1 is that the actualized refinement operation updates M's equations to include their *actual behavior*, and there is no actual behavior unless one specifies an actual world $(\vec{u}, \vec{v})$, so the operation can be ignored for **scc** and interventionist logics. Concretely, single world counterfactuals are statements about what is true in a particular, fully specified, world $(\vec{u}, \vec{v})$ that is governed by a causal model M, if we were to intervene to set some variables counter to fact.[4] Single context counterfactuals, on the other hand, are statements about what holds in all worlds that share a context $\vec{u}$ and are governed by the causal model M. Interventionist statements are of even wider scope, as they do not rely on any assumption except that the worlds are governed by the model M. In both of the latter cases, this means that we take into account *all* the possible ways in which M (or $(M, \vec{u})$) can be actualized, and thus the various possible actualized refinements ought to cancel each other out. We take the fact that our semantics bears out this intuition to be an important sanity check.

As far as we are aware, the distinction between **swc** and **scc** logics has gone unnoticed so far. In fact, recently (Halpern and Peters 2022) introduced generalized structural equation models – GSEMs – and they provide an axiomatization for the more general logic that is the result, but in doing so Halpern has moved entirely from an **swc** type of logic – that was the subject matter of (Halpern 2000) and (Galles and Pearl 1998) – towards a **scc** type of logic, for the logic for GSEMs does not allow evaluating formulas with respect to a single world. In the next section we conclude the definition of our **scw** and **scc** logics by offering a sound and complete axiomatization for both, and interpret the results in light of other recently proposed logics for nondeterministic causal reasoning.

[4]Usually one speaks of counterfactuals even when variables are set to their actual values by an intervention, but in such cases the intervention should have no impact on the truth of factual formulas, and Proposition 5.3 shows that this is indeed the case for our semantics.

5 Axiomatization and Recent Related Work

Throughout this section we hold fixed some finite signature $\mathcal{S} = (\mathcal{U}, \mathcal{F}, \mathcal{R})$, i.e., $\mathcal{U}$ and $\mathcal{V}$ are finite, and $\mathcal{R}(X)$ is finite for all $X \in \mathcal{U} \cup \mathcal{V}$. Let AX be the axiom system for the language $\mathcal{L}(\mathcal{S})$ that consists of the following list of axioms and inference rule MP.

D0. All instances of propositional tautologies.

D1. $[\vec{Y} \leftarrow \vec{y}](X = x \Rightarrow X \neq x')$ if $x, x' \in \mathcal{R}(X)$, $x \neq x'$ (functionality)

D2. $[\vec{Y} \leftarrow \vec{y}](\bigvee_{x \in \mathcal{R}(X)} X = x)$ (definiteness)

D3(a). $\langle \vec{X} \leftarrow \vec{x} \rangle (W = w \wedge \varphi) \Rightarrow \langle \vec{X} \leftarrow \vec{x}, W \leftarrow w \rangle (\varphi)$ if $W \notin \vec{X}$ (weak composition)

D3(b). $[\vec{X} \leftarrow \vec{x}](W = w \wedge \varphi) \Rightarrow [\vec{X} \leftarrow \vec{x}, W \leftarrow w](\varphi)$ if $W \notin \vec{X}$ (strong composition)

D4. $[\vec{X} \leftarrow \vec{x}](\vec{X} = \vec{x})$ (effectiveness)

D5. $(\langle \vec{X} \leftarrow \vec{x}, Y \leftarrow y \rangle (W = w \wedge \vec{Z} = \vec{z}) \wedge \langle \vec{X} \leftarrow \vec{x}, W \leftarrow w \rangle (Y = y \wedge \vec{Z} = \vec{z}))$
$\Rightarrow \langle \vec{X} \leftarrow \vec{x} \rangle (W = w \wedge Y = y \wedge \vec{Z} = \vec{z})$ if $\vec{Z} = \mathcal{V} - (\vec{X} \cup \{W, Y\})$ (reversibility)

D6. $(X_0 \rightsquigarrow X_1 \wedge \ldots \wedge X_{k-1} \rightsquigarrow X_k) \Rightarrow \neg(X_k \rightsquigarrow X_0)$ (recursiveness)

D7. $([\vec{X} \leftarrow \vec{x}]\varphi \wedge [\vec{X} \leftarrow \vec{x}](\varphi \Rightarrow \psi)) \Rightarrow [\vec{X} \leftarrow \vec{x}]\psi$ (distribution)

D8. $[\vec{X} \leftarrow \vec{x}]\varphi$ if φ is a propositional tautology (generalization)

D9. $\langle \vec{Y} \leftarrow \vec{y} \rangle true \wedge (\langle \vec{Y} \leftarrow \vec{y} \rangle \varphi \Rightarrow [\vec{Y} \leftarrow \vec{y}]\varphi)$ if $\vec{Y} = \mathcal{V}$ or, for some $X \in \mathcal{V}$, $\vec{Y} = \mathcal{V} - \{X\}$ (unique outcomes for $\mathcal{V}$ and $\mathcal{V} - \{X\}$)

D10(a). $\langle \vec{Y} \leftarrow \vec{y} \rangle true$ (at least one outcome)

D10(b). $\langle \vec{Y} \leftarrow \vec{y} \rangle \varphi \Rightarrow [\vec{Y} \leftarrow \vec{y}]\varphi$ (at most one outcome)

D10(c). $\langle\rangle\varphi \Rightarrow []\varphi$ (at most one actual outcome)

MP. From φ and $\varphi \Rightarrow \psi$, infer ψ (modus ponens)

Here, $Y \rightsquigarrow Z$ means that $Y \neq Z$ and $\vee_{\vec{X} \subseteq \mathcal{V}, \vec{x} \in \mathcal{R}(\vec{X}), y \neq y' \in \mathcal{R}(Y), z \in \mathcal{R}(Z)} (\langle \vec{X} \leftarrow \vec{x}, Y \leftarrow y \rangle (Z = z) \wedge [\vec{X} \leftarrow \vec{x}, Y \leftarrow y'](Z \neq z))$.

(Halpern and Peters 2022) define $\rightsquigarrow$ differently. However, our definition is easily seen to be equivalent to theirs in the presence of D10(b). Note that none of the axioms mentions basic formulas φ. For both the standard semantics and for ours this is without loss of generality, because φ is easily seen to be equivalent to $[]\varphi$. (Alternatively, we could have extended the causal language and add the axiom $\varphi \Leftrightarrow []\varphi$, which is the choice made by (Barbero 2024).) Lastly, note that in the presence of D10(b), AX contains redundant axioms: D3(b) is a consequence of D3(a) and D10(b), D9 is a consequence of D2 and D10(a,b), D10(c) is a consequence of D10(b), and – as shown by (Halpern and Peters 2022) – D5 is a consequence of D2, D3, D6, D7, D8, and D10(a,b).

(Halpern and Peters 2022) show that AX without D3(b) and D10(c) is a sound and complete axiomatization with respect to acyclic DSEMs that have signature $\mathcal{S}$, and thus the same holds for AX. Furthermore, none of the D10 axioms holds for cyclic DSEMs. In fact, (Halpern and Peters 2022) show that AX without D3(b), D10(a,b,c), and D6, is a sound and complete axiomatization for cyclic DSEMs.

D10(b) implies the *uniqueness* property mentioned earlier. To see why, consider some solution $(\vec{u}, \vec{v})$ (or context $\vec{u}$) relative to which we are evaluating formulas, and take φ to be $\mathcal{V} = \vec{v'}$ for some $\vec{v'} \in \mathcal{V}$. D10(b) then commits us to the claim that if the world $(\vec{u}, \vec{v'})$ is possible under the counterfactual supposition that $\vec{Y}$ were $\vec{y}$, then it is the only world that is possible under that supposition. Although cyclic DSEMs do not validate D10(b), they do validate D9, and that is conceptually not at all different: it simply restricts the above claim to the special case in which the counterfactual supposition contains all but one endogenous variable. As anticipated, neither axioms are sound for either our **swc** logic or our **scc** logic, and thus they free causal models from their commitment to this controversial property. (For sake of completeness, note that neither axiom is sound for our interventionist logic either, nor is it for the interventionist logic over DSEMs only.) The following model offers a simple counterexample.

Example 5.1: Consider the model M with two binary endogenous variables X, Y such that $\vec{Pa}_Y = \{X\}$, $\vec{Pa}_X = \emptyset$, the equation for X is $X = 1$, and the equation for Y is $Y \in \{0, 1\}$ if $X = 1$ and $Y = 0$ if $X = 0$. Then $(M, X = 0, Y = 0) \models \langle X \leftarrow 1 \rangle Y = 1$ and yet $(M, X = 0, Y = 0) \not\models [X \leftarrow 1] Y = 1$, since also $(M, X = 0, Y = 0) \models \langle X \leftarrow 1 \rangle Y = 0$. Also, $M \models \langle X \leftarrow 1 \rangle Y = 1$ and yet $M \not\models [X \leftarrow 1] Y = 1$.

We let AX^{scc}_{non} denote the axiom system D0-D8 and D10(a), and AX^{swc}_{non} denotes AX^{scc}_{non} plus D10(c). The following result shows that our **swc** logic is stronger than our **scc** logic, that both of them are weaker than the counterfactual logic for acyclic DSEMS, and that both of them are incomparable to the counterfactual logic for cyclic DSEMs.

Theorem 5.2: *AX^{swc}_{non} (resp. AX^{scc}_{non}) is a sound and complete axiomatization for the language $\mathcal{L}(\mathcal{S})$ with respect to the* **swc** *logic (resp. the* **scc** *logic) over acyclic NSEMS that have signature $\mathcal{S}$.*

D10(c) is what distinguishes our two logics, as it does not hold for the **scc** logic. Simply consider the model M consisting of a single binary endogenous variable X with an equation such that $X = \{0, 1\}$. Then we have that both $M \models \langle\rangle X = 1$ and $M \models \langle\rangle X = 0$ and thus also $M \not\models [] X = 1$. That D10(c) is not sound for our **scc** logic is what prevents the first problematic assumption of acyclic DSEMs mentioned in Section 1, namely that the values $\vec{u}$ for all exogenous variables *uniquely* determine an actual world $(\vec{u}, \vec{v})$. (As with D10(b), to see this take φ to be of the form $\mathcal{V} = \vec{v'}$.) In the context of our **swc** logic, D10(c) simply expresses the sensible property that if the world were as it actually is, then the actual world is the only world possible. To see this, note that for any $(M, \vec{u}, \vec{v})$, the only way for $\langle\rangle \mathcal{V} = \vec{v'}$ to hold is when $\vec{v'} = \vec{v}$.

Halpern & Peters. Theorem 5.2 shows that both of our counterfactual logics are stronger than the one introduced in (Halpern and Peters 2022) for so-called GSEMs, *Generalized Structural Equation Models*, even when restricting to the case that is most similar to ours (the case in which all interventions are considered, the signature is finite, and the graph is acyclic). Briefly, GSEMs do away with structural equations altogether, and simply define there to be some set of states $\vec{v}$ corresponding to each context-intervention pair $(\vec{u}, \vec{Y} \leftarrow \vec{y})$. As a result, this means that the connection between formulas that hold under some intervention $\vec{X} \leftarrow \vec{x}$ and those that should hold under an extension $\vec{X} \leftarrow \vec{x}, W \leftarrow w$ is lost, and thus the axioms D3 (both of them) and D5 are no longer sound, in addition to losing D9 and D10(a,b,c). (See their Theorem 5.3.)

Barbero. The closest logic to ours is that of (Barbero 2024), as it also generalizes structural equation models by allowing for multi-valued functions in the equations. (His work formed part of the motivation for the current paper.) Except for the fact that he also considers the cyclic case, there are two main differences with our proposal.

Firstly, his logic builds on *team semantics*, meaning that instead of evaluating formulas with respect to either a single solution $(\vec{u}, \vec{v})$ or a single context $\vec{u}$ of a model M, he does so with respect to a *team*, where a team is any subset of solutions of a model M. As a result, the logic is a mixture of all three of our logics: consider a singleton team and causal formulas are effectively single world counterfactual formulas, consider a team that consists of all solutions extending some context $\vec{u}$, and causal formulas are effectively single context counterfactual formulas, consider a team that consists of al solutions, and causal formulas are interventionist formulas, consider a team that belongs to none of these categories and the causal formula is some form of hybrid.

Second, although his semantics also aims to incorporate the actual behavior of the equations into all counterfactual worlds, his semantics is less demanding than our actualized refinements are. It merely requires that the values of all non-descendants of the variables in some intervention $\vec{X} \leftarrow \vec{x}$ remain identical to the actual ones in counterfactual worlds.[5] It is easy to show that this property also holds for our semantics, i.e., for any variable Y that is not a descendant of any

[5]At least, that is the intended semantics. When moving from the cyclic to the acyclic case, Barbero drops the requirement on non-descendants. Personal communication with the author confirms that this was unintended.

variable in $\vec{X}$, we have for any solution $(\vec{u}, \vec{v})$ of M that if $(M, \vec{u}, \vec{v}) \models Y = y$ then $(M, \vec{u}, \vec{v}) \models [\vec{X} \leftarrow \vec{x}]Y = y$.

As a result of the second difference, weak composition – D3(a) – comes apart from strong composition – D3(b) – for Barbero. Just as with our logics, D3(a) is sound and D10(b) is not. However, D3(b) is no longer sound for his logic. We believe this is an undesirable property. Consider again the model from Example 5.1, and consider the world $(M, X = 1, Y = 1)$. Then according to both Barbero and our counterfactual logic it holds that $(M, X = 1, Y = 1) \models []Y = 1$, yet for Barbero's logic we do not have that $(M, X = 1, Y = 1) \models [X \leftarrow 1]Y = 1$, the reason being that for his logic $(M, X = 1, Y = 1) \models \langle X \leftarrow 1 \rangle Y = 0$. As Barbero points out, this results in his logic not validating *conjunction conditionalization*, which is seen by many as a crucial property for a logic of counterfactuals to satisfy (see (Walters and Williams 2013) for a thorough defense).[6] We view it as an argument in favor of our approach that it does satisfy this property:

Proposition 5.3: *For a solution $(\vec{u}, \vec{v})$ of M, if $(M, \vec{u}, \vec{v}) \models \vec{X} = \vec{x} \wedge \varphi$ then $(M, \vec{u}, \vec{v}) \models [\vec{X} \leftarrow \vec{x}]\varphi$.*

As to the first difference, it is not clear to us what the notion of a team is supposed to capture. The most straightforward interpretation would be that it expresses the set of "possible worlds" in some sense or other. However, that interpretation conflicts with how teams behave. Again considering Example 5.1, we add that the team consists exclusively of the solution $(X = 1, Y = 1)$. If we then consider the intervention $X \leftarrow 1$, the team suddenly changes to also including the solution $(X = 1, Y = 0)$. We fail to understand why interventions that do not change anything should be able to result in changing what worlds we consider possible. Another consequence of his use of teams is that his logic invalidates D10(c), contrary to our **swc** logic.

Wysocki. (Wysocki 2023) also recently proposed to generalize structural equation models using multi-valued functions. His logic, however, only considers formulas with respect to an entire causal model M, and is thus what we have called a logic for interventionist reasoning rather than for counterfactual reasoning. There is only one important difference with our interventionist logic.

He follows Halpern's standard approach in that he *defines* the parent relation between variables rather than reading it of an additional graph $\mathcal{G}$ that we (and Barbero) have added to the definition of a causal model. Concretely, for DSEMs (Halpern 2016) defines the equation for a variable Y to consist of a function $f_Y : \mathcal{R}(\mathcal{U} \cup \mathcal{V} - \{Y\}) \rightarrow \mathcal{R}(Y)$, and then defines Y *depends* on X to mean that there exist $x \neq x' \in \mathcal{R}(X)$ and $\vec{z} \in \mathcal{R}(\mathcal{U} \cup \mathcal{V} - \{X, Y\})$ such that $f_Y(\vec{z}, x) \neq f_Y(\vec{z}, x')$. Wysocki uses the same definition, except that he generalizes it to the multi-valued case. Both Halpern and Wysocki then take "X is a parent of Y" to be synonymous with "Y depends on X". Clearly, if Y depends on X then it will also be a parent in our framework, but the reverse need not hold. In fact, following Barbero, we can distinguish between the graph $\mathcal{G}_M$ that comes with a model and the graph $\mathcal{G}_D$ that is built up out of all the dependence relations, noting that $\mathcal{G}_D$ is a subgraph of $\mathcal{G}_M$.[7] Although nothing prevents one from restricting our framework to models M in which $\mathcal{G}_M = \mathcal{G}_D$, we believe it is a mistake to enforce this stricter definition of a parent in general, because it prohibits a natural extension to probabilistic causal models. Viewed probabilistically, if conditional on all other variables, X can change the *distribution* of Y, then X is a parent of Y. Changing the distribution does not require than any value y changes from being possible $(p > 0)$ to impossible $(p = 0)$ or vice versa, and yet that is what is required for Y to depend on X (in the non-probabilistic sense defined above). Furthermore, one might want to allow for X being a parent of Y even if it *does not* meet this probabilistic requirement, because of a failure of *faithfulness* (Spirtes, Glymour, and Scheines 2001). This is consistent with our more permissive notion of a parent.

[6] For what it's worth, we also believe that Barbero's defense of this property is confused. He discusses exactly this example ($X = 1$ is Alice flipping a coin and $Y = 1$ is its landing heads) and he is explicit that the formula is a counterfactual statement, yet then says that it expresses what would happen if Alice flips the coin *again*, which is of course not a counterfactual statement.

[7] As we discuss in the Appendix, axiom D6 corresponds to the property that $\mathcal{G}_D$ is acyclic, which is a consequence of the acyclicity of $\mathcal{G}_M$. It is unclear whether it is possible to also express the acyclicity of $\mathcal{G}_M$ in the causal language.

6 Probabilistic Causal Models

As mentioned, DSEMs give rise to probabilistic DSEMs and – if Markovian – these in turn give rise to Causal Bayesian Networks – CBNs. (See (Pearl 2009, Sec. 1.4.2).) We here offer an initial study of the nondeterministic counterpart to this construction.

Definition 6.1: A *probabilistic causal model* (or *a PNSEM*) M is a triple $(\mathcal{S}, \mathcal{F}, \mathcal{G})$, where $\mathcal{S}$ is a signature, $\mathcal{G}$ is an acyclic directed graph such that there is one node for each variable in $\mathcal{S}$, and $\mathcal{F}$ defines a function that associates with each endogenous variable X a family of conditional probability distributions $P_X(X|\vec{Pa}_X)$ over $\mathcal{R}(X)$, giving the probability of the values of X in terms of the *parents* $\vec{Pa}_X$ of X as they appear in $\mathcal{G}$. Further, $\mathcal{F}$ associates with each exogenous variable X a probability distribution $P_X(X)$ over $\mathcal{R}(X)$, where $P_X(x) > 0$ for all $x \in \mathcal{R}(X)$.∎

A *solution* of P_X is a tuple $(x, \vec{pa}_X)$ such that $P_X(x|\vec{pa}_X) > 0$. We assume that PNSEMs satisfy the *Causal Markov Condition*, which means that we assume the joint distribution P_M over $\mathcal{R}(\mathcal{U} \times \mathcal{V})$ is given as: $P_M(\mathcal{U}, \mathcal{V}) = \prod_{X \in \mathcal{U} \cup \mathcal{V}} P_X(X|\vec{Pa}_X)$. (Here we abuse notation and write $P_X(X|\vec{Pa}_X)$ also for the exogenous case, with the understanding that $\vec{Pa}_X = \emptyset$.) A solution of M is a world $(\vec{u}, \vec{v})$ such that $P_M(\vec{u}, \vec{v}) > 0$.

In order to evaluate counterfactual probabilities we proceed analogously to the non-probabilistic case of Section 3. We extend the causal language $\mathcal{L}$ to a language $\mathcal{L}_{\mathcal{P}}$ by adding *probabilistic causal formulas* as follows: for each basic formula $\psi \in \mathcal{L}$ and each $p \in [0, 1]$, $\psi = p \in \mathcal{L}_{\mathcal{P}}$.

We then extend to non-basic formulas (i.e., Boolean combinations of basic formulas) in the standard way. We write $(M, \vec{u}, \vec{v}) \models \chi$ if the probabilistic causal formula χ is true in causal model M given world $(\vec{u}, \vec{v})$.

We first define the $\models$ relation for the case $\varphi = p$, where φ is a basic formula. $(M, \vec{u}, \vec{v}) \models \varphi = p$ if either $p = 1$ and $(M, \vec{u}, \vec{v}) \models \varphi$ or $p = 0$ and $(M, \vec{u}, \vec{v}) \not\models \varphi$.

We can define the $\models$ relation for basic probabilistic causal formulas by using the probabilistic counterparts of the two operations from Section 3, as follows.

Definition 6.2: Given a solution $(\vec{u}, \vec{v})$ of a probabilistic model $M = (\mathcal{S}, \mathcal{F}, \mathcal{G})$, we define the *actualized refinement* $M^{(\vec{u},\vec{v})}$ as the model M in which $\mathcal{F}$ is replaced by $\mathcal{F}^{(\vec{u},\vec{v})}$, as follows: for each variable $X \in (\mathcal{U} \cup \mathcal{V})$, its distribution P_X is replaced by $P_X^{(\vec{pa}_X,x)}(X|\vec{Pa}_X)$ that is identical to P_X for all $\vec{pa}_x{}'$ except for $\vec{pa}_X$. Instead, $P_X^{(\vec{pa}_X,x)}(x|\vec{pa}_X) = 1$, where x and $\vec{pa}_X$ are the respective restrictions of $(\vec{u}, \vec{v})$ to X and $\vec{Pa}_X$. ∎

Further, $M_{\vec{Y}\leftarrow\vec{y}}$ is the model that results from replacing P_Y with the unconditional point distribution that assigns probability 1 to $Y = y$ for each $Y \in \vec{Y}$, $y \in \vec{y}$ and replacing $\mathcal{G}$ with $\mathcal{G}^{\vec{Y}\leftarrow\vec{y}}$.

Finally, for any solution $(\vec{u}, \vec{v})$ of M, we define the $\models$ relation as follows. $(M, \vec{u}, \vec{v}) \models [\vec{Y} \leftarrow \vec{y}]\varphi = p$ iff $P_{M'}(\varphi) = p$, where $M' = (M^{(\vec{u},\vec{v})})_{\vec{Y}\leftarrow\vec{y}}$. We can also use the more standard notation, and write this as

$$P_M(\varphi|do(\vec{Y} \leftarrow \vec{y}), \vec{u}, \vec{v}) = p.$$

The extension to causal formulas is defined as before.

Definition 6.3: We say that a PNSEM M and a NSEM M' are *consistent* if both share the same signature and graph, and if for each $X \in \mathcal{V}$ we have that $P_X(x|\vec{pa}_X) > 0$ iff $x \in f_X(\vec{pa}_X)$. ∎

Importantly, the probabilistic **swc** logic over $\mathcal{L}_{\mathcal{P}}$ generalizes our original **swc** logic over $\mathcal{L}$, as it ought to do.

Theorem 6.4: *Given consistent models M and M^*, where M is probabilistic and M^* is not, for all worlds $(\vec{u}, \vec{v})$ and all basic causal formulas $\psi \in \mathcal{L}$ it holds that $(M, \vec{u}, \vec{v}) \models \psi = 1$ iff $(M^*, \vec{u}, \vec{v}) \models \psi$.*

Our computation of $P_M(\varphi|do(\vec{Y} \leftarrow \vec{y}), \vec{u}, \vec{v})$ is entirely analogous to Pearl's well-known 3-step procedure for computing counterfactual probabilities in DSEMs:

1. **Abduction** Update P_M by the evidence $(\vec{u}, \vec{v})$ to obtain $P_{M^{(\vec{u},\vec{v})}}$;
2. **Action** Modify this by the action $do(\vec{Y} \leftarrow \vec{y})$ to obtain $P_{M'}$;
3. **Prediction** Use $P_{M'}$ to compute the probability of φ.

Lastly, we briefly consider how this allows us to identify counterfactuals even in Causal Bayesian Networks. Recall that in the standard case, a PDSEM induces a CBN by marginalizing over the exogenous variables $\mathcal{U}$ under the assumption that the exogenous variables are independent and that no exogenous variable is a common parent of two endogenous ones. These assumptions together imply that the CBN satisfies the Markov Condition (Pearl 2009, Th. 1.4.1). In the nondeterministic case, we assumed the Markov Condition *itself*, at the level of the PNSEM. Other than that, in both cases we get that the marginal distribution over the endogenous variables also satisfies the Markov Condition:

$$P_M(\mathcal{V}) = \prod_{X\in\mathcal{V}} P'_X(X|\vec{EPa}_X)$$

where $\vec{EPa}_X$ are X's endogenous parents and P'_X is the result of marginalizing P_X over X's exogenous parents.[8]

Definition 6.5: Given a PNSEM $M = (\mathcal{S}, \mathcal{F}, \mathcal{G})$ such that no two endogenous variables share an exogenous parent, the *Causal Bayesian Network C_M* induced by M is the tuple $(\mathcal{S}_{\mathcal{V}}, P_{\mathcal{V}}, \mathcal{G}_{\mathcal{V}})$, where $\mathcal{S}_{\mathcal{V}}, \mathcal{G}_{\mathcal{V}}$ is obtained from $\mathcal{S}, \mathcal{G}$ by removing the exogenous variables, and $P_{\mathcal{V}}(\mathcal{V}) = P_M(\mathcal{V})$. ∎

Note that any CBN may be induced by either a PNSEM or a PDSEM, and thus at the level of CBNs it appears that nothing has changed. However, this is not true. Under the standard view, evaluating counterfactual probabilities cannot be done directly in a CBN, but instead requires computing them in the – unknown – underlying PDSEM. As a result, the standard orthodoxy is that they cannot in general be uniquely identified in CBNs. (See (Balke and Pearl 1994; Pearl 2009) for a study of their bounds.) Yet under the nondeterministic view, we *can* simply apply the procedure for computing the probability of a counterfactual directly to a CBN itself. Concretely, we can apply our 3-step procedure directly to the distribution $P_{\mathcal{V}}$, giving:

$$P_{\mathcal{V}}(\vec{V}|do(\vec{Y} \leftarrow \vec{y}), \vec{v}) = \prod_{X\in(\mathcal{V}-\vec{Y})} P_X'^{(\vec{epa}_X,x)}(X|\vec{EPa}_X)P^{\vec{Y}\leftarrow\vec{y}}(\vec{Y}).$$

Here $P^{\vec{Y}\leftarrow\vec{y}}$ is the point distribution assigning probability 1 to $\vec{y}$, and $P_X'^{(\vec{epa}_X,x)}$ results from applying Def. 6.2 to P'_X.

Although this computation itself does not require our nondeterministic framework (as it does not depend on knowledge of the underlying PNSEM), its explanation is possible only in our framework: it is the distribution that one would obtain for at least *some* set of all PNSEMs that could underlie the CBN, the extreme case being PNSEMs that have no exogenous variables at all. It remains to be investigated what other models belong to this set, and in what circumstances it is reasonable to assume that the true underlying model M belongs to it. We aim to do so in future work.

7 Conclusion

We here developed a nondeterministic generalization of causal models and offered an axiomatization for the two resulting counterfactual logics, arguing that their semantics are superior to other proposals. Crucially, they inherit all the benefits of Pearl's framework whilst dropping the controversial uniqueness property. We also initiated the probabilistic extension of our framework, but we consider it far from concluded. Among other things, we anticipate that it offers fruitful middle ground between the two camps disputing the role of counterfactuals in personalised decision-making.

[8] I.e., $P'_X(X|\vec{EPa}_X) = \sum_{\vec{w}\in\mathcal{R}(\vec{W})}(P_X(X|\vec{EPa}_X, \vec{w}) \prod_{w_i\in\vec{w}} P_{W_i}(w_i))$, where $\vec{W}$ are X's exogenous parents.

References

Balke, A.; and Pearl, J. 1994. Counterfactual Probabilities: Computational Methods, Bounds and Applications. In *Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence (UAI 1994)*, volume 10, 46–54.

Barbero, F. 2024. On the Logic of Interventionist Counterfactuals under Indeterministic Causal Laws. In *International Symposium on Foundations of Information and Knowledge Systems (FoIKS 2024)*, 203–221.

Beckers, S.; Halpern, J. Y.; and Hitchcock, C. 2023. Causal Models with Constraints. In *Proceedings of the 2nd Conference on Causal Learning and Reasoning (CLeaR 2023)*.

Bongers, S.; Forré, P.; Peters, J.; and Mooij, J. M. 2021. Foundations of Structural Causal Models with Cycles and Latent Variables. *The Annals of Statistics*, 49(5): 2885–2915.

Dawid, A. P. 2000. Causal Inference Without Counterfactuals (with Discussion). *Journal of the American Statistical Association*, 95: 465–480.

Dawid, A. P.; and Senn, S. 2023. Personalised Decision-Making without Counterfactuals. *arXiv preprint*, https://arxiv.org/abs/2301.11976.

Galles, D.; and Pearl, J. 1998. An Axiomatic Characterization of Causal Counterfactuals. *Foundations of Science*, 3(1): 151–182.

Halpern, J. Y. 2000. Axiomatizing Causal Reasoning. *Journal of Artificial Intelligence Research*, 12: 317–337.

Halpern, J. Y. 2016. *Actual Causality*. MIT Press.

Halpern, J. Y.; and Peters, S. 2022. Reasoning about Causal Models with Infinitely Many Variables. *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, 36(5668-5675).

Lewis, D. 1973. Causation. *Journal of Philosophy*, 70: 113–126.

Mueller, S.; and Pearl, J. 2023. Personalized Decision Making – A Conceptual Introduction. *Journal of Causal Inference*, 11(1).

Pearl, J. 2009. *Causality: Models, Reasoning, and Inference; 2nd edition*. Cambridge University Press.

Peters, S.; and Halpern, J. Y. 2021. Causal Modeling With Infinitely Many Variables. *arXiv preprint*, https://arxiv.org/abs/2112.09171.

Rubin, D. B. 1974. Estimating Causal Effects of Treatment in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66: 688–701.

Spirtes, P.; Glymour, C.; and Scheines, R. 2001. *Causation, Prediction, and Search*. MIT Press, 2nd edition edition.

Walters, L.; and Williams, J. R. G. 2013. An Argument for Conjunction Conditionalization. *The Review of Symbolic Logic*, 6(4): 573–588.

Wysocki, T. 2023. The Underdeterministic Framework. *British journal for the philosophy of science*, forthcoming.

A Appendix: Proofs of Theorems

Theorem A.1: *Given a nondeterministic causal model M, we have that for all $\vec{Y} \subseteq \mathcal{V}$, for all $\vec{y} \in \mathcal{R}(\vec{Y})$, and for all basic formulas φ:*

- *$M \models [\vec{Y} \leftarrow \vec{y}]\varphi$ iff $\vec{v} \models \varphi$ for all solutions $(\vec{u}, \vec{v})$ of $M_{\vec{Y} \leftarrow \vec{y}}$.*
- *For all contexts $\vec{u}$: $(M, \vec{u}) \models [\vec{Y} \leftarrow \vec{y}]\varphi$ iff $\vec{v} \models \varphi$ for all states $\vec{v}$ such that $(\vec{u}, \vec{v})$ is a solution of $M_{\vec{Y} \leftarrow \vec{y}}$.*

Proof: We first show that the two claims are equivalent. Per definition $M \models [\vec{Y} \leftarrow \vec{y}]\varphi$ iff $(M, \vec{u}, \vec{v}) \models [\vec{Y} \leftarrow \vec{y}]\varphi$ for all solutions $(\vec{u}, \vec{v})$ of M, which is equivalent to the statement that for all contexts $\vec{u}$: $(M, \vec{u}, \vec{v}) \models [\vec{Y} \leftarrow \vec{y}]\varphi$ for all states $\vec{v}$ such that $(\vec{u}, \vec{v})$ is a solution of M. In turn, the latter is per definition equivalent to the statement that for all contexts $\vec{u}$: $(M, \vec{u}) \models [\vec{Y} \leftarrow \vec{y}]\varphi$. By the second claim, the latter is equivalent to the statement that for all contexts $\vec{u}$: $\vec{v} \models \varphi$ for all states $\vec{v}$ such that $(\vec{u}, \vec{v})$ is a solution of $M_{\vec{Y} \leftarrow \vec{y}}$. Finally, this is equivalent to the right side of the first claim.

We now prove the first claim. Recall that, given a solution $(\vec{u}, \vec{v})$ of M, $(M, \vec{u}, \vec{v}) \models [\vec{Y} \leftarrow \vec{y}]\varphi$ iff $\vec{v}' \models \varphi$ for all states $\vec{v}'$ such that $(\vec{u}, \vec{v}')$ is a solution of $(M^{(\vec{u}, \vec{v})})_{\vec{Y} \leftarrow \vec{y}}$. Therefore we need to show that the following claims are equivalent:

- $\vec{v}' \models \varphi$ for all states $\vec{v}'$ for which there exists some solution $(\vec{u}, \vec{v})$ of M such that $(\vec{u}, \vec{v}')$ is a solution of $(M^{(\vec{u}, \vec{v})})_{\vec{Y} \leftarrow \vec{y}}$.
- $\vec{v}' \models \varphi$ for all solutions $(\vec{u}, \vec{v}')$ of $M_{\vec{Y} \leftarrow \vec{y}}$.

This equivalence follows directly from the statement that for all states $\vec{v}'$ and all contexts $\vec{u}$, the following claims are equivalent:

- $(\vec{u}, \vec{v}')$ is a solution of $(M^{(\vec{u}, \vec{v})})_{\vec{Y} \leftarrow \vec{y}}$ for some solution $(\vec{u}, \vec{v})$ of M.
- $(\vec{u}, \vec{v}')$ is a solution of $M_{\vec{Y} \leftarrow \vec{y}}$.

The implication from the first claim to the second is a direct consequence of the definition of an actualized refinement. Therefore we proceed with proving the reverse implication.

Assume $(\vec{u}, \vec{v}')$ is a solution of $M_{\vec{Y} \leftarrow \vec{y}}$. We need to show that there exists some $\vec{v}$ such that $(\vec{u}, \vec{v}')$ is a solution of $(M^{(\vec{u}, \vec{v})})_{\vec{Y} \leftarrow \vec{y}}$ and $(\vec{u}, \vec{v})$ is a solution of M. We proceed by induction over the partial order given by $\mathcal{G}_M$, by considering the claim: given an ancestrally closed set $\vec{W} \subseteq \mathcal{V}$, there exists some $\vec{w}$ such that $(\vec{u}, \vec{w}')$ appears in a solution of $(M^{(\vec{u}, \vec{w})})_{\vec{Y} \leftarrow \vec{y}}$ and the set of solutions extending $(\vec{u}, \vec{w})$ of M is non-empty. Here $\vec{w}'$ is the restriction of $\vec{v}'$ to $\vec{W}$, and $M^{(\vec{u}, \vec{w})}$ instantiates the obvious generalization of Definition 3.2 to ancestrally closed partial settings.

Given that we only consider total models, there exists at least one solution (to both M and to $M_{\vec{Y} \leftarrow \vec{y}}$) for each context $\vec{u}$, and thus the claim holds for the base case with $\vec{W} = \emptyset$.

We now proceed with the inductive case. Assume that the set of solutions of M extending $(\vec{u}, \vec{w})$ is non-empty and $(\vec{u}, \vec{w}')$ appears in some solution of $(M^{(\vec{u}, \vec{w})})_{\vec{Y} \leftarrow \vec{y}}$.

Say V is the next variable in $\mathcal{V} - \vec{W}$ according to the partial order of $\mathcal{G}_M$, and v' is its value in $\vec{v}'$. We need to prove that there exists some $v \in \mathcal{R}(V)$ so that the set of solutions of M extending $(\vec{u}, \vec{w}, v)$ is non-empty and such that $(\vec{u}, \vec{w}', v')$ appears in some solution of $(M^{(\vec{u},\vec{w},v)})_{\vec{Y}\leftarrow\vec{y}}$.

First we consider the case where $V \in \vec{Y}$. Since the set of solutions extending $(\vec{u}, \vec{w})$ is non-empty, there exists some v such that the set of solutions extending $(\vec{u}, \vec{w}, v)$ is non-empty as well. The equation for V in $(M^{(\vec{u},\vec{w},v)})_{\vec{Y}\leftarrow\vec{y}}$ will be the constant equation $V = v^*$, where v^* is the restriction of $\vec{y}$ to V. Furthermore, since $(\vec{u}, \vec{v}')$ is a solution of $M_{\vec{Y}\leftarrow\vec{y}}$, $v^* = v'$. It follows that $(\vec{u}, \vec{w}', v')$ appears in a solution of $(M^{(\vec{u},\vec{w},v)})_{\vec{Y}\leftarrow\vec{y}}$, as had to be shown.

Second we consider the case where $V \notin \vec{Y}$, which we separate into two sub-cases. Here $\vec{pa}_V{}'$ is the restriction of $\vec{v}'$ to $\vec{Pa}_V$. Note that, as $\vec{W}$ is ancestrally closed, $\vec{Pa}_V \subseteq \vec{W}$.

Consider the case where $\vec{pa}_V \neq \vec{pa}_V{}'$, where $\vec{pa}_V$ is the restriction of $\vec{w}$ to $\vec{Pa}_V$. As before, we can consider any v such that the set of solutions extending $(\vec{u}, \vec{w}, v)$ is non-empty. Per definition of an actualized refinement, we have that $f_V(\vec{pa}_V{}') = f_V^{(v,\vec{pa}_V)}(\vec{pa}_V{}')$. Furthermore, since $(\vec{u}, \vec{v}')$ is a solution of $M_{\vec{Y}\leftarrow\vec{y}}$, we have that $v' \in f_V(\vec{pa}_V{}')$, and thus also $v' \in f_V^{(v,\vec{pa}_V)}(\vec{pa}_V{}')$. It follows that $(\vec{u}, \vec{w}', v')$ appears in a solution of $(M^{(\vec{u},\vec{w},v)})_{\vec{Y}\leftarrow\vec{y}}$.

Lastly, consider the case where $\vec{pa}_V = \vec{pa}_V{}'$. As each $v \in f_V(\vec{pa}_V)$ is a solution of $V = f_V(\vec{Pa}_X)$, each solution of M that starts with $(\vec{u}, \vec{w})$ can be extended to a solution $(\vec{u}, \vec{w}, v)$ for each $v \in f_V(\vec{pa}_V)$. Therefore the set of solutions extending $(\vec{u}, \vec{w}, v)$ is non-empty for each $v \in f_V(\vec{pa}_V)$. Choosing $v = v'$, we get that there exists at least one solution that extends $(\vec{u}, \vec{w}, v)$, and $(\vec{u}, \vec{w}', v')$ appears in a solution of $(M^{(\vec{u},\vec{w},v)})_{\vec{Y}\leftarrow\vec{y}}$. This concludes the proof. ∎

Theorem A.2: *AX_{non}^{swc} (resp. AX_{non}^{scc}) is a sound and complete axiomatization for the language $\mathcal{L}(\mathcal{S})$ with respect to the* **swc** *logic (resp. the* **scc** *logic) over acyclic NSEMS that have signature $\mathcal{S}$.*

Proof: We start with completeness. First we consider AX_{non}^{swc} and the **swc** logic. It suffices to show that for any formula $\varphi \in \mathcal{L}(\mathcal{S})$ that is consistent with $AX_{non}^{swc}(\mathcal{S})$, there is an acyclic NSEM M such that $(M, \vec{u}, \vec{v}) \models \varphi$ for some causal setting $(M, \vec{u}, \vec{v})$. The proof follows the same technique as used by (Halpern 2000) and (Halpern and Peters 2022).

Suppose that $\varphi \in \mathcal{L}(\mathcal{S})$ is consistent with $AX_{non}^{swc}(\mathcal{S})$ (i.e., we cannot prove $\neg\varphi$ in $AX_{non}^{swc}(\mathcal{S})$). Then φ can be extended to a maximal consistent set C of formulas, meaning that $\varphi \in C$, every finite subset C' of C is consistent with $AX_{non}^{swc}(\mathcal{S})$, and no strict superset C^* of C has the property that every finite subset of C^* is consistent with $AX_{non}^{swc}(\mathcal{S})$. Standard arguments show that, for every formula $\psi \in \mathcal{L}(\mathcal{S})$, either ψ or $\neg\psi$ must be in C. Moreover, every instance of the axioms in $AX_{non}^{swc}(\mathcal{S})$ must be in C.

We now define an acyclic NSEM M with signature $\mathcal{S}$ and the requisite causal setting $(M, \vec{u}, \vec{v})$ as follows. Fix some context $\vec{u} \in \mathcal{R}(\mathcal{U})$. The following Lemma from (Halpern and Peters 2022) is useful, where AX_{bas} is the axiom system consisting of D0, D7, D8, and MP.

Lemma A.3:

(a) $AX_{bas} \vdash [\vec{Y} \leftarrow \vec{y}]\varphi_1 \wedge [\vec{Y} \leftarrow \vec{y}]\varphi_2 \Leftrightarrow [\vec{Y} \leftarrow \vec{y}](\varphi_1 \wedge \varphi_2)$

(b) $AX_{bas} \vdash \langle\vec{Y} \leftarrow \vec{y}\rangle(\varphi_1 \vee \varphi_2) \Leftrightarrow \langle\vec{Y} \leftarrow \vec{y}\rangle\varphi_1 \vee \langle\vec{Y} \leftarrow \vec{y}\rangle\varphi_2$

Applying Lemma A.3 and the axioms D1-2, D7, and D10(a) to the empty intervention, it follows by standard modal reasoning that there exists some $\vec{v} \in \mathcal{R}(\mathcal{V})$ so that $\langle\rangle\vec{V} = \vec{v} \in C$. We will construct M such that $(M, \vec{u}, \vec{v}) \models \varphi$.

For all $X \in \mathcal{V}$ we define $\vec{Pa}_X$ as the set of variables $Y \in \mathcal{V}$ such that $Y \rightsquigarrow X \in C$. Given D6, we have hereby defined an acyclic graph $\mathcal{G}$. We define F_X for $x \in \mathcal{R}(X)$ and $\vec{pa}_X \in \mathcal{R}(\vec{Pa}_X)$ by taking $x \in f_X(\vec{pa}_X)$ iff $\langle\vec{Pa}_X \leftarrow \vec{pa}_x\rangle X = x \in C$. As before, it follows from Lemma A.3 and the axioms D1-2, D7, and D10(a), that for each choice $\vec{pa}_X$ there will be at least one value x such that $x \in f_X(\vec{pa}_X)$, and thus our NSEM is total. Therefore M is an acyclic and total NSEM, as required.

Furthermore, $\langle\rangle\vec{V} = \vec{v} \in C$ combined with D10(c) gives $[]\vec{V} = \vec{v} \in C$. Applying D3(b) and Lemma A.3, for each X we get that $[\vec{Pa}_X \leftarrow \vec{pa}_X]X = x \in C$, where $\vec{pa}_X$ and x are the restrictions of $\vec{v}$ to their respective variables. Therefore the function f_X is deterministic for the actual parent values, (i.e, $x = f_X(\vec{pa}_X)$), and thus taking $f_X^{(\vec{pa}_X,x)}(\vec{pa}_X) = f_X(\vec{pa}_X)$ results in a deterministic actualized refinement for the actual parent values, as required for an NSEM. It remains to be shown that $(M, \vec{u}, \vec{v}) \models \varphi$.

Following exactly the same reasoning as in the proof of Theorem 5.2 in (Halpern and Peters 2022) and in (Beckers, Halpern, and Hitchcock 2023), it follows that this reduces to showing for all formulas of the form $\langle\vec{Y} \leftarrow \vec{y}\rangle\vec{X} = \vec{x}$ with $\vec{X} = \mathcal{V} - \vec{Y}$ that $\langle\vec{Y} \leftarrow \vec{y}\rangle\vec{X} = \vec{x} \in C$ iff $(M, \vec{u}, \vec{v}) \models \langle\vec{Y} \leftarrow \vec{y}\rangle\vec{X} = \vec{x}$.

Suppose that $\langle\vec{Y} \leftarrow \vec{y}\rangle\vec{X} = \vec{x} \in C$. Let $\vec{v'} = (\vec{y}, \vec{x})$. It suffices to show that for each $X \in \vec{X}$, $\langle\vec{Pa}_X \leftarrow \vec{pa}_x\rangle X = x \in C$, where $\vec{pa}_X$ and x are the restrictions of $\vec{v'}$ to their respective variables. Consider some X and the requisite values $\vec{pa}_X$ and x. By D3(a), we have that $\langle\vec{Pa}_X \leftarrow \vec{pa}_X, \vec{Z} \leftarrow \vec{z}\rangle X = x \in C$, where $\vec{Z} = \mathcal{V} - (\vec{Pa}_X \cup \{X\})$, and $\vec{z}$ is the restriction of $\vec{v'}$ to $\vec{Z}$.

Let us consider some $\vec{z''} \in \mathcal{R}(\vec{Z})$ such that $\langle\vec{Pa}_X \leftarrow \vec{pa}_X, X \leftarrow x\rangle(\vec{Z} = \vec{z''}) \in C$. (As before, such $z'' \in \mathcal{R}(Z)$ must exist.) If also $\langle\vec{Pa}_X \leftarrow \vec{pa}_X, \vec{Z} \leftarrow \vec{z''}\rangle X = x \in C$, then by D5 and D7 we get that $\langle\vec{Pa}_X \leftarrow \vec{pa}_X\rangle X = x \in C$, as required.

Remains to consider the case where $[\vec{Pa}_X \leftarrow \vec{pa}_X, \vec{Z} \leftarrow \vec{z''}]X \neq x \in C$. It follows that $\vec{z} \neq \vec{z''}$. Per construction of

$\vec{Pa}_X$, we have for each $Z \in \vec{Z}$ that $\neg(Z \rightsquigarrow X) \in C$. We show by induction that this results in a contradiction.

For the base case, take $\vec{A}_0 = \emptyset$, and $\vec{a}_0$ the restriction of $\vec{z}$ to $\vec{A}_0$. Let $\vec{W}_0 = \vec{Z} - \vec{A}_0$, and $\vec{w}_0, \vec{w''}_0$ the restrictions of $\vec{z}$ and $\vec{z''}$ to $\vec{W}_0$. We have that $[\vec{Pa}_X \leftarrow \vec{pa}_X, \vec{W}_0 \leftarrow \vec{w''}_0, \vec{A}_0 \leftarrow \vec{a}_0]X \neq x \in C$.

The inductive case consists of considering $\vec{A}_{k+1} = \vec{A}_k \cup \{Z\}$ for some $Z \in \vec{W}_k$. We let $\vec{W}_{k+1} = \vec{Z} - \vec{A}_{k+1}$, $\vec{a}_{k+1}$ is the restriction of $\vec{z}$ to $\vec{A}_{k+1}$, and $\vec{w}_{k+1}, \vec{w''}_{k+1}$ are the restrictions of $\vec{z}$ and $\vec{z''}$ to $\vec{W}_{k+1}$. By the induction hypothesis, we have that $[\vec{Pa}_X \leftarrow \vec{pa}_X, \vec{W}_k \leftarrow \vec{w''}_k, \vec{A}_k \leftarrow \vec{a}_k]X \neq x \in C$, which can be rewritten as $[\vec{Pa}_X \leftarrow \vec{pa}_X, \vec{W}_{k+1} \leftarrow \vec{w''}_{k+1}, \vec{A}_k \leftarrow \vec{a}_k, Z \leftarrow z'']X \neq x \in C$. If $\langle \vec{Pa}_X \leftarrow \vec{pa}_X, \vec{W}_{k+1} \leftarrow \vec{w''}_{k+1}, \vec{A}_k \leftarrow \vec{a}_k, Z \leftarrow z\rangle X = x \in C$, it follows that $Z \rightsquigarrow X \in C$. Therefore, $[\vec{Pa}_X \leftarrow \vec{pa}_X, \vec{W}_{k+1} \leftarrow \vec{w''}_{k+1}, \vec{A}_{k+1} \leftarrow \vec{a}_{k+1}]X \neq x \in C$. Given that $|\vec{Z}|$ is finite, for some $k \in \mathbb{N}$ this results in a contradiction.

For the other way, suppose that $(M, \vec{u}, \vec{v}) \models \langle \vec{Y} \leftarrow \vec{y}\rangle \vec{X} = \vec{x}$. Let $\vec{v'} = (\vec{y}, \vec{x})$. Per construction of M, we know that for each $X \in \vec{X}$, $\langle \vec{Pa}_X \leftarrow \vec{pa}_x\rangle X = x \in C$, where $\vec{pa}_X$ and x are the restrictions of $\vec{v'}$ to their respective variables.

For each $X \in \vec{X}$, we have that for any $\vec{Z} \subseteq \mathcal{V} - (\vec{Pa}_X \cup \{X\})$: $\neg(Z_i \rightsquigarrow X) \in C$ for all $Z_i \in \vec{Z}$. Therefore, for any $\vec{z}$ we have that $\langle \vec{Z} \leftarrow \vec{z}, \vec{Pa}_X \leftarrow \vec{pa}_x\rangle X = x \in C$. Letting $\vec{X} = \{X_1, \ldots, X_k\}$, we have in particular that for each $i \in \{1, \ldots, k\}$: $\langle \vec{Y} \leftarrow \vec{y}, \vec{X}^{-i} \leftarrow \vec{x}^{-i}\rangle X_i = x_i \in C$, where $\vec{X}^{-i} := (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k)$.

Taking $\langle \vec{Y} \leftarrow \vec{y}, \vec{X}^{-1} \leftarrow \vec{x}^{-1}\rangle X_1 = x_1 \in C$ and $\langle \vec{Y} \leftarrow \vec{y}, \vec{X}^{-2} \leftarrow \vec{x}^{-2}\rangle X_2 = x_2 \in C$, we can apply D5 to derive that $\langle \vec{Y} \leftarrow \vec{y}, \vec{X}^{-1,2} \leftarrow \vec{x}^{-1,2}\rangle (X_1 = x_1 \wedge X_2 = x_2) \in C$. By the same reasoning, we get that $\langle \vec{Y} \leftarrow \vec{y}, \vec{X}^{-2,3} \leftarrow \vec{x}^{-2,3}\rangle (X_2 = x_2 \wedge X_3 = x_3) \in C$. Again applying D5 to the last two statements, we get that $\langle \vec{Y} \leftarrow \vec{y}, \vec{X}^{-1,2,3} \leftarrow \vec{x}^{-1,2,3}\rangle (X_1 = x_1 \wedge X_2 = x_2 \wedge X_3 = x_3) \in C$. By straightforward induction, we get that $\langle \vec{Y} \leftarrow \vec{y}\rangle \vec{X} \leftarrow \vec{x} \in C$, which is what had to be shown. This concludes the proof of completeness.

For AX^{scc}_{non} and the **scc** logic, the proof proceeds identically except for three differences. The first difference is that we have no need for some $\vec{v} \in \mathcal{R}(\mathcal{V})$ so that $\langle\rangle \vec{V} = \vec{v} \in C$, as we only need to construct M and $\vec{u}$ such that $(M, \vec{u}) \models \varphi$. The second (related) difference is that in this case there is no need to consider some actual values $\vec{pa}_X$ for each $X \in \vec{X}$ and verify that $f_X(\vec{pa}_X)$ is deterministic, for the actualized refinement is not relevant to the semantics of $\models$ for $(M, \vec{u})$, as shown by Theorem A.1. This explains why D10(c) is not part of AX^{scc}_{non}. The third difference is that we here need to show $\langle \vec{Y} \leftarrow \vec{y}\rangle \vec{X} = \vec{x} \in C$ iff $(M, \vec{u}) \models \langle \vec{Y} \leftarrow \vec{y}\rangle \vec{X} = \vec{x}$ (as opposed to having $\vec{v}$ included on the RHS). By Theorem A.1, this is equivalent to showing that $\langle \vec{Y} \leftarrow \vec{y}\rangle \vec{X} = \vec{x} \in C$ iff $(\vec{u}, \vec{y}, \vec{x})$ is a solution of $M_{\vec{Y} \leftarrow \vec{y}}$. The remainder of the proof remains identical.

Now we prove soundness. We leave it as a simple exercise to the reader that D0-D1-D2-D4-D7-D8-D10(a) are sound for both of our logics, and that D10(c) is sound for our **swc** logic.

D6:

As we explain in the discussion of (Wysocki 2023)'s work in Section 5, contrary to our NSEMs, DSEMs do not come with a graph. Rather, a graph $\mathcal{G}_D$ is induced by invoking "Y depends on X": there is an edge from X to Y iff there exist settings $\vec{z} \in (\mathcal{U} \cup \mathcal{V} - \{X, Y\}$, and $x, x' \in \mathcal{R}(X)$, such that $f_Y(\vec{z}, x) \neq f_Y(\vec{z}, x')$. As mentioned in (Halpern and Peters 2022), D6 expresses the acyclicity of the induced graph $\mathcal{G}_D$. Of course NSEMs can be used to invoke a graph $\mathcal{G}_D$ in exactly the same manner: there is an edge from X to Y iff there exist settings $\vec{z} \in (\mathcal{U} \cup \mathcal{V} - \{X, Y\}$, and $x, x' \in \mathcal{R}(X)$, such that $f_Y(\vec{z}, x) \neq f_Y(\vec{z}, x')$. NSEMs also come with an explicit graph $\mathcal{G}_M$, and this need not be identical to $\mathcal{G}_D$. But the graph $\mathcal{G}_D$ is easily seen to be a subgraph of $\mathcal{G}_M$, since X has to be an argument of f_Y for Y to depend on X, and per definition this means that there is an edge from X to Y in $\mathcal{G}_M$. Therefore the acyclicity of $\mathcal{G}_M$ implies the acyclicity of $\mathcal{G}_D$, and thus D6 is sound for both of our logics. Concretely: if $X \rightsquigarrow Y$ for some solution $(\vec{u}, \vec{v})$, then X has to be an ancestor of Y in $\mathcal{G}_D$. So the falsity of D6 would imply that $\mathcal{G}_D$ is cyclic.

D3(a):

We start with soundness for $\models$ relative to $(M, \vec{u}, \vec{v})$. Assume $W \notin \vec{X}$, and $(\vec{u}, \vec{v})$ is a solution of M, and $(M, \vec{u}, \vec{v}) \models \langle \vec{X} \leftarrow \vec{x}\rangle (W = w \wedge \varphi)$. This means that there exists $\vec{v'}$ so that $(\vec{u}, \vec{v'})$ is a solution of $(M^{(\vec{u}, \vec{v})})_{\vec{X} \leftarrow \vec{x}}$ and $\vec{v'} \models (W = w \wedge \varphi)$. By D4, it directly follows that $(\vec{u}, \vec{v'})$ is also a solution of $(M^{(\vec{u}, \vec{v})})_{\vec{X} \leftarrow \vec{x}, W \leftarrow w}$, and we also have that $\vec{v'} \models \varphi$. This means precisely that $(M, \vec{u}, \vec{v}) \models \langle \vec{X} \leftarrow \vec{x}, W \leftarrow w\rangle \varphi$.

Soundness for $\models$ relative to $(M, \vec{u})$ is a consequence of the soundness for $\models$ relative to $(M, \vec{u}, \vec{v})$. For concreteness, we here write out the intermediate steps. Assume that $(M, \vec{u}) \models \langle \vec{X} \leftarrow \vec{x}\rangle (W = w \wedge \varphi)$. Per definition of $\langle\rangle$, this is equivalent to: $(M, \vec{u}) \models \neg[\vec{X} \leftarrow \vec{x}](W \neq w \vee \neg\varphi)$. In turn, this is equivalent to it not being the case that for all states $\vec{v}$ so that $(\vec{u}, \vec{v})$ is a solution of M, we have $(M, \vec{u}, \vec{v}) \models [\vec{X} \leftarrow \vec{x}](W \neq w \vee \neg\varphi)$. This is equivalent to there existing some state $\vec{v'}$ such that $(\vec{u}, \vec{v'})$ is a solution of M and $(M, \vec{u}, \vec{v'}) \models \langle \vec{X} \leftarrow \vec{x}\rangle (W = w \wedge \varphi)$. By D3(a) for $\models$ relative to causal settings, we get that for some solution $(\vec{u}, \vec{v''})$ of M, namely $(\vec{u}, \vec{v'})$, $(M, \vec{u}, \vec{v''}) \models \langle \vec{X} \leftarrow \vec{x}, W \leftarrow w\rangle \varphi$. Applying all of the above equivalences in the other direction, this is seen to be equivalent to $(M, \vec{u}) \models \langle \vec{X} \leftarrow \vec{x}, W \leftarrow w\rangle \varphi$, which is what had to be shown.

D3(b):

Soundness for $\models$ relative to $(M, \vec{u})$ is a direct consequence of the soundness for $\models$ relative to $(M, \vec{u}, \vec{v})$, so we proceed with the latter.

Assume $W \notin \vec{X}$, and $(\vec{u}, \vec{v})$ is a solution of M, and $(M, \vec{u}, \vec{v}) \models [\vec{X} \leftarrow \vec{x}](W = w \wedge \varphi)$. It suffices to show that for any state $\vec{v}'$, if $(\vec{u}, \vec{v}')$ is a solution of $(M^{(\vec{u},\vec{v})})_{\vec{X}\leftarrow\vec{x},W\leftarrow w}$, then it is a solution of $(M^{(\vec{u},\vec{v})})_{\vec{X}\leftarrow\vec{x}}$.

We proceed by a reductio. Assume that $(\vec{u}, \vec{v}')$ is a solution only of the former. Note that the two models have identical equations for all variables except W, and that the restriction of $\vec{v}'$ to W has to be w. Therefore it must be that $(w, \vec{pa_W})$ is not a solution of W's equation in $(M^{(\vec{u},\vec{v})})_{\vec{X}\leftarrow\vec{x}}$, where $\vec{pa_W}$ is the restriction of $\vec{v}'$ to $\vec{Pa_W}$. Given that, per the first assumption, $W = w$ for all solutions of $(M^{(\vec{u},\vec{v})})_{\vec{X}\leftarrow\vec{x}}$, it must be that $\vec{pa_W}$ does not appear in any such solution. However, given that both models are acyclic and have identical equations for all non-descendants of W (including $\vec{pa_W}$), they have exactly the same partial solutions $(\vec{u}, \vec{a})$, where $\vec{A} \subseteq \mathcal{V}$ consists of all non-descendants of W. Thus it cannot be that $\vec{pa_W}$ only appears in a solution to one of them.

D5:

We start with soundness for $\models$ relative to $(M, \vec{u}, \vec{v})$. Assume $(\vec{u}, \vec{v})$ is a solution of M, and $(M, \vec{u}, \vec{v}) \models \langle \vec{X} \leftarrow \vec{x}, Y \leftarrow y\rangle(W = w \wedge \vec{Z} = \vec{z}) \wedge \langle \vec{X} \leftarrow \vec{x}, W \leftarrow w\rangle(Y = y \wedge \vec{Z} = \vec{z})$, where $\vec{Z} = \mathcal{V} - (\vec{X} \cup \{W, Y\})$. We need to show that $(M, \vec{u}, \vec{v}) \models \langle \vec{X} \leftarrow \vec{x}\rangle(W = w \wedge Y = y \wedge \vec{Z} = \vec{z})$.

Per assumption, there exists a solution $(\vec{u}, \vec{v_1})$ of $(M^{(\vec{u},\vec{v})})_{\vec{X}\leftarrow\vec{x},Y\leftarrow y}$ such that its restriction to $(W, \vec{Z})$ is $(w, \vec{z})$, and there exists a solution $(\vec{u}, \vec{v_2})$ of $(M^{(\vec{u},\vec{v})})_{\vec{X}\leftarrow\vec{x},W\leftarrow w}$ such that its restriction to $(Y, \vec{Z})$ is $(y, \vec{z})$. Clearly also both solutions have that $\vec{X} = \vec{x}$, and the restriction of $\vec{v_1}$ to Y must be y, whereas the restriction of $\vec{v_2}$ to W must be w.

Since $\mathcal{V} = \vec{Z} \cup \vec{X} \cup \{W, Y\}$, the tuple $\vec{v_3} = (\vec{x}, \vec{z}, y, w)$ is a state. Therefore it follows that $\vec{v_1} = \vec{v_2} = \vec{v_3}$. It now suffices to show that $(\vec{u}, \vec{v_3})$ is a solution of $(M^{(\vec{u},\vec{v})})_{\vec{X}\leftarrow\vec{x}}$. As the equations for all variables in $\vec{Z} \cup \vec{X}$ are identical across the three models, $(\vec{u}, \vec{v_3})$ contains solutions to all these equations in $(M^{(\vec{u},\vec{v})})_{\vec{X}\leftarrow\vec{x}}$. As the equation for Y is identical across $(M^{(\vec{u},\vec{v})})_{\vec{X}\leftarrow\vec{x}}$ and $(M^{(\vec{u},\vec{v})})_{\vec{X}\leftarrow\vec{x},W\leftarrow w}$, $(\vec{u}, \vec{v_3})$ contains a solution for Y in $(M^{(\vec{u},\vec{v})})_{\vec{X}\leftarrow\vec{x}}$. The same holds for W and the other model, concluding the proof.

Now we prove soundness for $\models$ relative to $(M, \vec{u})$. Assume that $(M, \vec{u}) \models \langle \vec{X} \leftarrow \vec{x}, Y \leftarrow y\rangle(W = w \wedge \vec{Z} = \vec{z}) \wedge \langle \vec{X} \leftarrow \vec{x}, W \leftarrow w\rangle(Y = y \wedge \vec{Z} = \vec{z})$, where $\vec{Z} = \mathcal{V} - (\vec{X} \cup \{W, Y\})$. We need to show that $(M, \vec{u}) \models \langle \vec{X} \leftarrow \vec{x}\rangle(W = w \wedge Y = y \wedge \vec{Z} = \vec{z})$. By Theorem A.1, our assumption is equivalent to the statement that $(\vec{u}, \vec{x}, \vec{z}, y, w)$ is a solution of both $M_{\vec{X}\leftarrow\vec{x},Y\leftarrow y}$ and $M_{\vec{X}\leftarrow\vec{x},W\leftarrow w}$, and what we need to show is that $(\vec{u}, \vec{x}, \vec{z}, y, w)$ is a solution of $M_{\vec{X}\leftarrow\vec{x}}$. This follows by applying the reasoning from the previous paragraph to the three models at hand.

∎

Proposition A.4: *For a solution $(\vec{u}, \vec{v})$ of M, if $(M, \vec{u}, \vec{v}) \models \vec{X} = \vec{x} \wedge \varphi$ then $(M, \vec{u}, \vec{v}) \models [\vec{X} \leftarrow \vec{x}]\varphi$.*

Proof: This follows directly by induction using D3(b) for $\vec{X} = \emptyset$ and the fact that φ is equivalent to $[]\varphi$. ∎

Theorem A.5: *Given consistent models M and M^*, where M is probabilistic and M^* is not, for all basic causal formulas $\psi \in \mathcal{L}$ it holds that $(M, \vec{u}, \vec{v}) \models \psi = 1$ iff $(M^*, \vec{u}, \vec{v}) \models \psi$.*

Proof: Assume we have a probabilistic model M_1 and a non-probabilistic model M_2 as described, and consider a basic causal formula $[\vec{Y} \leftarrow \vec{y}]\varphi$ and a world $(\vec{u}, \vec{v})$. First we show that $(\vec{u}, \vec{v})$ is a solution of M_1 iff $(\vec{u}, \vec{v})$ is a solution of M_2.

$(\vec{u}, \vec{v})$ is a solution of M_1 iff $P_{M_1}(\vec{u}, \vec{v}) > 0$ iff $\prod_{X\in\mathcal{U}\cup\mathcal{V}} P_X(x|\vec{pa_X}) > 0$ iff for each $X \in \mathcal{V}$, $P_X(x|\vec{pa_X}) > 0$. (Throughout x and $\vec{pa_X}$ are the respective restrictions of $(\vec{u}, \vec{v})$.) By definition of consistency, the latter is equivalent to: for each $X \in \mathcal{V}$, $x \in f_X(\vec{pa_X})$ in M_2, and this in turn is equivalent to $(\vec{u}, \vec{v})$ being a solution of M_2.

What remains to be shown, is that $\vec{v}' \models \varphi$ for all states $\vec{v}'$ such that $(\vec{u}, \vec{v}')$ is a solution of $(M_2^{(\vec{u},\vec{v})})_{\vec{Y}\leftarrow\vec{y}}$ iff $P_{M'}(\varphi) = 1$, where $M' = (M_1^{(\vec{u},\vec{v})})_{\vec{Y}\leftarrow\vec{y}}$. This can be expressed equivalently as: there exists a state $\vec{v}'$ such that $\vec{v}' \models \neg\varphi$ and $(\vec{u}, \vec{v}')$ is a solution of $(M_2^{(\vec{u},\vec{v})})_{\vec{Y}\leftarrow\vec{y}}$ iff $P_{M'}(\neg\varphi) > 0$. The last statement simply means that there exists a state $\vec{v}'$ such that $\vec{v}' \models \neg\varphi$ and $(\vec{u}, \vec{v}')$ is a solution of M'. Therefore the results follows if we can show that for each state $\vec{v}'$, $(\vec{u}, \vec{v}')$ is a solution of $(M_2^{(\vec{u},\vec{v})})_{\vec{Y}\leftarrow\vec{y}}$ iff $(\vec{u}, \vec{v}')$ is a solution of M'. This follows from our previous result, if $(M_2^{(\vec{u},\vec{v})})_{\vec{Y}\leftarrow\vec{y}}$ is consistent with M'. It is an easy consequence of the definitions that this is the case. ∎