Proceedings of Machine Learning Research vol 275:514–532, 2025 4th Conference on Causal Learning and Reasoning

Actual Causation and Nondeterministic Causal Models

Sander Beckers SREKCEBREDNAS@GMAIL.COM
Cornell University

Editors: Biwei Huang and Mathias Drton

Abstract

In (Beckers, 2025) I introduced nondeterministic causal models as a generalization of Pearl's standard deterministic causal models. I here take advantage of the increased expressivity offered by these models to offer a novel definition of *actual causation* (that also applies to deterministic models). Instead of motivating the definition by way of (often subjective) intuitions about examples, I proceed by developing it based entirely on the unique function that it can fulfil in communicating and learning a causal model. First I generalize the more basic notion of *counterfactual dependence*, second I show how this notion has a vital role to play in the logic of causal discovery, third I introduce the notion of a *structural simplification* of a causal model, and lastly I bring both notions together in my definition of actual causation. Although novel, the resulting definition arrives at verdicts that are almost identical to those of my previous definition (Beckers, 2021, 2022).

Keywords: nondeterminism; counterfactuals; actual causation; counterfactual dependence

1. Introduction

In (Beckers, 2025) I introduced *Nondeterministic Structural Causal Models* (NSCMs) as a generalization of Pearl (2009)'s influential Deterministic Structural Causal Models (DSCMs). DSCMs form the mathematical bedrock of Pearl's widely influential causal modeling framework. These causal models are deterministic in the sense that they assume the values of causal parents uniquely determine the values of their children, both actually and counterfactually. As I argue in (Beckers, 2025), both assumptions are unnecessarily restrictive, and nondeterministic causal models are the result of dropping them.

I here take advantage of the increased expressivity offered by NSCMs to offer a novel definition of *actual causation* (that applies to both NSCMs and DSCMs), which is the causal relation that obtains between specific instances of events, such as when a specific patient's taking aspirin causes their headache to disappear. Not only has this concept been notoriously hard to define (in comparison to other causal relations, such as the average causal effect, for example), it has also proven difficult to objectively evaluate the quality of different proposed definitions (Glymour et al., 2010; Halpern, 2016; Beckers, 2021). Instead of starting from paradigmatic test cases that a definition should "get right", I here develop a definition by focussing exclusively on the vital function that statements of actual causation fulfil in efficiently communicating and learning a causal model.

The paper proceeds as follows. The next section formally introduces NSCMs (Sec. 2), followed by an informal summary and motivation for the main idea behind my definition of actual causation (Sec. 3). This sets the stage for taking all the necessary steps towards formally developing this idea, starting with an investigation of the more basic concept of *counterfactual dependence* in the context of NSCMs (Sec. 4). This reveals a hitherto ignored notion of *interventional dependence* (Sec. 4.1) that can play a crucial role in the causal discovery of causal models from an idealized logical perspective (Sec. 5). I then introduce the idea of structurally simplifying a causal model (Sec. 5.1), to arrive at a definition of actual causation as the presence of counterfactual dependence in some *structural simplification* of a causal model (Sec. 6).

© 2025 S. Beckers.

2. Nondeterministic Structural Causal Models[1]

Definition 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 $\boldsymbol{X} = (X_1, \ldots, X_n)$, $\mathcal{R}(\boldsymbol{X})$ denotes the Cartesian product $\mathcal{R}(X_1) \times \cdots \times \mathcal{R}(X_n)$.*

Nondeterministic models generalize Halpern (2000, 2016)'s definition of deterministic causal models by using multi-valued functions. 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 *A* causal model *(or a* Nondeterministic Structural Causal Model – NSCM*) M is a triple $(\mathcal{S}, \mathcal{F}, \mathcal{G})$, where $\mathcal{S}$ is a signature, $\mathcal{G}$ is a directed acyclic graph – a DAG – such that there is one node for each variable in $\mathcal{S}$, and $\mathcal{F}$ associates with each endogenous variable X a* structural equation *F_X that takes on the form $X = f_X(\boldsymbol{Pa_X})$, where $\boldsymbol{Pa_X} \subseteq (\mathcal{U} \cup \mathcal{V})$ are the* parents *of X in $\mathcal{G}$, and $f_X : \mathcal{R}(\boldsymbol{Pa_X}) \to \mathcal{P}(\mathcal{R}(X))_\emptyset$. (Here $\mathcal{P}(\mathcal{R}(X))_\emptyset$ is the* powerset *of $\mathcal{R}(X)$: the set that contains as its elements all subsets of $\mathcal{R}(X)$, except for $\emptyset$.)*

It will prove useful to adopt Halpern's – by now standard – graphical convention that for any *deterministic* equation $Y = f_Y(\boldsymbol{Pa_Y})$, Y functionally depends on each of its parents $X \in \boldsymbol{Pa_Y}$, where functional dependence means that there exists some setting $\boldsymbol{z}$ of all the other parents and values x, x' such that $f_Y(\boldsymbol{z}, x) \neq f_Y(\boldsymbol{z}, x')$.

The exogenous variables $\mathcal{U}$ are taken to represent the background conditions that are simply given. We call $\boldsymbol{u} \in \mathcal{R}(\mathcal{U})$ a *context*, $\boldsymbol{v} \in \mathcal{R}(\mathcal{V})$ a *state*, a $(\boldsymbol{u}, \boldsymbol{v}) \in \mathcal{R}(\mathcal{U} \cup \mathcal{V})$ is a *world*, and $(M, \boldsymbol{u}, \boldsymbol{v})$ a *causal setting*. 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 $\boldsymbol{pa_X}$. In nondeterministic models, a *solution* of the equation $X = f_X(\boldsymbol{Pa_X})$ is a tuple $(x, \boldsymbol{pa_X})$ such that $x \in f_X(\boldsymbol{pa_X})$. A solution of M is a world $(\boldsymbol{u}, \boldsymbol{v})$ that is a solution of all equations in $\mathcal{F}$.

2.1. The Causal Language and Semantics

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 $\boldsymbol{Y} \leftarrow \boldsymbol{y}$, 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 $[\boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$. The special case where $k = 0$ is abbreviated as φ.[2] 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 setting. We write $(M, \boldsymbol{u}, \boldsymbol{v}) \models \psi$ if the causal formula ψ is true in causal model M given world $(\boldsymbol{u}, \boldsymbol{v})$.

We first define the $\models$ relation for basic formulas. $(M, \boldsymbol{u}, \boldsymbol{v}) \models X = x$ if x is the restriction of $\boldsymbol{v}$ 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 $\boldsymbol{v}$, and thus we often also write $\boldsymbol{v} \models \varphi$.

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

1. This section is based almost ad verbatim on (Beckers, 2025).
2. This is warranted because φ and $[]\varphi$ turn out to be equivalent, just as was the case for Halpern (2000, 2016).

as observed in a world $(\boldsymbol{u}, \boldsymbol{v})$ 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. Given $\boldsymbol{X}$ and $\boldsymbol{Y}$, we let $\boldsymbol{X}_{\boldsymbol{Y}}$ denote the restriction of $\boldsymbol{X}$ to $\boldsymbol{X} \cap \boldsymbol{Y}$, and thus $\boldsymbol{x}_{\boldsymbol{Y}} \in \mathcal{R}(\boldsymbol{X} \cap \boldsymbol{Y})$.

Definition 3 *Given a solution $(\boldsymbol{u}, \boldsymbol{v})$ of a model $M = (\mathcal{S}, \mathcal{F}, \mathcal{G})$, we define the* actualized refinement *$M^{(\boldsymbol{u},\boldsymbol{v})}$ as the model in which $\mathcal{F}$ is replaced by $\mathcal{F}^{(\boldsymbol{u},\boldsymbol{v})}$, as follows: for each variable $X \in \mathcal{V}$, its function f_X is replaced by $f_X^{(\boldsymbol{pa}_{\boldsymbol{X}},x)}$ that behaves identically to f_X for all inputs except for $\boldsymbol{pa}_{\boldsymbol{X}}$. Instead, $f_X^{(\boldsymbol{pa}_{\boldsymbol{X}},x)}(\boldsymbol{pa}_{\boldsymbol{X}}) = x$. Here $x = (\boldsymbol{u}, \boldsymbol{v})_X$ and $\boldsymbol{pa}_{\boldsymbol{X}} = (\boldsymbol{u}, \boldsymbol{v})_{\boldsymbol{Pa}_{\boldsymbol{X}}}$.*

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

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

We define $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle \varphi$ as an abbreviation of $\neg[\boldsymbol{Y} \leftarrow \boldsymbol{y}]\neg\varphi$. So $(M, \boldsymbol{u}, \boldsymbol{v}) \models \langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle \varphi$ iff $\boldsymbol{v}' \models \varphi$ for some solution $(\boldsymbol{u}, \boldsymbol{v}')$ of $(M^{(\boldsymbol{u},\boldsymbol{v})})_{\boldsymbol{Y} \leftarrow \boldsymbol{y}}$.

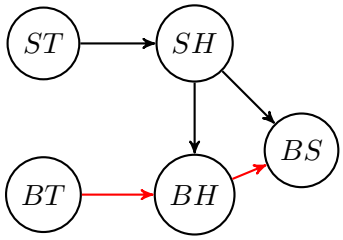
We can also evaluate formulas with respect to a partial causal setting $(M, \boldsymbol{u})$, or with respect to a model M. For basic causal formulas, $(M, \boldsymbol{u}) \models [\boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$ iff $(M, \boldsymbol{u}, \boldsymbol{v}) \models [\boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$ holds for all states $\boldsymbol{v}$ such that $(\boldsymbol{u}, \boldsymbol{v})$ is a solution of M. In a similar fasion, we define that $M \models [\boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$ iff $(M, \boldsymbol{u}, \boldsymbol{v}) \models [\boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$ holds for all solutions $(\boldsymbol{u}, \boldsymbol{v})$ of M. We again inductively extend to causal formulas in the standard way: $(M, \boldsymbol{u}) \models [\boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi_1 \wedge [\boldsymbol{Z} \leftarrow \boldsymbol{z}]\varphi_2$ iff $(M, \boldsymbol{u}) \models [\boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi_1$ and $(M, \boldsymbol{u}) \models [\boldsymbol{Z} \leftarrow \boldsymbol{z}]\varphi_2$, 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 $(M, \boldsymbol{u}, \boldsymbol{v})$, the second with respect to partial causal settings $(M, \boldsymbol{u})$, and the third with respect to M. I refer the reader to (Beckers, 2025) for a discussion and comparison of these semantics.

3. Informal Summary and Motivation

In order to get a sense of where we are heading, I here present the informal idea behind the new definition of actual causation that will be developed. Say we observe two events $X = x$ and $Y = y$. The most popular approach to defining what it means for $X = x$ to be an actual cause of $Y = y$ relative to a causal setting $(M, \boldsymbol{u}, \boldsymbol{v})$ is to construct it as a more nuanced version of *counterfactual dependence*: if X had been x' instead of x, then Y might have been y' instead of y, for some $x' \neq x$ and $y' \neq y$. The reason that we need to add nuance is that there exist compelling examples illustrating that counterfactual dependence is not necessary for actual causation (despite it arguably being sufficient). Preemption cases are by far the most famous such examples, and therefore I use a preemption classic that is discussed by Halpern (2016) at length to illustrate the problem.

Example 1 (Late Preemption) *Suzy and Billy both throw a rock at a bottle. Suzy's rock gets there first, shattering the bottle. However Billy's throw was also accurate, and would have shattered the bottle had it not been preempted by Suzy's throw. (See Figure 1.)*



$\mathcal{F}$:
$ST = 1$
$BT = 1$
$SH = ST$
$BH = BT \wedge \neg SH$
$BS = SH \vee BH.$

$\mathcal{S}$:
ST: Suzy Throwing
BT: Billy Throwing
SH: Suzy Hitting
BH: Billy Hitting
BS: Bottle Shattering.

Figure 1: M for Example 1 (Removing the red edges results in structural simplifications of $M^{\boldsymbol{v}}$.)

Intuitively, Suzy's throw caused the bottle to shatter and Billy's throw did not. Yet if Suzy had not thrown, the bottle would have shattered nonetheless, and thus there is no counterfactual dependence. Note that Billy's throw serves as a backup process here, and if it weren't for that backup process, then there would be counterfactual dependence. So a natural idea to pursue for defining actual causation is to consider whether there would be counterfactual dependence if the backup process were somehow ignored. To implement this idea requires answering two questions: what exactly does it mean to "ignore" the backup process, and why should we be at all interested in considering counterfactual dependence in such a hypothetical situation of ignorance, given that the causal model is as it is? So far the debate on defining actual causation has centered exclusively around the various technical answers to the first question, and almost no attention has been paid to the second question. (Two notable exceptions are (Hitchcock, 2017; Beckers, 2022).) Instead, I propose that answering the second question is what should guide us towards answering the first, for it is only by understanding what *function* the notion of actual causation fulfils that we can get a handle on formally defining it. In fact, as we will see, addressing the second question offers insights into the logic of causal discovery that are of interest quite aside from the issue of actual causation.

Coming back to our example, what if we are not fully confident about all details of the causal model, but rather, the model as we outlined it is merely the simplest model that is consistent with all the available observational and experimental data? Erring on the side of caution, in that case we should consider the possibility that the real model allows for a richer set of data. In particular, as it turns out, we can systematically increase the set of data consistent with the model by simplifying the structure of the model, and thus any such simplification represents a model that we still consider plausible. It's just that the more we simplify the structure the less plausible the model becomes.

More specifically, given a causal setting $(M, \boldsymbol{u}, \boldsymbol{v})$, we should extend our interest from relations of counterfactual dependence relative to $(M, \boldsymbol{u}, \boldsymbol{v})$ to such relations relative to any $(M^{simp}, \boldsymbol{u}, \boldsymbol{v})$, where M^{simp} is an appropriate simplification of M. (As we will see in Section 4, that relations of counterfactual dependence are themselves of interest is not hard to show.) I claim that actual causation consists precisely of all relations of counterfactual dependence in such simplifications.

Informally, these simplifications are the result of removing edges from $\mathcal{G}_M$ in such a way that (1) each edge removal is also a removal of an ancestor-descendant pair; (2) no cases of counterfactual dependence w.r.t. $(\boldsymbol{u}, \boldsymbol{v})$ are removed; and (3) the increase in cases of counterfactual dependence w.r.t. $(\boldsymbol{u}, \boldsymbol{v})$ is minimal with respect to the resulting graph.

As it turns out, in Example 1 the only edges that can be so removed are those in red. In this case, ignoring the backup process coincides with removing the edge $BH \rightarrow BS$, which results in the equation for BS taking on the nondeterministic form: $BS = 1$ if $SH = 1$ else $BS \in \{0, 1\}$.

(Roughly: BH's role in the equation is replaced with nondeterminism.) According to that equation, we get that if Suzy had not thrown, then the bottle might have not shattered, and hence Suzy's throw counts as an actual cause of the shattering. We revisit this example in detail in Section 6.

4. Counterfactual Dependence

At the heart of the interventionist approach to causation that was mathematically developed by Pearl (2009) and philosophically developed by Woodward (2003) lies the general idea that for X to be a cause of Y means for X to be a tool that allows one a certain amount of control over Y: in some situation, interventions on X from one value to another possibly result in changes to the value of Y. In other words, X is a cause of Y if in some situation Y depends on X, where dependence is to be understood in terms of interventions. *Counterfactual dependence* amounts to making this idea precise when considering a particular, actual, situation, in which the causal mechanisms have run their course and produced particular values for all the variables. Within the context of causal models, counterfactual dependence is thus a statement about what is the case in an actual world $(\boldsymbol{u}, \boldsymbol{v})$ that is governed by a causal model M, i.e., it is relative to a causal setting $(M, \boldsymbol{u}, \boldsymbol{v})$.

I first present the standard definition of counterfactual dependence as it appears in Halpern (2016), where it was defined for deterministic causal models, meaning that we are in a situation of full control: intervening to change X from its actual value to a specific other value results in *certainly* changing the value of Y. In the deterministic case a partial causal setting $(M, \boldsymbol{u})$ has only one solution $(\boldsymbol{u}, \boldsymbol{v})$, and therefore it can be identified with the full causal setting $(M, \boldsymbol{u}, \boldsymbol{v})$. We label this notion *dependence** to indicate its preliminary nature. (We assume throughout that $X \neq Y$.)

Definition 4 $Y = y$ depends* *on* $X = x$ *in* $(M, \boldsymbol{u})$ *if there exists* $x' \in \mathcal{R}(X)$ *s.t.:*
[CD1.] $(M, \boldsymbol{u}) \models \langle\rangle(X = x \wedge Y = y)$ *and* [CD2.] $(M, \boldsymbol{u}) \models [X \leftarrow x']Y \neq y$.

In the deterministic case, Def. 4 states informally that: (CD1) $X = x$ and $Y = y$ are the values of X and Y in the actual world, and (CD2) if X had been x', then Y would certainly have taken on a non-actual value. In the nondeterministic case, Def. 4 states that *possibly*, there is dependence, in the sense just described. Indeed, we have the following trivial proposition.

Proposition 5 *If* $Y = y$ *depends* on* $X = x$ *in* $(M, \boldsymbol{u})$ *then there exists some* $\boldsymbol{v} \in \mathcal{R}(\mathcal{V})$ *and* $x' \in \mathcal{R}(X)$ *s.t.* $(M, \boldsymbol{u}, \boldsymbol{v}) \models (X = x \wedge Y = y)$ *and* $(M, \boldsymbol{u}, \boldsymbol{v}) \models [X \leftarrow x']Y \neq y$.

The idea of dependence that we started out with above, however, merely requires the *possibility* of changing the value of Y. Therefore the generalization of counterfactual dependence to nondeterministic models is as follows.

Definition 6 $Y = y$ counterfactually depends* *on* $X = x$ *in* $(M, \boldsymbol{u}, \boldsymbol{v})$ *if there exists* $x' \in \mathcal{R}(X)$ *s.t.:*
[CD1.] $(M, \boldsymbol{u}, \boldsymbol{v}) \models (X = x \wedge Y = y)$ *and* [CD2.] $(M, \boldsymbol{u}, \boldsymbol{v}) \models \langle X \leftarrow x'\rangle Y \neq y$.

In the acyclic deterministic case, Definitions 4 and 6 are easily seen to be equivalent, whereas in the acyclic nondeterministic case, Definition 4 implies Definition 6 (for some $\boldsymbol{v}$) but not the reverse.

Both of these definitions can be understood as capturing *observational dependence*, in the sense that they consist of dependence relative to an *observed* causal model M. Put differently, the actual world $(\boldsymbol{u}, \boldsymbol{v})$ belongs to the observational rung of Pearl's three-fold causal hierarchy (Bareinboim

et al., 2022). We can also consider dependence in a world that is brought about *after* M has been intervened upon, so that the causal model takes on the form $M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}$. (For example, we can imagine that the actual world takes place in an experimental set-up in which some variables are being held fixed.) Except for the fact that some of its equations are constant equations, a causal model $M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}$ is just that, a causal model. Hence we can generalize our definitions to include them.

4.1. Interventional Dependence

Definition 7 $Y = y$ depends *on* $X = x$ *in* $(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}, \boldsymbol{u})$ *if* $X \notin \boldsymbol{Z}$ *and there exists* $x' \in \mathcal{R}(X)$ *s.t.:*
[CD1.] $(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}, \boldsymbol{u}) \models \langle\rangle(X = x \wedge Y = y)$ *and* [CD2.] $(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}, \boldsymbol{u}) \models [X \leftarrow x']Y \neq y$.

Definition 8 $Y = y$ counterfactually depends *on* $X = x$ *in* $(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v})$ *if* $X \notin \boldsymbol{Z}$ *and there exists* $x' \in \mathcal{R}(X)$ *s.t.:*
[CD1.] $(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v}) \models (X = x \wedge Y = y)$ *and* [CD2.] $(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v}) \models \langle X \leftarrow x'\rangle Y \neq y$.
We say that Y depends on X *if there exists some setting* $(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v})$ *and some values such that* $Y = y$ *counterfactually depends on* $X = x$.
We say that Y directly depends on X *if* Y *depends on* X *for* $\boldsymbol{Z} = \mathcal{V} - \{X, Y\}$.

In Definitions 7 and 8 the actual world $(\boldsymbol{u}, \boldsymbol{v})$ belongs to the interventional rung of Pearl's causal hierarchy: it represents a world that is observed after an intervention took place. For all variants of dependence, the formula in CD2 belongs to the counterfactual rung, in the sense that it is a statement about worlds that are counterfactual relative to $(\boldsymbol{u}, \boldsymbol{v})$. As with our original notions of dependence, it easily follows that Definition 7 implies the existence of some $\boldsymbol{v}$ such that Definition 8 holds, and thus also that if $Y = y$ depends on $X = x$ for some $(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}, \boldsymbol{u})$ then Y depends on X.

Statements of dependence matter because they are highly informative for learning the graph $\mathcal{G}_M$ of a causal model.

Theorem 9 *Given a causal model* M *and* $X, Y \in \mathcal{V}$*, if* Y *depends on* X *then* X *is an ancestor of* Y *in* $\mathcal{G}_M$.

Proof: Proofs of all Theorems are to be found in the Appendix. ∎

The reverse direction does not hold because the ancestor relation is transitive (per definition), whereas dependence is not. (Here is a simple counterexample. $Y = abs(X)$, $X = A$ if $Z = 1$ and $X = -A$ if $Z = 0$. Then Z is an ancestor of Y, but Y does not depend on Z.) If we restrict ourselves to the direct case, the two notions *do* coincide.

Theorem 10 *Given a causal model* M *and* $X, Y \in \mathcal{V}$*,* $X \rightarrow Y \in \mathcal{G}_M$ *iff* Y *directly depends on* X.

These results help to shed light on the importance of statements of counterfactual dependence in both scientific practice and in everyday life: they offer a concise and efficient means of communicating the causal structure of a model. We explore this insight further in Sections 5 and 6, leading to my proposal for defining actual causation.

Interventional dependence opens up an interesting conceptual distinction that – as far as I am aware – has been overlooked so far, namely the distinction between $(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v}) \models \langle \boldsymbol{W} \leftarrow \boldsymbol{w}\rangle\varphi$ and $(M, \boldsymbol{u}, \boldsymbol{v'}) \models \langle \boldsymbol{W} \leftarrow \boldsymbol{w}, \boldsymbol{Z} \leftarrow \boldsymbol{z}\rangle\varphi$. (Throughout the following we assume that $\boldsymbol{Z} \cap \boldsymbol{W} = \emptyset$.)

The reason that it has been overlooked is simple: it cannot be made visible using deterministic models, for there it holds that

$$(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v}) \models \langle \boldsymbol{W} \leftarrow \boldsymbol{w}\rangle\varphi \text{ iff } (M, \boldsymbol{u}, \boldsymbol{v'}) \models \langle \boldsymbol{Z} \leftarrow \boldsymbol{z}, \boldsymbol{W} \leftarrow \boldsymbol{w}\rangle\varphi. \tag{1}$$

And yet conceptually these two types of statements are quite different. For sake of simplicity, for now let us assume that $\boldsymbol{W} = \emptyset$. Then the LHS of (1) states that if the causal model is such that it contains the equations $\mathcal{F}^{\boldsymbol{Z}\leftarrow\boldsymbol{z}}$ and if this has produced the actual world $(\boldsymbol{u}, \boldsymbol{v})$, then it might be the case that φ. Simply put: intervention first, observation second. The RHS, on the other hand, states that if the causal model is such that it contains the equations $\mathcal{F}$ and if this has produced the actual world $(\boldsymbol{u}, \boldsymbol{v'})$, then if $\boldsymbol{Z}$ were $\boldsymbol{z}$ it might have been the case that φ. Simply put: observation first, intervention second. So the former is a statement about what is the case in the *actual* world (that just happens to have been brought about after an intervention took place) and the latter is a statement about what is the case in a *counterfactual* world. The equivalence (1) breaks down in the nondeterministic case, as it should. Here is a very simple example.

Example 2 *Imagine two situations sharing identical background conditions $\boldsymbol{u}$ that involve the possible recovery Y of a patient who can undergo one of two possible treatments X, each of which is sometimes effective. Assume that one treatment ($X = 1$) is the default in such situations. In situation 1, the patient is given the deviant treatment ($X \leftarrow 0$), and we observe that the patient does not recover ($Y = 0$). In this situation, it plainly holds that the patient certainly does not recover. In situation 2, we observe that the patient receives the default treatment ($X = 1$) and does not recover, and we then consider whether they would have recovered if they had received the deviant treatment instead. In this situation, the patient might have recovered. Formally this translates into a model in which the equation for Y is $Y \in \{0, 1\}$ for both of the parent values $X = 1$ and $X = 0$, and the equation for X is $X = 1$. Then – situation 1 – $(M_{X\leftarrow 0}, \boldsymbol{u}, X = 0, Y = 0) \models Y = 0$ but also – situation 2 – $(M, \boldsymbol{u}, X = 1, Y = 0) \models \langle X \leftarrow 0\rangle Y = 1$.* ∎

Nonetheless, equivalence (1) remains also for the nondeterministic case when considering partial settings (which is the case in Definition 7). Intuitively, the explanation for this goes as follows. Recall from Section 2.1 that causal formulas (relative to the original model M) are evaluated after two operations on M have taken place, first refinement and then intervention, expressed by $(M^{(\boldsymbol{u},\boldsymbol{v})})_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}$. The refinement operation updates the equations in line with the observation $\mathcal{V} = \boldsymbol{v}$, and thus this ordering aligns with the "observation first, intervention second" motto. That the equivalence breaks down for full causal settings is because the refinement and intervention operations do not commute: $(M^{(\boldsymbol{u},\boldsymbol{v})})_{\boldsymbol{Z}\leftarrow\boldsymbol{z}} \neq (M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}})^{(\boldsymbol{u},\boldsymbol{v})}$. Moreover, as a partial causal setting $(M, \boldsymbol{u})$ does not contain an observation, we evaluate formulas by considering all possible observations, which effectively neutralizes the refinement operation and thus the issue does not arise.

Theorem 11 *Given a causal model M, for any basic causal formula φ and for any choice of $\boldsymbol{Z}, \boldsymbol{W} \subseteq \mathcal{V}$, $\boldsymbol{z} \in \mathcal{R}(\boldsymbol{Z})$, $\boldsymbol{w} \in \mathcal{R}(\boldsymbol{W})$, and $\boldsymbol{u} \in \mathcal{R}(\mathcal{U})$, we have that*

$$(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}, \boldsymbol{u}) \models \langle \boldsymbol{W} \leftarrow \boldsymbol{w}\rangle\varphi \textit{ iff } (M, \boldsymbol{u}) \models \langle \boldsymbol{Z} \leftarrow \boldsymbol{z}, \boldsymbol{W} \leftarrow \boldsymbol{w}\rangle\varphi.$$

Although overlooked within the context of DSCMs, the distinction between the two types of statement is in fact a very important one within the context of Pearl's *probabilistic* treatment of causation. The probabilistic counterpart of the "intervention-first" statement is the subject of Pearl's

well-known *do*-calculus, which offers us rules that allow us to reduce interventional probabilities to probabilities with fewer interventions. Concretely, it is concerned with statements of the form $P_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}(Y=y|\boldsymbol{X}=\boldsymbol{x})$, where the subscript $\boldsymbol{Z}\leftarrow\boldsymbol{z}$ indicates that P is the probability distribution over $\mathcal{V}$ *after* the intervention $\boldsymbol{Z}\leftarrow\boldsymbol{z}$. (As a consequence, the observed condition $\boldsymbol{X}=\boldsymbol{x}$ is an observation after the intervention, and thus $P_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{X}=\boldsymbol{x})=1$.) The probabilistic counterpart of the "observation-first" statement is the subject of Pearl's probabilities of counterfactuals, which are statements of the form $P(Y_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}=y|\boldsymbol{X}=\boldsymbol{x})$. Here the appearance of the subscript $\boldsymbol{Z}\leftarrow\boldsymbol{z}$ next to Y (instead of P) indicates that P is the distribution after *first* having observed $\boldsymbol{X}=\boldsymbol{x}$, and *then* applying the intervention $\boldsymbol{Z}\leftarrow\boldsymbol{z}$. (As a consequence, in general we do not have that $P(\boldsymbol{X}_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}=\boldsymbol{x}|\boldsymbol{X}=\boldsymbol{x})=1$.) More generally,

$$P_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}(Y=y|\boldsymbol{X}=\boldsymbol{x})\neq P(Y_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}=y|\boldsymbol{X}=\boldsymbol{x}). \tag{2}$$

So why does a distinction that is fundamental to Pearl's probabilistic framework for causation not have a counterpart in his (and Halpern's) logical framework for causation? Because the former is concerned with the setting in which there is uncertainty over the context $\boldsymbol{u}$, whereas the latter is formulated relative to a known $\boldsymbol{u}$, and for DSCMs the distinction is visible only in the first of these. It suffices to condition the above probabilities on a known context $\boldsymbol{u}$ and then invoke our earlier equivalence to make this clear (under the assumption that M is deterministic):[3]

$$P_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}(Y=y|\mathcal{U}=\boldsymbol{u},\mathcal{V}=\boldsymbol{v})=1_{\{(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}},\boldsymbol{u},\boldsymbol{v})\models Y=y\}}=1_{\{(M,\boldsymbol{u},\boldsymbol{v}')\models[\boldsymbol{Z}\leftarrow\boldsymbol{z}]Y=y\}}=P(Y_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}=y|\mathcal{U}=\boldsymbol{u},\mathcal{V}=\boldsymbol{v}').$$

Since the equivalence breaks down in the nondeterministic case, so do these equalities, and therefore the probabilistic and logical treatments of causation are more closely aligned in the nondeterministic case than they are in the standard deterministic case.

Still, even in the deterministic case interventional dependence opens up a novel subject of investigation, namely the probabilistic counterpart to interventional dependence that we arrive at when we drop our earlier assumption that $\boldsymbol{W}=\emptyset$. In that case, the LHS of (1) represents a situation in which we have "intervention first, observation second, intervention third", which is a *counterfactual* statement relative to an *actual world* that was brought about *after an intervention*. Sticking to the notation used above, the probabilistic counterpart of such statements would be $P_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}(Y_{\boldsymbol{W}\leftarrow\boldsymbol{w}}=y|\boldsymbol{X}=\boldsymbol{x})$. So although such statements are perfectly sensible within Pearl's existing framework, Pearl does not discuss them.

I can only speculate, but I suspect the reason for this oversight lies in Pearl's inconsistent use of notation throughout his book (Pearl, 2009). Within the context of the *do*-calculus (p. 67), he equates the following notation: $P_{Z\leftarrow z}(Y=y)=P(Y=y|\hat{Z}=\hat{z})=P(Y=y|do(Z=z))$. This seems to imply that $P(.|do(.))$ is his notation for the "intervention-first, observation-second" type of statement. Yet within the context of counterfactuals (p. 221 and p. 232), he equates $P(Y=y|do(Z=z))$ with $P(Y_{Z\leftarrow z}=y)$, implying instead that $P(.|do(.))$ is his notation for the "observation-first, intervention-second" type of statement. This ambiguity is harmless in two extreme cases: the case in which we do not condition on any observation, or the case in which we condition on a complete context $\boldsymbol{u}$ (in the deterministic case, that is). Outside of these extremes, however, inequality (2) shows that this ambiguity results in a contradiction.[4]

3. Here $1_{\{condition\}}$ denotes the indicator function that returns 1 if the *condition* is true and 0 otherwise.
4. Interestingly, in the last chapter Pearl replies to comments by readers. In one comment (p. 354) a reader points out that Pearl lacks the notation to distinguish between an observation that occurs before, and one that occurs after, an intervention. In his reply, Pearl claims that he does make such a distinction in the chapter on counterfactuals. Yet such a distinction is nowhere to be found, further confirming that Pearl was overlooking the nuances here discussed.

5. The Logic of Causal Discovery

We now take advantage of Theorems 9, 10, and 11, to move on to considering the process of causal discovery of an NSCM M from an idealized logical perspective. Although obviously extremely important, we here set aside entirely all statistical issues justifying the inductive leap from a finite sample to a general theory, as well as all issues regarding the impracticality of performing perfect interventions, and criteria for inferring that two observations occurred under identical conditions. Instead, we start from the idealized assumption that through rigorous experimentation and well-justified assumptions about the background conditions that make up a context $\boldsymbol{u}$, we have learned a set of "interventionist statements of possibility" that hold in the ground truth model M, i.e., we have learned a set of statements S of the form $(M_{\boldsymbol{X}\leftarrow\boldsymbol{x}}, \boldsymbol{u}) \models \langle\rangle\mathcal{V} = \boldsymbol{v}$, or equivalently (Thm. 11), $(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x}\rangle\mathcal{V} = \boldsymbol{v}$, as well as the maximal set of consequences that can be derived from such statements (through the axiomatization given in (Beckers, 2025)). Furthermore, assume that S contains at least one element for each combination of a context $\boldsymbol{u}$ and an intervention $\boldsymbol{X} \leftarrow \boldsymbol{x}$, so that we can imagine having performed each possible experiment in each possible context at least once. Per assumption, if $(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x}\rangle\varphi \in S$, then $(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x}\rangle\varphi$. In order to fully identify M, we need at least the further assumption that our sample of experiments was sufficiently large for S to contain *all* possibilities, so that if $(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x}\rangle\varphi \notin S$ then $(M, \boldsymbol{u}) \models [\boldsymbol{X} \leftarrow \boldsymbol{x}]\neg\varphi$. Let us call this the **exhaustivity assumption**.

Under the **exhaustivity assumption** such a set S gets us quite a way towards identifying $\mathcal{G}_M$: S contains all statements of dependence relative to partial settings (Def. 7), both direct and not direct, so using Theorem 10 this allows us to identify a partial graph $\mathcal{G}_S \subseteq \mathcal{G}_M$. However, completely identifying $\mathcal{G}_M$ requires knowledge of all statements of *counterfactual* dependence (Def. 8), and those are not exhausted by dependence relative to partial settings. Example 2 offers an illustration: we cannot identify the edge $X \rightarrow Y \in \mathcal{G}_M$ from statements of dependence relative to partial settings alone, for those are equally compatible with the equation for Y simply being $Y \in \{0, 1\}$, so that $X \not\rightarrow Y \in \mathcal{G}_M$. Only upon learning a counterfactual statement such as $(M, \boldsymbol{u}, X = 1, Y = 0) \models \langle X \leftarrow 0\rangle Y = 1$ can we rule out the latter. In fact, $\mathcal{G}_S \subseteq \mathcal{G}_M$ is *exactly* what we learn about the graph from S and **exhaustivity**, in the sense that we cannot rule out any acyclic supergraph of $\mathcal{G}_S$ as the complete graph $\mathcal{G}_M$. Concretely, as the following result shows, the problem of identifying the model M reduces to the problem of identifying those edges in $\mathcal{G}_M$ that go beyond $\mathcal{G}_S$.

Theorem 12 *Given a set of interventionist statements S as described above and an acyclic graph $\mathcal{G}_M \supseteq \mathcal{G}_S$ over the same signature $\mathcal{S}$, the following construction uniquely defines a causal model $M = (\mathcal{S}, \mathcal{F}, \mathcal{G}_M)$ such that for each $\boldsymbol{u} \in \mathcal{R}(\mathcal{U})$, $\boldsymbol{X} \subseteq \mathcal{V}$, $\boldsymbol{x} \in \mathcal{R}(\boldsymbol{X})$, and basic formula $\varphi \in \mathcal{L}(\mathcal{S})$:*

$$(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x}\rangle\varphi \in S \text{ iff } (M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x}\rangle\varphi.$$

For all $X \in \mathcal{V}$, $x \in \mathcal{R}(X)$, $\boldsymbol{a} \in \mathcal{R}(\boldsymbol{Pa_X} \cap \mathcal{V})$ and $\boldsymbol{b} \in \mathcal{R}(\boldsymbol{Pa_X} \cap \mathcal{U})$, we have that $x \in f_X(\boldsymbol{a}, \boldsymbol{b})$ iff $(M, \boldsymbol{u}) \models \langle \boldsymbol{A} \leftarrow \boldsymbol{a}\rangle X = x \in S$ for some $\boldsymbol{u} \in \mathcal{R}(\mathcal{U})$ such that $\boldsymbol{u_B} = \boldsymbol{b}$.

We let M_S denote the model corresponding to $\mathcal{G}_S$, and call it the **default model** *for S.*

We do get unique identifiability of M just from a set S if we assume that it is a *deterministic* SCM, since in that case dependence and counterfactual dependence coincide and thus $\mathcal{G}_S = \mathcal{G}_M$. This allows us to reconceptualize the assumption of determinism as simply a more extreme version of the **exhaustivity assumption**: it amounts to the assumption that exhaustivity is reached after

doing each experiment in each context only once, because it is assumed that each context and intervention results in a *unique* solution. As a consequence, under the assumption of determinism, just a single instance of observing two solutions for the same experiment in the same context forces us to conclude that we *must* have somehow made a mistake in our experimental setup or our observations. NSCMs allow us to avoid that conclusion. (Indeed, it was the restrictiveness of the uniqueness property that motivated me in (Beckers, 2025) to develop NSCMs in the first place.)

Still, even without assuming determinism, **exhaustivity** is a strong assumption, and learning just a single novel interventionist possibility results in its violation. If we just drop it, we are unable to even rule out the extreme model M that allows for *all* possibilities, meaning the completely disconnected model in which each variable $X \in \mathcal{V}$ takes on the equation $X \in \mathcal{R}(X)$. Nor are we able to rule out any edge $X \rightarrow Y \in \mathcal{G}_M$. Therefore it would be good to replace it with an assumption that takes **exhaustivity** as the default, but is compatible with it being violated. To do so, consider that the underlying motivation for **exhaustivity** is in fact an appeal to a certain kind of simplicity assumption, namely the assumption that the best explanation for the data is the simplest model that is able to explain it. Such appeals to the best explanation are standard within scientific methodology, as is the view that simplicity is one of the most important indicators of good explanations. (Philosophers of science call this "inference to the best explanation", but among scientists it is more often referred to as "abduction" (Douven, 2021).)

Simplicity in this sense corresponds to the set-inclusion partial order for sets of interventionist possibilities: all else being equal, A is simpler than B iff $A \subseteq B$. Let us call this *int-set simplicity*. However, the complexity of causal models is also characterized by their causal structure, which is encoded in the graph $\mathcal{G}_M$. There exist many measures of relative graph complexity, but within the context of causal models we require a measure that focuses on the relations of causal dependence. In light of Theorem 9, I suggest using the ancestor-descendant relation as a measure of simplicity: all else being equal, $\mathcal{G}_A$ is simpler than $\mathcal{G}_B$ iff $\mathcal{G}_A$ is constructed out of $\mathcal{G}_B$ by removing ancestor-descendant pairs. (I explain such constructions in Definition 13 below.) Let us call this *structural simplicity*. Since both measures of simplicity pull in different directions, the most conservative approach is to only rule out those models that are dominated along the combination of both measures, meaning that there exists an alternative model explaining S which is strictly simpler along one measure and is at least as simple along the other. Furthermore, in line with – but much weaker than – **exhaustivity**, we assume a preference for int-set simplicity, so that models which are structurally simpler but less int-set simple become evermore abnormal as edges are removed. Let us call all of this the **simplicity assumption**.

Under **simplicity** the most normal models are the structurally simplest, int-set minimal, models. All the models identified in Theorem 12 are int-set minimal, but M_S is the only one that does not contain any redundant edges, and that is why I take it to be the *default* model. There do, however, also exist less normal models consistent with **simplicity**, and those will be structurally simpler than M_S. Concretely, for any graph $\mathcal{G}_T$ constructed out of $\mathcal{G}_S$, such a corresponding model M_T has to contain a minimal set of interventionist possibilities extending S. We work our way up to identifying these models as the *structural simplifications* of the model M_S.

5.1. Structural Simplifications

Structurally simplifying a graph comes down to removing edges in order to remove some ancestor-descendant pairs. Although very similar to removing edges simpliciter, things are complicated

slightly by the fact that the removal of an edge $X \to Y$ does not necessarily result in the removal of an ancestor-descendant pair, for there might remain an indirect path from X to Y. (Consider the edge $SH \to BS$ in Figure 1 for an illustration.) As the removal of an edge from $\mathcal{G}_S$ comes at the valuable expense of increasing the int-set complexity, we only consider subgraphs in which the removal of edges was worth the price (i.e., it bought us the removal of ancestor-descendant pairs).

Definition 13 *Given graphs $\mathcal{G}_1$ and $\mathcal{G}_2$, we say that $\mathcal{G}_2$ is a* structural simplification *of $\mathcal{G}_1$ if $\mathcal{G}_2 \subseteq \mathcal{G}_1$, $Anc(\mathcal{G}_2) \subseteq Anc(\mathcal{G}_1)$, and for any $X \to Y \in (\mathcal{G}_1 - \mathcal{G}_2)$: $(X, Y) \notin Anc(\mathcal{G}_2)$.*

Int-set simplicity is assessed by considering whether one model *interventionally extends* another.

Definition 14 *Given models M_1 and M_2 over a signature $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R})$, we say that M_2 is an* interventional extension *of M_1 if for all $\boldsymbol{u} \in \mathcal{R}(\mathcal{U})$, $\boldsymbol{X} \subseteq \mathcal{V}$, $\boldsymbol{x} \in \mathcal{R}(\boldsymbol{X})$, and basic formulas $\varphi \in \mathcal{L}(\mathcal{S})$, if $(M_1, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x} \rangle \varphi$ then $(M_2, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x} \rangle \varphi$.*

Definition 15 *Given a function $f : \mathcal{R}(\boldsymbol{X}) \to \mathcal{P}(\mathcal{R}(Y))$, we define the* generalized function *$f^{\boldsymbol{Z}}$ for any $\boldsymbol{Z}$ as $f^{\boldsymbol{Z}} : \mathcal{R}(\boldsymbol{Z}) \to \mathcal{P}(\mathcal{R}(Y))$ and $f^{\boldsymbol{Z}}(\boldsymbol{z}) := \bigcup_{\{\boldsymbol{x} | \boldsymbol{x}_{\boldsymbol{Z}} = \boldsymbol{z}_{\boldsymbol{X}}\}} f(\boldsymbol{x})$.*

Since $\boldsymbol{Z}$ can be deduced from the argument $\boldsymbol{z}$, we overload notation and write $f(\boldsymbol{z})$ instead of $f^{\boldsymbol{Z}}(\boldsymbol{z})$. Thus we are effectively interpreting f as a family of multi-valued functions with co-domain $\mathcal{P}(\mathcal{R}(Y))$ and different domains. This idea comes in handy when comparing functions whose domains consist of different parents, which is what is required to construct structural simplifications.

Definition 16 *Given models M_1 and M_2 over a signature $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R})$, we say that M_2 is a* structural simplification *of M_1 if $\mathcal{G}_{M_2}$ is a structural simplification of $\mathcal{G}_{M_1}$ and for each $X \in \mathcal{V}$ we have that: $f_X(\boldsymbol{Pa}^{\mathbf{2}}_{\boldsymbol{X}}) = g_X(\boldsymbol{Pa}^{\mathbf{2}}_{\boldsymbol{X}})$, where f_X and g_X respectively denote X's function in M_2 and M_1, and $\boldsymbol{Pa}^{\mathbf{2}}_{\boldsymbol{X}}$ are X's parents in M_2.*

Informally, structurally simplifying a model by removing an edge $Z \to X$ means to re-interpret f_X so that for any given input of the remaining parents of X, we return all values of X that the original f_X could have given for each value of Z. (Alternatively, one could implement the same idea by replacing $Z \to X$ with $Z' \to X$ for some novel exogenous variable Z' that takes over the role of Z but is independent of it. That is in fact the definition of edge removal used by Janzing et al. (2013) – Def 4. – within the context of measuring causal influence.)

Structural simplifications satisfy our requirement of adding a minimal set of interventionist possibilities to the original model in order to account for the removal of ancestor-descendant pairs:

Theorem 17 *Given models M_1, M_2, M_3 over a signature $\mathcal{S}$ such that M_2 is a structural simplification of M_1 and $\mathcal{G}_{M_3} = \mathcal{G}_{M_2}$, we have that M_2 is an interventional extension of M_1, and either M_3 is not an interventional extension of M_1 or M_3 is an interventional extension of M_2.*

Hence we conclude that, for a given set of interventionist possibilities S, under **simplicity** the possible causal models are identified as all structural simplifications of the default model M_S. Note that the extreme, completely disconnected, model mentioned earlier is a structural simplification of any model, and thus we cannot rule it out entirely. Still, as mentioned the degree of structural simplification is also a degree of abnormality, making the extreme model maximally abnormal.

6. Actual Causation

This is as far as we can get with just knowledge of interventionist possibilities. But what if we were to learn just a *single* counterfactual possibility? Contrary to the interventionist case, just a single instance can offer vital novel information without requiring any further assumption. For example, learning that Y counterfactually depends on X in some setting $(M, \boldsymbol{u}, \boldsymbol{v})$ offers us a way of using Theorem 9 to establish that X is an ancestor of Y. Doing so allows us to discard all structural simplifications of M_S in which X is not an ancestor of Y, or if we like, the generally larger set of simplifications in which Y does not counterfactually depend on X.

Note that if X is not ancestor of Y in $\mathcal{G}_S$, this means discarding all of our possible models. In that case, the counterfactual statement was incompatible with the weakest hypothesis that the above discovery process resulted in, namely that the ground truth M is a structural simplification of M_S, and thus we have to re-evaluate. I propose that statements of *actual causation* between X and Y correspond to the weakest statement regarding the counterfactual dependence of Y on X in $(M, \boldsymbol{u}, \boldsymbol{v})$ such that it can falsify our weakest hypothesis, and that is the statement that Y counterfactually depends on X in *some* structural simplification of the ground truth $(M, \boldsymbol{u}, \boldsymbol{v})$.

To make this precise, we first define structural simplifications relative to a specific full causal setting $(M, \boldsymbol{u}, \boldsymbol{v})$, as opposed to an entire model M or to a specific partial setting $(M, \boldsymbol{u})$.

Definition 18 *Given models M_1 and M_2 over a signature $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R})$, we say that $(M_2, \boldsymbol{u}, \boldsymbol{v})$ is a* structural simplification *of $(M_1, \boldsymbol{u}, \boldsymbol{v})$ if $M_2^{(\boldsymbol{u},\boldsymbol{v})}$ is a structural simplification of $M_1^{(\boldsymbol{u},\boldsymbol{v})}$.*

Obviously we would like that, in addition, M_2 is a structural simplification of M_1, and this is not guaranteed to be the case. Simply consider the trivial one-variable model M_1 with equation $Y \in \{0, 1\}$. Then the model M_2 with equation $Y = 1$ is not a structural simplification of M_1, but $(M_2, Y = 1)$ is a structural simplification of $(M_1, Y = 1)$. Still, Proposition 20 shows that this technicality can be ignored when defining actual causation.

Definition 19 *$X = x$ is an* actual cause *of $Y = y$ in $(M_1, \boldsymbol{u}, \boldsymbol{v})$ if there exists a structural simplification $(M_2, \boldsymbol{u}, \boldsymbol{v})$ of $(M_1, \boldsymbol{u}, \boldsymbol{v})$ so that $Y = y$ counterfactually depends on $X = x$ in $(M_2, \boldsymbol{u}, \boldsymbol{v})$.*

Proposition 20 *If $X = x$ is an actual cause of $Y = y$ in $(M_1, \boldsymbol{u}, \boldsymbol{v})$ then there exists a structural simplification $(M_2, \boldsymbol{u}, \boldsymbol{v})$ of $(M_1, \boldsymbol{u}, \boldsymbol{v})$ so that M_2 is a structural simplification of M_1 and $Y = y$ counterfactually depends on $X = x$ in $(M_2, \boldsymbol{u}, \boldsymbol{v})$.*

Note that (by applying Theorem 17 twice, once to M_2 and M_1 and once more to $M_2^{(\boldsymbol{u},\boldsymbol{v})}$ and $M_1^{(\boldsymbol{u},\boldsymbol{v})}$) this means such structural simplifications result in minimal increases – given $\mathcal{G}_{M_2}$ – both of interventionist possibilities in M_1 and counterfactual dependence w.r.t. $(M_1, \boldsymbol{u}, \boldsymbol{v})$.

I have here offered a novel definition of actual causation based entirely on the unique function that it can fulfil in learning a causal model. Moreover, it is the first definition of actual causation that applies equally to DSCMs and NSCMs. Still, for it to be a contender as a definition of *actual causation*, as opposed to some other interesting concept, it should offer defensible verdicts of causation across the enormous variety of examples that are discussed in the literature. That this is indeed the case follows from the fact that, when restricted to DSCMs, my definition behaves almost identically to the one I first developed in (Beckers, 2021) and later reformulated in (Beckers, 2022). Concretely, the new definition agrees with my previous one on *all* of the examples I discuss in the

former, and therefore I refer the reader to that work for a comprehensive defense of the verdicts reached by my definition. Here I simply illustrate the definition using Example 1 and a variant of it that is a rare case where my new definition differs from – and improves upon – my old definition.

Applying Definition 19 to Example 1 (where $\mathcal{U} = \emptyset$) we get that $ST = 1$ is an actual cause of $BS = 1$ by considering the simplification M' that is the result of removing the edge $BH \rightarrow BS$. As mentioned, this results in the equation for BS becoming $BS = 1$ if $SH = 1$ else $BS \in \{0, 1\}$ (because in M we had $\{1\}_{BS} = 1_{SH} \vee \{1, 0\}_{BH}$ and $\{0, 1\}_{BS} = 0_{SH} \vee \{1, 0\}_{BH}$). Observe that there is indeed counterfactual dependence in M', since $(M', \boldsymbol{v}) \models \langle ST \leftarrow 0 \rangle BS = 0$. Lastly, we need to verify that $M'^{\boldsymbol{v}}$ is a structural simplification of $M^{\boldsymbol{v}}$, which reduces to verifying that $f^{\boldsymbol{v}}_{BS}(SH) = g^{\boldsymbol{v}}_{BS}(SH)$, and that is a direct consequence of the fact that $f^{\boldsymbol{v}}_{BS}(SH) = f_{BS}(SH)$, $g_{BS}(SH) = g^{\boldsymbol{v}}_{BS}(SH)$, and the first observation (where, as in Definition 16, f_{BS} and g_{BS} are the respective functions in M and M'). Now I show that $BT = 1$ is not an actual cause of $BS = 1$. Clearly $BS = 1$ does not counterfactually depend on $BT = 1$ in the original model. I leave it to the reader to verify that the only other structural simplifications of $M^{\boldsymbol{v}}$ are those in which either $BT \rightarrow BH$ or $BH \rightarrow BS$ (or both) is removed. Since BT is not an ancestor of BS in any of those models, there is no counterfactual dependence either.

Lastly, I discuss a problem for my previous definition that the new definition does not suffer from. Imagine the same setup from Example 1, except that this time Suzy does not throw. Intuitively, we would not want to consider her failure to throw ($ST = 0$) to be considered a cause of the bottle shattering ($BS = 1$), and neither of the above definitions do so. Now imagine that Suzy is not always accurate (SA) when throwing. All it takes for my old definition to change its verdict is to make explicit the obvious fact that Suzy would only hit the bottle if she were accurate, so that the equation for SH becomes $SH = ST \wedge SA$. That seems an undesirable result, and the current definition does not suffer from it. (For sake of completeness I mention that the undesirable result only shows up in case SA is modelled as an endogenous variable, which is arguably not the best choice. Instead, I would of course recommend modelling the above using the *nondeterministic* equation $SH \in \{0, 1\}$ if $ST = 1$ and $SH = 0$ if $ST = 0$.)

It is worth pointing out that my definition can be interpreted as vindicating the intuitive but relentlessly discredited idea that actual causation *is* counterfactual dependence after all, although this requires a very broad interpretation of possibility. Given that we consider all structural simplifications possible, actual causation amounts to stating that changing $X = x$ possibly might have resulted in changing $Y = y$. The latter statement is precisely how counterfactual dependence is understood in the counterfactual tradition on actual causation, for it builds on the possible world semantics to counterfactuals, both of which can be traced back to the seminal work of Lewis (1973a,b).

7. Conclusion

Taking advantage of the increased generality offered by nondeterministic causal models, I have developed a *functional* account of actual causation: statements of actual causation test the hypothesis that the set of causal models we consider plausible contains the ground truth. This functional approach is a departure from the existing literature that instead focuses on ambiguous examples and inconsistent intuitions, and owes its inspiration to Woodward (2021)'s longstanding project of offering a functional account of causal reasoning more generally. In future work I aim to extend this account to a probabilistic one by extending the idea of structural simplifications to the *probabilistic* causal models I defined in (Beckers, 2025), including Causal Bayesian Networks.

Acknowledgments

I thank the reviewers for their insightful reviews. This work was funded by ARO grant: W911NF-22-1-0061.

References

Elias Bareinboim, Juan D. Correa, Duligur Ibeling, and Thomas Icard. On pearl's hierarchy and the foundations of causal inference. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, pages 507–556. Association for Computing Machinery, 2022.

Sander Beckers. Causal sufficiency and actual causation. *Journal of Philosophical Logic*, 50:1341–1374, 2021.

Sander Beckers. Causal explanations and XAI. In *Proceedings of the First Conference on Causal Learning and Reasoning*, volume 177, pages 90–109. PMLR, 2022.

Sander Beckers. Nondeterministic causal models. In *Proceedings of the Fourth Conference on Causal Learning and Reasoning*, 2025.

Igor Douven. Abduction. *Stanford Encyclopedia of Philosophy*, 2021.

Clark Glymour, David Danks, Bruce Glymour, Frederick Eberhardt, Joseph Ramsey, Richard Scheines, Peter Spirtes, Choh Man Teng, and Jiji Zhang. Actual causation: a stone soup essay. *Synthese*, 2:169–192, 2010.

Joseph Y. Halpern. Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research*, 12: 317–337, 2000.

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

Christopher Hitchcock. Actual causation: What's the use? In Christopher Hitchcock Helen Beebee and Huw Price, editors, *Making a Difference: Essays on the Philosophy of Causation*, pages 116–131. Oxford University Press, 2017.

Dominik Janzing, David Balduzzi, Moritz Grosse-Wentrup, and Bernhard Schölkopf. Quantifying causal influences. *The Annals of Statistics*, 41(5):2324–2358, 2013.

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

David Lewis. *Counterfactuals*. Blackwell Publishers and Harvard University Press, 1973b.

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

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

James Woodward. *Causation with a Human Face: Normative Theory and Descriptive Psychology*. Oxford University Press, 2021.

Proofs of Theorems

Theorem 9 *Given a causal model M and $X, Y \in \mathcal{V}$, if Y (deterministically) depends on X then X is an ancestor of Y in $\mathcal{G}_M$.*

Proof: Assume Y depends on X in M. Recall our assumption that $X \neq Y$. We show that either $X \in \boldsymbol{Pa_Y}$, or there exists some $P \in \boldsymbol{Pa_Y}$ such that P depends on X. The result then follows by straightforward induction.

If $X \in \boldsymbol{Pa_Y}$ then per definition $X \to Y \in \mathcal{G}_M$. So let us assume $X \notin \boldsymbol{Pa_Y}$. We know that for some $(M_{\boldsymbol{Z} \leftarrow \boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v})$, $X \notin \boldsymbol{Z}$ and values x, x', y we have:

- $(M_{\boldsymbol{Z} \leftarrow \boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v}) \models (X = x \wedge Y = y)$.
- $(M_{\boldsymbol{Z} \leftarrow \boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v}) \models \langle X \leftarrow x' \rangle Y \neq y$.

Since we have that both $(M_{\boldsymbol{Z} \leftarrow \boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v}) \models \boldsymbol{Z} = \boldsymbol{z}$ and $(M_{\boldsymbol{Z} \leftarrow \boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v}) \models [X \leftarrow x'] \boldsymbol{Z} = \boldsymbol{z}$, it follows that $Y \notin \boldsymbol{Z}$. Therefore by the first condition, $y \in f(\boldsymbol{pa_Y})$, where $\boldsymbol{pa_Y} = (\boldsymbol{u}, \boldsymbol{v})_{\boldsymbol{Pa_Y}}$. Also, by the second condition, there exists some $\boldsymbol{v'}$ such that $\boldsymbol{v'}_X = x'$, $\boldsymbol{v'}_Y = y'$ for some $y \neq y'$, and $y' \in f^{(\boldsymbol{pa_Y}, y)}(\boldsymbol{pa'_Y})$, where $\boldsymbol{pa'_Y} = (\boldsymbol{u}, \boldsymbol{v'})_{\boldsymbol{Pa_Y}}$. Combining both claims, we get that $\boldsymbol{pa_Y} \neq \boldsymbol{pa'_Y}$, and thus there exists some $P \in \boldsymbol{Pa_Y}$ such that $p = \boldsymbol{pa_Y}_P \neq \boldsymbol{pa'_Y}_P = p'$. Furthermore, $P \in \mathcal{V}$, since $\mathcal{U} = \boldsymbol{u}$ across both solutions. So $p = \boldsymbol{v}_P$ and $p' = \boldsymbol{v'}_P$.

Therefore we have the following two conditions, which imply that P depends on X, as had to be shown:

- $(M_{\boldsymbol{Z} \leftarrow \boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v}) \models (X = x \wedge P = p)$.
- $(M_{\boldsymbol{Z} \leftarrow \boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v}) \models \langle X \leftarrow x' \rangle P \neq p$.

∎

Theorem 10 *Given a causal model M and $X, Y \in \mathcal{V}$, $X \to Y \in \mathcal{G}_M$ iff Y directly depends on X.*

Proof: We start with the implication from right to left. Assume Y directly depends on X. Then just as in the proof of Theorem 9 we can conclude that either $X \in \boldsymbol{Pa_Y}$ or there exists some $P \in \boldsymbol{Pa_Y}$ which changes value in the solutions used in the two conditions, and $P \in \mathcal{V}$, $P \notin \boldsymbol{Z}$. Given that we have direct dependence, also $\boldsymbol{Z} = \mathcal{V} - \{X, Y\}$, and thus $P = X$.

Now the implication from left to right. First let us assume that f_Y is a deterministic function, recalling our convention that in that case, if $X \in \boldsymbol{Pa_Y}$ then there exists some setting $\boldsymbol{w}$ of all the other parents and values x, x', y, y' such that $y = f_Y(\boldsymbol{w}, x) \neq f_Y(\boldsymbol{w}, x') = y'$. Let $\boldsymbol{A} = \boldsymbol{Pa_Y}_{\mathcal{V}}$ and $\boldsymbol{B} = \boldsymbol{Pa_Y}_{\mathcal{U}}$. Let $\boldsymbol{u} \in \mathcal{R}(\mathcal{U})$ denote some setting such that $\boldsymbol{u_B} = \boldsymbol{w_B}$, and let $\boldsymbol{z} \in \mathcal{R}(\mathcal{V} - \{X, Y\})$ denote some setting such that $\boldsymbol{z_A} = \boldsymbol{w_A}$.

Say $(\boldsymbol{u}, \boldsymbol{v})$ is a solution to the equations in $M_{\boldsymbol{Z} \leftarrow \boldsymbol{z}}$, and $\boldsymbol{v}_X = x''$. If $f_Y(\boldsymbol{w}, x'') = y = f_Y(\boldsymbol{w}, x)$, then we have the following two conditions, which is precisely what we had to show:

- $(M_{\boldsymbol{Z} \leftarrow \boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v}) \models (X = x'' \wedge Y = y)$.
- $(M_{\boldsymbol{Z} \leftarrow \boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v}) \models [X \leftarrow x'] Y \neq y$.

If $y'' = f_Y(\boldsymbol{w}, x'') \neq y = f_Y(\boldsymbol{w}, x)$, then the result follows from the following two conditions:

- $(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v}) \models (X = x'' \wedge Y = y'')$.
- $(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v}) \models [X \leftarrow x]Y \neq y''$.

Second we assume that f_Y is a multi-valued function, meaning that there exists at least one setting $(\boldsymbol{w}, x)$ such that there exist $y \neq y'$ for which $\{y, y'\} \subseteq f_Y(x, \boldsymbol{w})$. Let $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{u}$, and $\boldsymbol{z}$ be as before.

Say $(\boldsymbol{u}, \boldsymbol{v})$ is a solution to the equations in $M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}$, and $\boldsymbol{v}_X = x''$, $\boldsymbol{v}_Y = y''$, so: $(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v}) \models (X = x'' \wedge Y = y'')$.

If $x'' \neq x$, we get that $(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v}) \models \langle X \leftarrow x\rangle Y = y \wedge \langle X \leftarrow x\rangle Y = y'$. Since we cannot have both $y'' = y$ and $y'' = y'$, the result follows.

If $x'' = x$, we consider two cases.

First, consider the case such that there exists $x' \neq x$ for which $f_Y(x', \boldsymbol{w}) = f_Y(x, \boldsymbol{w})$. We have that $\{y, y'\} \subseteq f_Y(x', \boldsymbol{w})$. As before, since we cannot have both $y'' = y$ and $y'' = y'$, the result follows from $(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v}) \models \langle X \leftarrow x'\rangle Y = y \wedge \langle X \leftarrow x'\rangle Y = y'$.

Second, consider the case where there exists some $x' \neq x$ such that $f_Y(x', \boldsymbol{w}) \neq f_Y(x, \boldsymbol{w})$. If there exists some $y^* \in f_Y(x', \boldsymbol{w})$ so that $y^* \notin f_Y(x, \boldsymbol{w})$, then $(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v}) \models \langle X \leftarrow x'\rangle Y = y^*$ and the result follows. If there exists some $y^* \in f_Y(x, \boldsymbol{w})$ so that $y^* \notin f_Y(x', \boldsymbol{w})$, then there must also exist some solution $(\boldsymbol{u}, \boldsymbol{v'})$ to the equations in $M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}$ so that $\boldsymbol{v'}_X = x$ and $\boldsymbol{v}_Y = y^*$. (Note that $\boldsymbol{v}$ and $\boldsymbol{v'}$ differ only on Y.) The result then follows from observing that $(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}, \boldsymbol{u}, \boldsymbol{v'}) \models [X \leftarrow x']Y \neq y^*$. ∎

Theorem 11 *Given a causal model M, for any basic causal formula φ and for any choice of $\boldsymbol{Z}, \boldsymbol{W} \subseteq \mathcal{V}$, $\boldsymbol{z} \in \mathcal{R}(\boldsymbol{Z})$, $\boldsymbol{w} \in \mathcal{R}(\boldsymbol{W})$, and $\boldsymbol{u} \in \mathcal{R}(\mathcal{U})$, we have that*

$$(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}, \boldsymbol{u}) \models \langle \boldsymbol{W} \leftarrow \boldsymbol{w}\rangle\varphi \text{ iff } (M, \boldsymbol{u}) \models \langle \boldsymbol{Z} \leftarrow \boldsymbol{z}, \boldsymbol{W} \leftarrow \boldsymbol{w}\rangle\varphi.$$

Proof: This is a direct consequence of the following result from (Beckers, 2025).

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

$$(M, \boldsymbol{u}) \models [\boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi \text{ iff } \boldsymbol{v} \models \varphi \text{ for all states } \boldsymbol{v} \text{ such that } (\boldsymbol{u}, \boldsymbol{v}) \text{ is a solution of } M_{\boldsymbol{Y}\leftarrow\boldsymbol{y}}.$$

First note that under our assumption of $\boldsymbol{W} \cap \boldsymbol{Z} = \emptyset$, it holds that $(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}})_{\boldsymbol{W}\leftarrow\boldsymbol{w}} = M_{(\boldsymbol{Z}\leftarrow\boldsymbol{z}, \boldsymbol{W}\leftarrow\boldsymbol{w})}$.

The result then follows from applying this theorem to both sides of the equivalence in Theorem 11, giving:

For the left part, we get that $(M_{\boldsymbol{Z}\leftarrow\boldsymbol{z}}, \boldsymbol{u}) \models \langle \boldsymbol{W} \leftarrow \boldsymbol{w}\rangle\varphi$ iff there exists a state $\boldsymbol{v}$ such that $\boldsymbol{v} \models \varphi$ and $(\boldsymbol{u}, \boldsymbol{v})$ is a solution to $M_{(\boldsymbol{Z}\leftarrow\boldsymbol{z}, \boldsymbol{W}\leftarrow\boldsymbol{w})}$.

For the right part, we get that $(M, \boldsymbol{u}) \models \langle \boldsymbol{W} \leftarrow \boldsymbol{w}, \boldsymbol{Z} \leftarrow \boldsymbol{z}\rangle\varphi$ iff there exists a state $\boldsymbol{v}$ such that $\boldsymbol{v} \models \varphi$ and $(\boldsymbol{u}, \boldsymbol{v})$ is a solution to $M_{(\boldsymbol{Z}\leftarrow\boldsymbol{z}, \boldsymbol{W}\leftarrow\boldsymbol{w})}$.

∎

Theorem 12 *Given a set of interventionist statements S as described above and an acyclic graph $\mathcal{G}_M \supseteq \mathcal{G}_S$ over the same signature $\mathcal{S}$, the following construction uniquely defines a causal model $M = (\mathcal{S}, \mathcal{F}, \mathcal{G}_M)$ such that for each $\boldsymbol{u} \in \mathcal{R}(\mathcal{U})$, $\boldsymbol{X} \subseteq \mathcal{V}$, $\boldsymbol{x} \in \mathcal{R}(\boldsymbol{X})$, and basic formula $\varphi \in \mathcal{L}(\mathcal{S})$:*

$$(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x}\rangle\varphi \in S \text{ iff } (M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x}\rangle\varphi.$$

For all $X \in \mathcal{V}$, $x \in \mathcal{R}(X)$, $\boldsymbol{a} \in \mathcal{R}(\boldsymbol{Pa_X} \cap \mathcal{V})$ and $\boldsymbol{b} \in \mathcal{R}(\boldsymbol{Pa_X} \cap \mathcal{U})$, we have that $x \in f_X(\boldsymbol{a}, \boldsymbol{b})$ iff $(M, \boldsymbol{u}) \models \langle \boldsymbol{A} \leftarrow \boldsymbol{a} \rangle X = x \in S$ for some $\boldsymbol{u} \in \mathcal{R}(\mathcal{U})$ such that $\boldsymbol{u_B} = \boldsymbol{b}$.

We let M_S denote the model corresponding to $\mathcal{G}_S$, and call it the **default model** *for S.*

Proof: Assume we have some $M = (\mathcal{S}, \mathcal{F}, \mathcal{G}_M)$ and a set S as described. Concretely, for each $\boldsymbol{u} \in \mathcal{R}(\mathcal{U})$, $\boldsymbol{X} \subseteq \mathcal{V}$, $\boldsymbol{x} \in \mathcal{R}(\boldsymbol{X})$, there exists at least one $\mathcal{V} = \boldsymbol{v}$ such that $(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x} \rangle \mathcal{V} = \boldsymbol{v} \in S$. Furthermore, S contains all consequences derived from such statements using the axiomatization $AX^{\mathbf{scc}}_{non}$ of the **single context counterfactual** logic from (Beckers, 2025), which is the logic of NSCMs relative to partial settings $(M, \boldsymbol{u})$.

Since, per assumption about S, for all $X \in \mathcal{V}$, $\boldsymbol{a} \in \mathcal{R}(\boldsymbol{Pa_X} \cap \mathcal{V})$ and $\boldsymbol{b} \in \mathcal{R}(\boldsymbol{Pa_X} \cap \mathcal{U})$, there exists at least one statement of the form $(M, \boldsymbol{u}) \models \langle \boldsymbol{A} \leftarrow \boldsymbol{a} \rangle \mathcal{V} = \boldsymbol{v} \in S$ with $\boldsymbol{b} = \boldsymbol{u_B}$, we have by axiom D7 that $(M, \boldsymbol{u}) \models \langle \boldsymbol{A} \leftarrow \boldsymbol{a} \rangle X = \boldsymbol{v}_X \in S$ and thus there exists at least one value x so that $x \in f_X(\boldsymbol{a}, \boldsymbol{b})$, and thus the above construction results in a total NSCM. The rest of the proof proceeds very similarly to the completeness proof for $AX^{\mathbf{scc}}_{non}$ as it appears in (Beckers, 2025).

$\Rightarrow$:

We start with the implication from left to right. Consider some $(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x} \rangle \varphi \in S$. Given the construction of S, this means that there exists some set of statements of the form $(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x} \rangle \mathcal{V} = \boldsymbol{v} \in S$ from which the former statement can be derived using $AX^{\mathbf{scc}}_{non}$. Therefore the same holds for any NSCM for which that set of statements is true, and thus we can assume without loss of generality that φ is of the form $\mathcal{V} = \boldsymbol{v}$.

We need to show that $(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x} \rangle \mathcal{V} = \boldsymbol{v}$, which by the theorem from (Beckers, 2025) that we stated in the proof of Theorem 11 reduces to showing that $(\boldsymbol{u}, \boldsymbol{v})$ is a solution of $M_{\boldsymbol{X} \leftarrow \boldsymbol{x}}$. By axiom D4, $\boldsymbol{v_X} = \boldsymbol{x}$. This means we need to show that for each $Y \in \mathcal{V} - \boldsymbol{X}$, $y \in f_Y(\boldsymbol{pa_Y})$, where $y = \boldsymbol{v}_Y$ and $\boldsymbol{pa_Y} = (\boldsymbol{u}, \boldsymbol{v})_{\boldsymbol{pa_Y}}$. Using the notation from the theorem, let $\boldsymbol{pa_Y} = (\boldsymbol{a}, \boldsymbol{b})$.

Per construction of M, we need to show that $(M, \boldsymbol{u}) \models \langle \boldsymbol{A} \leftarrow \boldsymbol{a} \rangle Y = y \in S$.

By D3(a), we know that $(M, \boldsymbol{u}) \models \langle \boldsymbol{C} \leftarrow \boldsymbol{c}, \boldsymbol{A} \leftarrow \boldsymbol{a} \rangle \mathcal{V} = \boldsymbol{v} \in S$, where $\boldsymbol{C} = \mathcal{V} - (\{Y\} \cup \boldsymbol{A})$ and $\boldsymbol{c} = \boldsymbol{v_C}$. By D7, we get that $(M, \boldsymbol{u}) \models \langle \boldsymbol{C} \leftarrow \boldsymbol{c}, \boldsymbol{A} \leftarrow \boldsymbol{a} \rangle Y = y \in S$.

Let us consider some $\boldsymbol{c'} \in \mathcal{R}(\boldsymbol{C})$ such that $(M, \boldsymbol{u}) \models \langle \boldsymbol{A} \leftarrow \boldsymbol{a}, Y \leftarrow y \rangle \boldsymbol{C} = \boldsymbol{c'} \in S$. (Per assumption about S, such $\boldsymbol{c'}$ must exist.) If also $(M, \boldsymbol{u}) \models \langle \boldsymbol{A} \leftarrow \boldsymbol{a}, \boldsymbol{C} \leftarrow \boldsymbol{c'} \rangle Y = y \in S$, then by D5 and D7 we get that $(M, \boldsymbol{u}) \models \langle \boldsymbol{A} \leftarrow \boldsymbol{a} \rangle Y = y \in S$, as required.

Remains to consider the case where $(M, \boldsymbol{u}) \models \langle \boldsymbol{A} \leftarrow \boldsymbol{a}, \boldsymbol{C} \leftarrow \boldsymbol{c'} \rangle Y = y \notin S$. We know that for each $C \in \boldsymbol{C}$, $C \rightarrow Y \notin \mathcal{G}_S$. We show by induction that this results in a contradiction.

Consider some $C \in \boldsymbol{C}$. $C \rightarrow Y \notin \mathcal{G}_S$ implies that for all $c, c' \in \mathcal{R}(C)$, all $y \in \mathcal{R}(Y)$, $\boldsymbol{z} \in \mathcal{R}(\boldsymbol{Z})$ with $\boldsymbol{Z} = \mathcal{V} - \{C, Y\}$, if $(M, \boldsymbol{u}) \models \langle \boldsymbol{Z} \leftarrow \boldsymbol{z} \rangle (C = c \wedge Y = y) \in S$ then $(M, \boldsymbol{u}) \models \langle \boldsymbol{Z} \leftarrow \boldsymbol{z}, C \leftarrow c' \rangle Y = y \in S$. (To see why, note that due to D3(a) otherwise it must be that $c \neq c'$ and thus Y directly deterministically depends on C, and so by Theorem 10 and per construction of $\mathcal{G}_S$, $C \rightarrow Y \in \mathcal{G}_S$.) As a consequence, if $(M, \boldsymbol{u}) \models \langle \boldsymbol{Z} \leftarrow \boldsymbol{z} \rangle Y = y \in S$ then for all $c' \in \mathcal{R}(C)$, $(M, \boldsymbol{u}) \models \langle \boldsymbol{Z} \leftarrow \boldsymbol{z}, C \leftarrow c' \rangle Y = y \in S$. Taking $\boldsymbol{D} = \boldsymbol{C} - \{C\}$, $\boldsymbol{d} = \boldsymbol{c_D}$, and $c = \boldsymbol{c}_C$, from the above we know that $(M, \boldsymbol{u}) \models \langle \boldsymbol{A} \leftarrow \boldsymbol{a}, \boldsymbol{D} \leftarrow \boldsymbol{d}, C \leftarrow c \rangle Y = y \in S$. Therefore also $(M, \boldsymbol{u}) \models \langle \boldsymbol{A} \leftarrow \boldsymbol{a}, \boldsymbol{D} \leftarrow \boldsymbol{d}, C \leftarrow c' \rangle Y = y \in S$, where $c' = \boldsymbol{c'}_C$.

We can now apply the same reasoning we applied to C to some $D \in \boldsymbol{D}$, so that by induction we arrive at $(M, \boldsymbol{u}) \models \langle \boldsymbol{A} \leftarrow \boldsymbol{a}, \boldsymbol{C} \leftarrow \boldsymbol{c'} \rangle Y = y \in S$, contradicting our starting point for this case.

$\Leftarrow$:

We now move on to the implication from right to left. Consider some $(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x} \rangle \varphi$. This means there exists some $\boldsymbol{v} \models \varphi$ such that $(\boldsymbol{u}, \boldsymbol{v})$ is a solution of $M_{\boldsymbol{X} \leftarrow \boldsymbol{x}}$. So it suffices to show

that $(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x} \rangle \mathcal{V} = \boldsymbol{v} \in S$, which reduces to showing that $(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x} \rangle \boldsymbol{Y} = \boldsymbol{y} \in S$, with $\boldsymbol{Y} = \mathcal{V} - \boldsymbol{X}$ and $\boldsymbol{y} = \boldsymbol{v_Y}$.

Per construction of M, we have that for each $Y \in \boldsymbol{Y}$, $y \in f_Y(\boldsymbol{pa_Y})$, where $y = \boldsymbol{v}_Y$ and $\boldsymbol{pa_Y} = (\boldsymbol{u}, \boldsymbol{v})_{\boldsymbol{pa_Y}}$. Let $\boldsymbol{A} = \mathcal{V} \cap \boldsymbol{Pa_Y}$ and $\boldsymbol{B} = \mathcal{U} \cap \boldsymbol{Pa_Y}$. As a result, we have that $(M, \boldsymbol{u}) \models \langle \boldsymbol{A} \leftarrow \boldsymbol{v_A} \rangle Y = y \in S$.

For each $Y \in \boldsymbol{Y}$, we have that for any $\boldsymbol{Z} \subseteq \mathcal{V} - (\boldsymbol{A} \cup \{Y\})$: $Z_i \rightarrow Y \notin \mathcal{G}_S$ for all $Z_i \in \boldsymbol{Z}$. Therefore (by the same reasoning as above for $C \in \boldsymbol{C}$) for any $\boldsymbol{z}$ we have that $(M, \boldsymbol{u}) \models \langle \boldsymbol{Z} \leftarrow \boldsymbol{z}, \boldsymbol{A} \leftarrow \boldsymbol{v_A} \rangle Y = y \in S$. Letting $\boldsymbol{Y} = \{Y_1, \ldots, Y_k\}$, we have in particular that for each $i \in \{1, \ldots, k\}$: $(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x}, \boldsymbol{Y}^{-i} \leftarrow \boldsymbol{y}^{-i} \rangle Y_i = \boldsymbol{v}_{Y_i} \in S$, where $\boldsymbol{Y}^{-i} := (Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_k)$ and likewise for $\boldsymbol{y}^{-i}$.

Taking $(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x}, \boldsymbol{Y}^{-1} \leftarrow \boldsymbol{y}^{-1} \rangle Y_1 = y_1 \in S$ and $(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x}, \boldsymbol{Y}^{-2} \leftarrow \boldsymbol{y}^{-2} \rangle Y_2 = y_2 \in S$, we can apply D5 to derive that $(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x}, \boldsymbol{Y}^{-1,2} \leftarrow \boldsymbol{y}^{-1,2} \rangle (Y_1 = y_1 \wedge Y_2 = y_2) \in S$. By the same reasoning, we get that $(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x}, \boldsymbol{Y}^{-2,3} \leftarrow \boldsymbol{y}^{-2,3} \rangle (Y_2 = y_2 \wedge Y_3 = y_3) \in S$. Again applying D5 to the last two statements, we get that $(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x}, \boldsymbol{Y}^{-1,2,3} \leftarrow \boldsymbol{y}^{-1,2,3} \rangle (Y_1 = y_1 \wedge Y_2 = y_2 \wedge Y_3 = y_3) \in S$. By straightforward induction, we get that $(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x} \rangle \boldsymbol{Y} \leftarrow \boldsymbol{y} \in S$, which is what had to be shown. This concludes the proof. ∎

Theorem 17 *Given models M_1, M_2, M_3 over a signature $\mathcal{S}$ such that M_2 is a structural simplification of M_1 and $\mathcal{G}_{M_3} = \mathcal{G}_{M_2}$, we have that M_2 is an interventional extension of M_1, and either M_3 is not an interventional extension of M_1 or M_3 is an interventional extension of M_2.*

Proof: Assume we have M_1, M_2, M_3 as described. First we prove that M_2 is an interventional extension of M_1. Assume $(M_1, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x} \rangle \varphi$. Just as with the proof of Theorem 12, we can assume without loss of generality that φ is of the form $\mathcal{V} = \boldsymbol{v}$. We need to show that $(M_2, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x} \rangle \mathcal{V} = \boldsymbol{v}$. By the theorem from (Beckers, 2025) that we stated in the proof of Theorem 11 this reduces to showing that if $(\boldsymbol{u}, \boldsymbol{v})$ is a solution of $M_{1,\boldsymbol{X} \leftarrow \boldsymbol{x}}$ then $(\boldsymbol{u}, \boldsymbol{v})$ is a solution of $M_{2,\boldsymbol{X} \leftarrow \boldsymbol{x}}$.

Consider some $Y \in \mathcal{V} - \boldsymbol{X}$, $y \in \mathcal{R}(Y)$, $\boldsymbol{a} \in \mathcal{R}(\boldsymbol{Pa^1_Y})$, $\boldsymbol{b} \in \mathcal{R}(\boldsymbol{Pa^2_Y})$, such that $y = \boldsymbol{v}_Y$, $\boldsymbol{a} = (\boldsymbol{u}, \boldsymbol{v})_{\boldsymbol{A}}$, $\boldsymbol{b} = (\boldsymbol{u}, \boldsymbol{v})_{\boldsymbol{B}}$, and let g_Y/f_Y denote the function for Y in M_1/M_2 respectively. We know that $y \in g_Y(\boldsymbol{a})$ and we need to show that $y \in f_Y(\boldsymbol{b})$. Per definition of a structural simplification, $f_Y(\boldsymbol{b}) = g_Y(\boldsymbol{b})$. Per definition of a generalized function, $g_Y(\boldsymbol{b}) = \bigcup_{\{\boldsymbol{a'} | \boldsymbol{a'}_{\boldsymbol{B}} = \boldsymbol{b}\}} g_Y(\boldsymbol{a'}) \supseteq g_Y(\boldsymbol{a}) \supseteq \{y\}$. Hence $y \in f_Y(\boldsymbol{b})$. This concludes the first part of the proof.

Second we prove that either M_3 is not an interventional extension of M_1 or M_3 is an interventional extension of M_2. We do so by a reductio: assume that M_3 is an interventional extension of M_1 and M_3 is not an interventional extension of M_2.

The latter statement means that there is some $(M_2, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x} \rangle \varphi$ and $(M_3, \boldsymbol{u}) \models [\boldsymbol{X} \leftarrow \boldsymbol{x}] \neg \varphi$. This means there exists some solution $(\boldsymbol{u}, \boldsymbol{v})$ of $M_{2,\boldsymbol{X} \leftarrow \boldsymbol{x}}$ that is not a solution of $M_{3,\boldsymbol{X} \leftarrow \boldsymbol{x}}$.

So there exists some $Y \in \mathcal{V} - \boldsymbol{X}$, $y \in \mathcal{R}(Y)$, and $\boldsymbol{b} \in \mathcal{R}(\boldsymbol{Pa^2_Y})$, such that $y = \boldsymbol{v}_Y$, $\boldsymbol{b} = (\boldsymbol{u}, \boldsymbol{v})_{\boldsymbol{B}}$, $y \in f_Y(\boldsymbol{b})$ and $y \notin h_Y(\boldsymbol{b})$. Here f_Y is Y's function in M_2, and h_Y is Y's function in M_3.

Letting g_Y denote Y's function in M_1, since M_2 is a structural simplification of M_1, we also have that $f_Y(\boldsymbol{b}) = g_Y(\boldsymbol{b})$ and thus $y \in g_Y(\boldsymbol{b})$. Therefore there exists some $\boldsymbol{a} \in \mathcal{R}(\boldsymbol{Pa^1_Y})$ such that $\boldsymbol{a}_{\boldsymbol{B}} = \boldsymbol{b}$ and $y \in g_Y(\boldsymbol{a})$. This implies that for some $\boldsymbol{u'} \in \mathcal{R}(\mathcal{U})$ with $\boldsymbol{u'}_{\boldsymbol{A}} = \boldsymbol{a}_{\mathcal{U}}$ and some $\boldsymbol{z} \in \mathcal{R}(\mathcal{V} - \{Y\})$ with $\boldsymbol{z}_{\boldsymbol{A}} = \boldsymbol{a}_{\mathcal{V}}$, we have $(M_1, \boldsymbol{u'}) \models \langle \boldsymbol{Z} \leftarrow \boldsymbol{z} \rangle Y = y$.

Per our first assumption, also $(M_3, \boldsymbol{u'}) \models \langle \boldsymbol{Z} \leftarrow \boldsymbol{z} \rangle Y = y$. Given that $\boldsymbol{B} \subseteq \boldsymbol{A}$, it follows that $(\boldsymbol{u'}, \boldsymbol{z}, y)_{\boldsymbol{B}} = \boldsymbol{a}_{\boldsymbol{B}} = \boldsymbol{b}$. Since $(\boldsymbol{u'}, \boldsymbol{z}, y)$ is a solution of $M_{3,\boldsymbol{Z} \leftarrow \boldsymbol{z}}$, this implies that $y \in h_Y(\boldsymbol{b})$, contradicting our earlier statement. This concludes the proof. ∎

Proposition 20 *If $X = x$ is an actual cause of $Y = y$ in $(M_1, \boldsymbol{u}, \boldsymbol{v})$ then there exists a structural simplification $(M_2, \boldsymbol{u}, \boldsymbol{v})$ of $(M_1, \boldsymbol{u}, \boldsymbol{v})$ so that M_2 is a structural simplification of M_1 and $Y = y$ counterfactually depends on $X = x$ in $(M_2, \boldsymbol{u}, \boldsymbol{v})$.*

Proof: Assume $Y = y$ counterfactually depends on $X = x$ in $(M_2, \boldsymbol{u}, \boldsymbol{v})$, where $M_2^{(\boldsymbol{u},\boldsymbol{v})}$ is a structural simplification of $M_1^{(\boldsymbol{u},\boldsymbol{v})}$, but M_2 is not a structural simplification of M_1. We need to show that there exists some M_3 over the same signature $\mathcal{S}$ such that:

1. M_3 is a structural simplification of M_1,
2. $M_3^{(\boldsymbol{u},\boldsymbol{v})}$ is a structural simplification of $M_1^{(\boldsymbol{u},\boldsymbol{v})}$,
3. $(M_3, \boldsymbol{u}, \boldsymbol{v}) \models (X = x \wedge Y = y)$,
4. $(M_3, \boldsymbol{u}, \boldsymbol{v}) \models \langle X \leftarrow x' \rangle Y \neq y$.

Note that $\mathcal{G}_{M_1} = \mathcal{G}_{M_1^{(\boldsymbol{u},\boldsymbol{v})}}$ and $\mathcal{G}_{M_2} = \mathcal{G}_{M_2^{(\boldsymbol{u},\boldsymbol{v})}}$, so $\mathcal{G}_{M_2}$ is a structural simplification of $\mathcal{G}_{M_1}$.

Let M_3 be the unique structural simplification of M_1 such that $\mathcal{G}_{M_3} = \mathcal{G}_{M_2}$. We claim that M_3 also satisfies the remaining three requirements.

Condition 3:

Given that $(M_2, \boldsymbol{u}, \boldsymbol{v})$ is a structural simplification of $(M_1, \boldsymbol{u}, \boldsymbol{v})$, $(\boldsymbol{u}, \boldsymbol{v})$ is a solution of M_1. The latter statement is equivalent to $(M_1, \boldsymbol{u}) \models \langle\rangle \mathcal{V} = \boldsymbol{v}$. By Theorem 17, M_3 is an interventional extension of M_1, and thus $(\boldsymbol{u}, \boldsymbol{v})$ is also a solution of M_3, which implies the result.

Condition 2:

Consider some $Z \in \mathcal{V}$. We know that $h_Z(\boldsymbol{Pa_Z^2}) = g_Z(\boldsymbol{Pa_Z^2})$, where h_Z/g_Z is Z's function in M_3/M_1. We know that $f_Z^{(\boldsymbol{pa_Z^2},z)}(\boldsymbol{Pa_Z^2}) = g_Z^{(\boldsymbol{pa_Z^1},z)}(\boldsymbol{Pa_Z^2})$, where f_Z is Z's function in M_2, and $\boldsymbol{pa_Z^2} = (\boldsymbol{u}, \boldsymbol{v})_{\boldsymbol{Pa_Z^2}}$, $\boldsymbol{pa_Z^1} = (\boldsymbol{u}, \boldsymbol{v})_{\boldsymbol{Pa_Z^1}}$, $z = \boldsymbol{v}_Z$. We need to show that $h_Z^{(\boldsymbol{pa_Z^2},z)}(\boldsymbol{Pa_Z^2}) = g_Z^{(\boldsymbol{pa_Z^1},z)}(\boldsymbol{Pa_Z^2})$.

First consider some $\boldsymbol{pa_Z^2}' \neq \boldsymbol{pa_Z^2}$. Then $h_Z^{(\boldsymbol{pa_Z^2},z)}(\boldsymbol{pa_Z^2}') = h_Z(\boldsymbol{pa_Z^2}')$ and $g_Z^{(\boldsymbol{pa_Z^1},z)}(\boldsymbol{pa_Z^2}') = g_Z(\boldsymbol{pa_Z^2}')$, from which the result follows.

Second consider $\boldsymbol{pa_Z^2}$. We have that $h_Z^{(\boldsymbol{pa_Z^2},z)}(\boldsymbol{pa_Z^2}) = \{z\} = f_Z^{(\boldsymbol{pa_Z^2},z)}(\boldsymbol{pa_Z^2}) = g_Z^{(\boldsymbol{pa_Z^1},z)}(\boldsymbol{pa_Z^2})$.

Condition 4:

Since both $M_2^{(\boldsymbol{u},\boldsymbol{v})}$ and $M_3^{(\boldsymbol{u},\boldsymbol{v})}$ are structural simplifications of $M_1^{(\boldsymbol{u},\boldsymbol{v})}$, and $\mathcal{G}_{M_2^{(\boldsymbol{u},\boldsymbol{v})}} = \mathcal{G}_{M_3^{(\boldsymbol{u},\boldsymbol{v})}}$, it follows that $M_2^{(\boldsymbol{u},\boldsymbol{v})} = M_3^{(\boldsymbol{u},\boldsymbol{v})}$, implying the result.∎