Causal Models with Constraints

Sander Beckers SREKCEBREDNAS@GMAIL.COM
University of Tübingen

Joseph Y. Halpern HALPERN@CS.CORNELL.EDU
Cornell University

Christopher Hitchcock CRICKY@CALTECH.EDU
California Institute of Technology

Editors: Mihaela van der Schaar, Dominik Janzing and Cheng Zhang

Abstract

Causal models have proven extremely useful in offering formal representations of causal relationships between a set of variables. Yet in many situations, there are non-causal relationships among variables. For example, we may want variables LDL, HDL, and TOT that represent the level of low-density lipoprotein cholesterol, the level of high-density lipoprotein cholesterol, and total cholesterol level, with the relation $LDL + HDL = TOT$. This cannot be done in standard causal models, because we can intervene simultaneously on all three variables. The goal of this paper is to extend standard causal models to allow for constraints on settings of variables. Although the extension is relatively straightforward, to make it useful we have to define a new intervention operation that *disconnects* a variable from a causal equation. We give examples showing the usefulness of this extension, and provide a sound and complete axiomatization for causal models with constraints.

Keywords: Causality; Constraints; Interventions; Abstractions

1. Introduction

Causal models have proven extremely useful in offering formal representations of causal relationships between a set of variables. Yet in many situations we want to study both causal and non-causal relationships between a single set of variables; this cannot be done in a standard causal model. For example, a standard causal model cannot talk simultaneously about the level of high-density lipoprotein cholesterol (HDL), the level of low-density lipoprotein cholesterol (LDL), and the level of total cholesterol (TOT), although this seems quite natural. One can imagine a situation where we only have data regarding the level of total cholesterol, even though our causal model may say that certain health conditions depend on the amount of LDL. The problem is that standard causal models allow simultaneous interventions to all variables in the model. But we cannot intervene to simultaneously set LDL to 120 mg/dL, HDL to 70, and TOT to 180, for that is logically inconsistent! In this example, the variables have a part-whole relationship, rather than a causal relationship. Other kinds of non-causal constraints giving rise to similar problems include:

- Unit transformations; for example, having variables that describe weight in pounds and weight in kilograms.
- Mathematical relationships; for example, having variables for both Cartesian co-ordinates and polar co-ordinates.

- Microscopic/macroscopic relationships; for example, having variables for chemical compositions combined with variables indicating whether a liquid sample is water, hydrogen peroxide, or sulphuric acid; or variables representing the distribution of molecular velocities in a sample of gas, together with variables representing temperature and pressure.

Representing any of these using standard causal models would require having separate causal models for each separate description, thereby ignoring the important non-causal relationships between the variables in the distinct models.

Allowing models with non-causal constraints increases the expressive power of causal models in important ways. For one thing, we can represent *ambiguous* interventions. For example, if we change only TOT, rather than changing the levels of LDL and HDL separately, then such a change is ambiguous, because it can be realized in a number of different ways, corresponding to different (and perhaps unknown) interventions on LDL and HDL. (This terminology, as well as the cholesterol example, are taken from (Spirtes and Scheines, 2004).) Having constraints also gives us a way of effectively disallowing certain interventions, by stipulating that certain settings of the variables are disallowed, such as setting TOT below the sum of LDL and HDL.

Moreover, causal models with constraints have an important practical application. In many cases, different institutions or researchers study the same causal domain using non-causally related sets of variables. These relationships can be as trivial as the unit transformations mentioned above, but can also be far more complicated, such as the relationship between particular settings and outputs of fMRI machines produced by different companies, the translation of terminology used in the financial reporting of different countries, or more generally, the relationship between datasets that encode observations of the same kind using different conventions. We cannot combine the causal models used by such groups into one (standard) causal model, because of the relationships between the variables used in different models. On the other hand, causal models with constraints allow for the integration of the causal knowledge of the individual models into one combined model.

The goal of this paper is to show how all of this (and more) can be accomplished by extending causal models with constraints on settings of variables. Although the extension is relatively straightforward, to make it useful we have to define a new operation. Specifically, we need to be able to *disconnect* a variable from a causal equation. We provide examples that illustrate how causal models with constraints can capture many situations of interest.

We are not the first to suggest moving beyond standard causal models. In many ways, our framework can be seen as formalizing the informal suggestions of Woodward (2015). In addition, Blom, Bongers, and Mooij (2019) consider *causal constraint models*, which also allow non-causally related variables, but their emphasis lies on extending causal models with additional *causal* constraints, rather than the non-causal constraints that we consider. (Concretely, they focus exclusively on causal representations of dynamic systems, and consider the constraints that arise in equilibrium.) Our work differs from theirs in several respects (see Section 5); the approaches can be viewed as complementary.

The rest of this paper is structured as follows. The next section reviews the formalism of causal models. Section 3 introduces our new formalism for representing non-causal constraints. In Section 4, we provide a sound and complete axiomatization for causal models with constraints, in the spirit of that provided by Halpern (2000) for causal models. We conclude with some discussion in Section 5.

2. Causal Models

Before getting to the new definitions, we review the standard definition of a causal model (Halpern, 2000, 2016) (with a slight modification; see below). A *causal model* M is a pair $(\mathcal{S}, \mathcal{F})$, where $\mathcal{S}$ is a *signature*, which explicitly lists the endogenous and exogenous variables and characterizes their possible values, and $\mathcal{F}$ defines a set of *structural equations*, relating the values of the variables. Formally, 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 $\mathcal{R}$ 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*).

For some endogenous variables $X \in \mathcal{V}$, $\mathcal{F}$ associates a function denoted F_X such that F_X maps $\mathcal{R}(\mathcal{U} \cup \mathcal{V} - \{X\})$ to $\mathcal{R}(X)$ (where, if $\boldsymbol{Y}$ is a set of variables, we take $\mathcal{R}(\boldsymbol{Y})$ to be an abbreviation for $\times_{Y \in \boldsymbol{Y}} \mathcal{R}(Y)$); that is, F_X takes as input the values of the variables in $\mathcal{U} \cup \mathcal{V}$ other than X, and returns a value in the range of X. Note that we have departed from standard causal models (Halpern, 2000, 2016) by not requiring $\mathcal{F}$ to associate a function F_X with *every* variable $X \in \mathcal{V}$, only some of them. This turns out to be critical when we add constraints.

If the value F_X depends only on the variables in some subset $\boldsymbol{W} \subseteq \mathcal{U} \cup \mathcal{V} - \{X\}$, we often write $F_X(\boldsymbol{w}) = x$ or $X = \mathcal{F}_X(\boldsymbol{W})$. For example, if we have an exogenous variable U and endogenous variables $X_1, \ldots, X_5$, and X_3 is the sum of X_1 and X_2, we write $X_3 = X_2 + X_1$, omitting X_4, X_5, and U. Formally, if $\boldsymbol{Y} = \mathcal{U} \cup \mathcal{V} - \{X\} - \boldsymbol{W}$, then $F_X(\boldsymbol{w}) = x$ is an abbrevation of $F_X(\boldsymbol{w}, \boldsymbol{y}) = x$ for all $\boldsymbol{y} \in \boldsymbol{Y}$. While this shorthand is quite common, as we will see below it is particularly useful in the presence of constraints.

Much of the work on causality has focused on *recursive* or *acyclic* models, where there are no dependency cycles between variables, and the values of all endogenous variables are ultimately determined by the *context*, that is, an assignment of values to the exogenous variables. As we shall see, once we allow constraints, even in acyclic models, the values of the endogenous variables may not be determined by the context; we also need a *state*, that is, an assignment of values to the endogenous variables. A context-state pair is called an *extended state*. Our approach for dealing with constraints generalizes the way Halpern (2016) deals with cyclic models, so we allow cyclic models from the start. Given a signature $\mathcal{S}$, let $\mathcal{M}^{\mathcal{S}}$ denote all causal models of the form $(\mathcal{S}, \mathcal{F})$, where $\mathcal{F}$ can be arbitrary.

It is useful to have a language for reasoning about causality. The language that has been used in earlier papers is defined as follows: Given a signature $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R})$, a *primitive event* is a formula of the form $X = x$, for $X \in \mathcal{V}$ and $x \in \mathcal{R}(X)$. A *basic causal formula (over $\mathcal{S}$)* is one of the form $[Y_1 \leftarrow y_1, \ldots, Y_k \leftarrow y_k]\varphi$, where

- φ is a Boolean combination of primitive events,
- $Y_1, \ldots, Y_k$ are distinct variables in $\mathcal{V}$, and
- $y_i \in \mathcal{R}(Y_i)$, for $i = 1, \ldots, k$.

Such a formula is abbreviated as $[\boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$, using the vector notation. The special case where $k = 0$ is abbreviated as $[\,]\varphi$.[1] We assume for simplicity that the variables in $\mathcal{V}$ are ordered, and, no matter in what order the variables appear in an intervention, the resulting formula is syntactic sugar for the formula where the variables appear in order. For example if Y_1 is earlier in the order than

1. In standard acyclic models (where there is no disconnection and an equation for each endogenous variable), we can identify $[\,]\varphi$ and the formula φ, but in our setting, we cannot do so.

Y_2, then $[Y_2 \leftarrow y_2, Y_1 \leftarrow y_1]\varphi$ is syntactic sugar for $[Y_1 \leftarrow y_1, Y_2 \leftarrow y_2]\varphi$. (This assumption is made implicitly in (Galles and Pearl, 1998; Halpern, 2000; Halpern and Peters, 2022), the papers that we are aware of that provide axiomatizations for causal models. Without it, the axiomatizations they provide would not be complete: we would need an axiom that allows us to rearrange the order of interventions.) Intuitively, $[Y_1 \leftarrow y_1, \ldots, Y_k \leftarrow y_k]\varphi$ says that φ would hold if Y_i were set to y_i, for $i = 1, \ldots, k$. A *causal formula* is a Boolean combination of basic causal formulas. For $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R})$, let $\mathcal{L}(\mathcal{S})$ consist of all causal formulas where the variables in the formulas are taken from $\mathcal{V}$ and their possible values are determined by $\mathcal{R}$.

A causal formula ψ is true or false in a causal model, given an extended state. We write $(M, \boldsymbol{u}, \boldsymbol{v}) \models \psi$ if the causal formula ψ is true in causal model M given extended state $(\boldsymbol{u}, \boldsymbol{v})$. The $\models$ relation is defined inductively (see (Halpern and Pearl, 2005; Halpern, 2016)). $(M, \boldsymbol{u}, \boldsymbol{v}) \models X = x$ if $(\boldsymbol{u}, \boldsymbol{v})$ satisfies all the equations in $\mathcal{F}$ and $X = x$ in state $\boldsymbol{v}$. We extend $\models$ to conjunctions and negations in the standard way. Finally, $(M, \boldsymbol{u}, \boldsymbol{v}) \models [\boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$ iff $(M_{\boldsymbol{Y}\leftarrow\boldsymbol{y}}, \boldsymbol{u}, \boldsymbol{v}') \models \varphi$ for all states $\boldsymbol{v}'$ such that $(\boldsymbol{u}, \boldsymbol{v}')$ satisfies all the equations in $\mathcal{F}_{\boldsymbol{Y}\leftarrow\boldsymbol{y}}$, where $M_{\boldsymbol{Y}\leftarrow\boldsymbol{y}} = (\mathcal{S}, \mathcal{F}_{\boldsymbol{Y}\leftarrow\boldsymbol{y}})$, and $\mathcal{F}_{\boldsymbol{Y}\leftarrow\boldsymbol{y}}$ is identical to $\mathcal{F}$, except that for each Y_i in $\boldsymbol{Y}$ and corresponding y_i in $\boldsymbol{y}$, the causal equation for Y_i is replaced by $Y_i = y_i$ (or $Y_i = y_i$ is added if there was no equation for Y_i in $\mathcal{F}$). We write $(M, \boldsymbol{u}) \models \psi$ if the truth of ψ depends only on the context $\boldsymbol{u}$, which is easily seen to be the case for formulas of the form $[\boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$ and write $\boldsymbol{v} \models \psi$ if ψ is a Boolean combination of primitive events that is true in state $\boldsymbol{v}$ (note that the truth of Boolean combinations of primitive events is completely determined by the state).

Some comments:

- In a standard acyclic causal model, there is a unique $\boldsymbol{v}$ such that $(\boldsymbol{u}, \boldsymbol{v})$ satisfies the equations in $\mathcal{F}$. That is why, in the standard semantics for causal formulas in acyclic causal models, there is no mention of the state $\boldsymbol{v}$; cf. (Halpern, 2016). In cyclic causal models there may be more than one such $\boldsymbol{v}$ such that $(\boldsymbol{u}, \boldsymbol{v})$ satisfies the equations in $\mathcal{F}$, or none. Once we drop the requirement that there is an equation for each endogenous variable, there may again be more than one such $\boldsymbol{v}$, even in acyclic models.

- It is easy to check that this definition is equivalent to the standard definition of $\models$ in acyclic causal models.

- If we define $\langle \boldsymbol{X} \leftarrow \boldsymbol{x} \rangle\varphi$ as an abbreviation of $\neg[\boldsymbol{X} \leftarrow \boldsymbol{x}]\neg\varphi$, then $(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x} \rangle \mathit{true}$ iff there is some state $\boldsymbol{v}$ such that $(\boldsymbol{u}, \boldsymbol{v})$ satisfies all the causal equations in $\mathcal{F}_{\boldsymbol{X}\leftarrow\boldsymbol{x}}$. In this case we say that $\boldsymbol{v}$ is a *solution* of $(M_{[\boldsymbol{X}\leftarrow\boldsymbol{x}]}, \boldsymbol{u})$, meaning that (with the obvious abuse of notation) $(M, \boldsymbol{u}) \models \langle \boldsymbol{X} \leftarrow \boldsymbol{x} \rangle \mathcal{V} = \boldsymbol{v}$.

3. Causal Models With Constraints

We now extend causal models by allowing constraints. Some of the constraints we are interested in are defined by equations, such as $TOT = HDL + LDL$. But we also want to allow constraints such as (1) $X \leq Y$, (2) $X - Y \in \mathbf{S}$ (where $\mathbf{S}$ is a set of values), and (3) X and Y are either both positive or both negative. Thus, we take a *causal model with constraints* to be a triple $(\mathcal{S}, \mathcal{F}, \mathcal{C})$, where, as before, $\mathcal{S}$ is a signature and $\mathcal{F}$ is a collection of equations, and $\mathcal{C}$ is a set of extended states (intuitively, the extended states that satisfy the constraints). In the special case where $\mathcal{C}$ contains all possible extended states (i.e., where $\mathcal{C} = \times_{Z\in\mathcal{U}\cup\mathcal{V}}\mathcal{R}(Z)$, so $\mathcal{C}$ places no constraints) and $\mathcal{F}$ associates an equation with each variable in $\mathcal{V}$, the causal model with constraints $(\mathcal{S}, \mathcal{F}, \mathcal{C})$ is equivalent to the standard causal model $(\mathcal{S}, \mathcal{F})$. Given a signature $\mathcal{S}$, let $\mathcal{M}_c^{\mathcal{S}}$ consist of all causal models with constraints of the form $(\mathcal{S}, \mathcal{F}, \mathcal{C})$, where $\mathcal{S}$ is fixed and $\mathcal{F}$ and $\mathcal{C}$ are arbitrary.

We give semantics to formulas in $\mathcal{L}(\mathcal{S})$ just as before, except that we take $\mathcal{C}$ into account. Specifically, $(M, \boldsymbol{u}) \models [\boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$ iff $(M_{\boldsymbol{Y}\leftarrow\boldsymbol{y}}, \boldsymbol{u}, \boldsymbol{v}) \models \varphi$ for all states $\boldsymbol{v}$ such that $(\boldsymbol{u}, \boldsymbol{v}) \in \mathcal{C}$ and $(\boldsymbol{u}, \boldsymbol{v})$ satisfies all the causal equations in $\mathcal{F}_{\boldsymbol{Y}\leftarrow\boldsymbol{y}}$.

Note that, crucially, the causal equations only matter for extended states $(\boldsymbol{u}, \boldsymbol{v})$ that satisfy the constraints. This explains why we often need not write an equation F_X as depending on all other variables, and why not all endogenous variables require a causal equation. Consider a model where D is diet, W_P is weight in pounds, W_K is weight in kilograms, and the constraints $\mathcal{C}$ implement the obvious logical constraint that relates W_P and W_K (meaning that W_P and W_K fully determine each other). It does not matter whether we write F_{W_P} as a function of the values of both D and W_K or as a function only of D, since for all extended states (w_P, w_K, d) and (w_P, w'_K, d), if both (w_P, w_K, d) and (w_P, w'_K, d) are in $\mathcal{C}$, then $w_P = F_{W_P}(w_K, d)$ iff $w_P = F_{W_P}(w'_K, d)$. Moreover, it is unnecessary to write an additional causal equation for W_K; it is far more natural for W_K to be determined by the logical constraint that relates W_P and W_K.

We find it useful to extend the language $\mathcal{L}(\mathcal{S})$ a little further, to allow us to disconnect some variables $\boldsymbol{X}$ from their causal equations, so that the values of the variables in $\boldsymbol{X}$ are determined only by the constraints. Specifically, we allow formulas of the form $[disc(\boldsymbol{X}), \boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$, where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are disjoint, and either of $\boldsymbol{X}$ or $\boldsymbol{Y}$ may be empty. $(M, \boldsymbol{u}) \models [disc(\boldsymbol{X}), \boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$ iff $(M_{-\boldsymbol{X}}, \boldsymbol{u}) \models [\boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$, where $M_{-\boldsymbol{X}}$ is the model that is just like M, except that all causal equations for variables in $\boldsymbol{X}$ are removed from $\mathcal{F}$.[2] Let $\mathcal{L}^d(\mathcal{S})$ be the language that extends $\mathcal{L}(\mathcal{S})$ by allowing disconnection.

Causal models with constraints, as the name suggests, extend causal models by adding constraints on possible solutions to the structural equations. While, at some level, this is a straightforward extension, as the examples we present below show, it actually adds significant expressive power, letting us capture realistic situations that cannot be captured in standard causal models. The extension also brings out some subtle issues regarding the relationship between exogenous and endogenous variables and how the value of an endogenous variable is determined that we briefly discuss here.

- In some respects, an endogenous variable for which there is no equation behaves similarly to an exogenous variable: neither is determined by the structural equations, and they can both be restricted by the constraints. However, in other respects, they behave quite differently: the value of an exogenous variable is assumed to be simply given, as it's determined by factors that are not part of our model, whereas the value of an endogenous variable that does not have an equation is either free to take on any value that is allowed by the constraints, or is set to some value by means of an intervention.

- We could further generalize the way that the values of endogenous variables are determined. Instead of having to choose between an endogenous variable X being uniquely determined by its equation or not being determined by an equation at all, we could have an equation F_X such that F_X maps $\mathcal{R}(\mathcal{U} \cup \mathcal{V} - \{X\})$ to $\mathcal{P}(\mathcal{R}(X))$. (Peters and Halpern (2021) go even further and

2. Requiring that $\boldsymbol{X}$ and $\boldsymbol{Y}$ be disjoint does not lose expressive power. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ were not disjoint, we would want $(M, \boldsymbol{u}) \models [disc(\boldsymbol{X}), \boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$ iff $(M, \boldsymbol{u}) \models [disc(\boldsymbol{X} - \boldsymbol{Y}), \boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$.

abandon equations altogether, taking a causal model to simply be a mapping from context-intervention pairs to states.) Although we believe that this is a sensible generalization, we believe that the current framework is already sufficiently expressive to merit a discussion of its own.

Example 1 *Suppose that two different researchers study the effect of temperature on heat stroke in vulnerable populations. One expresses temperature in Celsius (and uses a variable TC to represent temperature in Celsius) while the other uses a variable TF to represent temperature in Fahrenheit. We can combine their two models into a single causal model M that includes the constraint $TF = 1.8TC + 32$ (which means that $\mathcal{C}$ consists of all those extended states where the equation holds). For simplicity, suppose that the value of TC is determined by an exogenous variable U according to the causal equation $TC = U$. There is no causal equation for TF (whose value is determined by the constraint). There is one other variable HS (the patient will suffer heatstroke), with the causal equation $HS = 1$ if $TC \geq 40$, and $HS = 0$ otherwise. Consider the context u where $U = 35$, so that $TC = 35$, $TF = 95$, and $HS = 0$. Clearly we have that $(M, \boldsymbol{u}) \models \langle TC \leftarrow 40\rangle(HS = 1)$; if we set TC to 40, there is a unique solution to the equations, and in that solution $HS = 1$. On the other hand, we do not have $(M, \boldsymbol{u}) \models \langle TF \leftarrow 104\rangle(HS = 1)$. If we set TF to 104 degrees in context u, then TC remains at 35 degrees (since the value of TC is determined by the context u, which has not changed). The resulting state is not in $\mathcal{C}$; there are no solutions to the equations in $\mathcal{C}$ where $TC = 35$ and $TF = 104$. Thus, $[TF \leftarrow 104](HS = 0)$ is vacuously true in all these solutions; that is, $(M, \boldsymbol{u}) \models [TF \leftarrow 104](HS = 0)$. On the other hand, we have $(M, \boldsymbol{u}) \models \langle disc(TC), TF \leftarrow 104\rangle(HS = 1)$. Once we disconnect the equation for TC, there is a (unique) solution to the equations where $TF = 104$; in that solution, $TC = 40$ (because of the constraint) and $HS = 1$. The key point here is that we need to disconnect TC to get the desired effect of intervening on TF.*

Now consider a formalization of the cholesterol example.

Example 2

Consider a model M that represents the impact of cholesterol on atherosclerosis in a particular patient. While it is normal for physicians to report total cholesterol level, total cholesterol includes three different kinds of cholesterol: HDL cholesterol, LDL cholesterol, and very low-density lipoproteins (VLDL cholesterol). LDL cholesterol is harmful,contributing to the buildup of plaque in arteries. By contrast, HDL cholesterol is beneficial, since it helps to clear LDL cholesterol out of the arteries. VLDL cholesterol has little direct impact on the arteries, but it contributes to levels of triglycerides, which are harmful. In practice, it is very difficult to directly measure LDL and VLDL cholesterol. Instead, VLDL cholesterol is inferred from observed triglyceride levels, and this inferred value is used together with measured values of HDL and total cholesterol to estimate the value of LDL cholesterol. For this reason, it may be useful to be able to include all of these variables together in the same causal model. The model has the following endogenous variables:

- AS *– atherosclerosis, level of plaque build-up in arteries*
- HDL *– level of HDL cholesterol*
- LDL *– level of LDL cholesterol*
- $VLDL$ *– level of VLDL cholesterol*

- *TOT – total cholesterol level*
- *TRI – level of triglycerides*
- *D – dietary factors that affect cholesterol.*

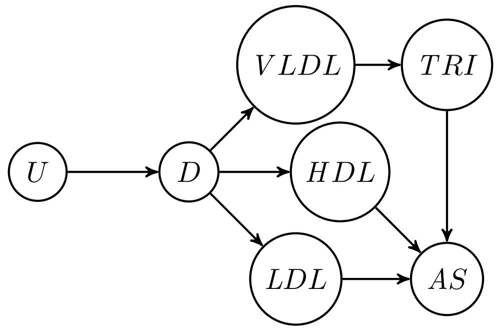

Figure 1: Causal graph for the cholesterol example (Ex. 2).

There is one exogenous variable, U. The causal equations are

- $D = F_D(U)$
- $HDL = F_{HDL}(D)$
- $LDL = F_{LDL}(D)$
- $VLDL = F_{VLDL}(D)$
- $TRI = F_{TRI}(VLDL)$
- $AS = F_{AS}(HDL, LDL, TRI)$

D is determined by the exogenous variable (i.e., the context). We do not specify the precise equations, but assume that F_{AS} is a decreasing function of HDL and increasing in LDL and TRI; we also assume that F_{TRI} is an increasing function of $VLDL$. The constraint $\mathcal{C}$ consists of all the states where $TOT = HDL + LDL + VLDL$.

In this model, we can freely intervene on HDL, LDL, and $VLDL$; the value of TOT will change in the appropriate way, so as to maintain the constraint. Of course, if we intervene to set $HDL = hdl$, $LDL = ldl$, $VLDL = vldl$ and $TOT = tot$ simultaneously, then unless the intervention is such that $tot = ldl + hdl + vldl$, there will be no states satisfying the constraints, so all formulas of the form $[LDL \leftarrow ldl, HDL \leftarrow hdl, VLDL \leftarrow vldl, TOT \leftarrow tot]\varphi$ will be vacuously true. Indeed, in a context $\boldsymbol{u}$ where $LDL = ldl^*$, $HDL = hdl^*$, $VLDL = vldl^*$, and $TOT = tot^*$, an intervention that sets TOT to $tot' > tot^*$ will also lead to an inconsistency, unless we disconnect the equation for at least one of LDL, HDL, or $VLDL$.

In the context just described, if we intervene to set $TOT = tot'$, while disconnecting the equation for LDL (but not the equations for HDL and $VLDL$), there will be a unique solution to the equations, where $HDL = hdl^*$, $VLDL = vldl^*$, $TOT = tot'$, and $LDL = tot' - hdl^* - vldl^*$. That is, intervening on TOT while disconnecting LDL results in the values of HDL and $VLDL$

remaining fixed, while LDL changes to maintain the constraint. Similarly, if we disconnect only HDL or only $VLDL$. If we disconnect all of HDL, LDL, and $VLDL$ while setting $TOT = tot'$, then there will be multiple solutions to the equations: HDL, LDL, and $VLDL$ can take arbitrary values that add up to tot'. This makes $(disc(LDL, HDL, VLDL), TOT = tot')$ what Spirtes and Scheines (2004) call an ambiguous intervention*.*

The next example shows that using the disconnection operation allows us to distinguish different ways of implementing an intervention on a variable.

Example 3 *A point is confined to the first quadrant of the Cartesian plane. We can represent its position using Cartesian coordinates X and Y, with $0 < X, Y$. We can also represent its position using polar coordinates R and θ, with $0 < R$ and $0 < \theta < \frac{\pi}{2}$. The model requires X, Y, R, and θ to satisfy the usual constraints:*

- $R = \sqrt{X^2 + Y^2}$
- $\theta = \arctan(\frac{Y}{X})$.

In the absence of intervention, the point will remain in place. Thus, we can have as our exogenous variable the previous position of the point $U = (U_X, U_Y)$. The causal equations are

- $X = U_X$
- $Y = U_Y$
- $R = \sqrt{U_X^2 + U_Y^2}$
- $\theta = \arctan(\frac{U_Y}{U_X})$.

If we want to set the value of X in a meaningful way, we need to either disconnect both R and θ or disconnect Y and R. That is, we consider interventions of the form

- $disc(R, \theta), X \leftarrow x$ *and*
- $disc(Y, R), X \leftarrow x$.

The first intervention sets the value of X while leaving Y alone; technically, this means that Y takes the value determined by the causal equations. This corresponds to sliding the point horizontally until the desired value of X is reached. This intervention removes R and θ from the influence of their causal equations, effectively forcing them to take the values determined by the constraints. The second intervention, $disc(Y, R), X \leftarrow x$, sets the value of X while leaving θ alone. This corresponds to sliding the point along the ray connecting its current position to the origin, until the desired value of X is reached. We can also consider the intervention $disc(Y, \theta), X \leftarrow x$. This corresponds to rotating the point around the origin until $X = x$. In context (u_X, u_Y), this intervention only yields solutions consistent with the constraint when $x < \sqrt{u_X^2 + u_Y^2}$.

4. A sound and complete axiomatization for causal models with constraints

In this section we provide a sound and complete axiomatization for the language $\mathcal{L}^d(\mathcal{S})$ with respect to $\mathcal{M}_c^{\mathcal{S}}$. Following (Halpern, 2000), we restrict to the case that $\mathcal{S} = (\mathcal{U}, V, \mathcal{R})$ is finite, that is, $\mathcal{U}$ is finite, $\mathcal{V}$ is finite, and $\mathcal{R}(X)$ is finite for all $X \in \mathcal{U} \cup \mathcal{V}$.

Halpern (2000) considers a somewhat richer language than we do, where the context $\boldsymbol{u}$ is part of the formula, not on the left-hand side of the $\models$. Specifically, Halpern considers primitive events of the form $X(\boldsymbol{u}) = x$, where $M \models X(\boldsymbol{u}) = x$ in Halpern's semantics iff $(M, \boldsymbol{u}) \models X = x$ in our semantics. We follow what is now the more standard usage, with the context $\boldsymbol{u}$ on the left of $\models$. We thus follow (Halpern and Peters, 2022) and consider a variant of Halpern's axioms more appropriate for our language.

Here are Halpern's axioms, as given in (Halpern and Peters, 2022) (we keep the same numbering):[3]

D0. All instances of propositional tautologies.

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

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

D3. $\langle \boldsymbol{X} \leftarrow \boldsymbol{x} \rangle (W = w \wedge \varphi) \Rightarrow \langle \boldsymbol{X} \leftarrow \boldsymbol{x}, W \leftarrow w \rangle (\varphi)$ if $W \notin \boldsymbol{X}$[4] (composition)

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

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

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

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

D9. $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle true \wedge (\langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle \varphi \Rightarrow [\boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi)$ if $\boldsymbol{Y} = \mathcal{V}$ or, for some $X \in \mathcal{V}$, $\boldsymbol{Y} = \mathcal{V} - \{X\}$ (unique outcomes for $\mathcal{V}$ and $\mathcal{V} - \{X\}$)[5]

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

We refer the reader to (Halpern and Peters, 2022) for a detailed discussion of how these axioms compare to those of Halpern (2000).

Let AX^+ consist of axiom schema D0-D5 and D7-D9, and inference rule MP.

Theorem 1 (Halpern, 2000) *AX^+ is a sound and complete axiomatization for the language $\mathcal{L}(\mathcal{S})$ with respect to $\mathcal{M}^{\mathcal{S}}$.*

We now want to extend this result to causal models with constraints. The first step is to deal with disconnection, which can be done using the following surprisingly simple axiom, where $\mathcal{R}(\boldsymbol{X}) = \times_{X \in \boldsymbol{X}} \mathcal{R}(X)$.

3. The axiom D6 that we omit is for axiomatizing acyclic models, since our focus is on general models here.
4. The requirement $W \notin \boldsymbol{X}$ is not explicit in (Halpern, 2000), but is needed to ensure that the variables in $\langle \boldsymbol{X} \leftarrow \boldsymbol{x}, W \leftarrow w \rangle$ are distinct.
5. Halpern (2000) did not include the case that $\boldsymbol{Y} = \mathcal{V}$, but it seems necessary for completeness.

DSC. $[disc(\boldsymbol{X}), \boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi \Leftrightarrow \bigwedge_{\boldsymbol{x} \in \mathcal{R}(\boldsymbol{X})} [\boldsymbol{X} \leftarrow \boldsymbol{x}, \boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$.

Roughly speaking, DSC says that disconnecting all the variables in $\boldsymbol{X}$ is the same as nondeterministically assigning the variables in $\boldsymbol{X}$ an arbitrary value in their range. As we shall see, DSC is exactly what we need to capture disconnection.

We also need to modify D9. We break the modification up into two parts, which we discuss further below.

D9′. $(\langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (X = x) \wedge \langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (X = x') \wedge \langle \boldsymbol{Y} \leftarrow \boldsymbol{y}^*, X \leftarrow x'' \rangle true) \Rightarrow \langle \boldsymbol{Y} \leftarrow \boldsymbol{y}^* \rangle (X = x'')$ if $\boldsymbol{Y} = \mathcal{V} - \{X\}$ and $x \neq x'$.

D9″. $\wedge_{x \in \mathcal{R}(X)} \langle \boldsymbol{Y} \leftarrow \boldsymbol{y}, X \leftarrow x \rangle true \Rightarrow \langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle true$, where $\boldsymbol{Y} = \mathcal{V} - \{X\}$.

D9′ is intended to deal with the case that F_X is undefined (i.e., $\mathcal{F}$ does not associate a function F_X with the variable X) in a causal model M. This must be the case if there are two distinct values $x, x' \in \mathcal{R}(X)$ such that $(\langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (X = x) \wedge \langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (X = x')$ is true in $(M, \boldsymbol{u})$ for some context $\boldsymbol{u}$. In that case, $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y}^* \rangle (X = x'')$ must be true in $(M, \boldsymbol{u})$ for all $\boldsymbol{y}^* \in \mathcal{R}(\boldsymbol{Y})$ and $x'' \in \mathcal{R}(X)$ such that $(\boldsymbol{u}, \boldsymbol{y}^*, x'') \in \mathcal{C}$, which will be the case exactly if $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y}^*, X \leftarrow x'' \rangle true$ is true in $(M, \boldsymbol{u})$. D9″ says that, for a fixed setting $\boldsymbol{y}$ of the variables in $\boldsymbol{Y} = \mathcal{V} - \{X\}$, if the constraints do not preclude X from taking any value, then there is some solution to the equations $\mathcal{F}_{\boldsymbol{Y} \leftarrow \boldsymbol{y}}$, whether or not F_X is defined.

Let $AX^{+,d}$ be the result of adding axiom DSC to AX^+ and replacing D9 by D9′ and D9″.

Theorem 2 *$AX^{+,d}$ is a sound and complete axiomatization for the language $\mathcal{L}^d(\mathcal{S})$ with respect to $\mathcal{M}_c^{\mathcal{S}}$.*

Proof We here focus on the parts of the proof that differ from that of (Halpern, 2000).

4.1. Completeness

For completeness, using DSC, we can eliminate all occurrences of $disc(\boldsymbol{X})$ from formulas, so it suffices to show that if a formula $\varphi \in \mathcal{L}(\mathcal{S})$ is valid in $\mathcal{M}_c^{\mathcal{S}}$, then it is provable from AX', where AX' is identical to AX^+ except that D9 is replaced by D9′ and D9″. The steps of the argument are standard: It suffices to show that if a formula $\varphi \in \mathcal{L}(\mathcal{S})$ is consistent with respect to AX' (i.e., we cannot prove $\neg\varphi$ in AX'), then there is a causal model with constraints $M \in \mathcal{M}_c^{\mathcal{S}}$ and a context $\boldsymbol{u}$ such that $(M, \boldsymbol{u}) \models \varphi$.

We extend $\{\varphi\}$ to a maximal set C of formulas consistent with AX'. We then use the formulas in C to define a model $M = (\mathcal{S}, \mathcal{F}, \mathcal{C}) \in \mathcal{M}_c^{\mathcal{S}}$ such that in all contexts $\boldsymbol{u}$ of M and for all formulas $\psi \in \mathcal{L}(\mathcal{S})$, we have that $(M, \boldsymbol{u}) \models \psi$ iff $\psi \in C$. Halpern (2000) used the formulas in C to define $\mathcal{F}$, by taking $F_X(\boldsymbol{u}, \boldsymbol{y}) = x$ if $\boldsymbol{y} \in \mathcal{R}(\mathcal{V} - \{X\})$ and $\langle \boldsymbol{Y} = \boldsymbol{y} \rangle (X = x) \in C$. It follows easily from D1, D2, and D9 that F_X is well defined: there is a unique value $x \in \mathcal{R}(X)$ such that $\langle \boldsymbol{Y} = \boldsymbol{y} \rangle (X = x) \in C$ for $\boldsymbol{Y} = \mathcal{V} - \{X\}$. We must work harder here, since we do not have axiom D9, only axioms D9′ and D9″. For each variable $X \in \mathcal{V}$, there may be a unique $x \in \mathcal{R}(X)$ such that $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (X = x) \in C$, but there may not be any such value x, and there may be more than one. We have to define $\mathcal{F}$ in all these cases.

We proceed as follows. We define $\mathcal{C}$ to consist of all extended states $(\boldsymbol{u}, \boldsymbol{v})$ such that $\langle \mathcal{V} \leftarrow \boldsymbol{v} \rangle true \in C$. To define $\mathcal{F}$, for each variable $X \in \mathcal{V}$ we consider three cases. Given X, if for some

$\boldsymbol{y} \in \boldsymbol{Y} = \mathcal{V} - \{X\}$ there are two values x and x' in $\mathcal{R}(X)$ such that both $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (\boldsymbol{X} = \boldsymbol{x}) \in C$ and $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (\boldsymbol{X} = \boldsymbol{x'}) \in C$, then F_X is undefined. Otherwise, for all $\boldsymbol{y} \in \mathcal{R}(\boldsymbol{Y})$, there is at most one $x \in \mathcal{R}(X)$ such that $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (X = x) \in C$. Thus, if there is some $x \in \mathcal{R}(X)$ such that $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (X = x) \in C$, then x is unique, and we take $F_X(\boldsymbol{u}, \boldsymbol{y}) = x$ for all contexts $\boldsymbol{u}$. Finally, if there are no values $x \in \mathcal{R}(X)$ such that $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (X = x) \in C$, then there must be some $x \in \mathcal{R}(X)$ such that $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y}, \boldsymbol{X} \leftarrow \boldsymbol{x} \rangle true \notin C$, for otherwise, by D9$''$, $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle true \in C$, and it follows by standard modal reasoning, using D2, D7, D8, and MP, that $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (X = x) \in C$ for some $x \in \mathcal{R}(X)$. We define $F_X(\boldsymbol{u}, \boldsymbol{y}) = x$ for all contexts $\boldsymbol{u}$. (If there is more than one value x such that $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y}, X \leftarrow x \rangle true \notin C$, we can choose one arbitrarily.) Let $M = (\mathcal{S}, \mathcal{F}, \mathcal{C})$, for this definition of $\mathcal{F}$ and $\mathcal{C}$.

Since F_X (if it is defined) is independent of $\boldsymbol{u}$, it follows that for all formulas $\psi \in \mathcal{L}(\mathcal{S})$, $(M, \boldsymbol{u}) \models \psi$ for some context $\boldsymbol{u}$ iff $(M, \boldsymbol{u}) \models \psi$ for all contexts $\boldsymbol{u}$. We show that for all $\psi \in \mathcal{L}(\mathcal{S})$, we have that $(M, \boldsymbol{u}) \models \psi$ for some (and hence all) contexts $\boldsymbol{u}$ iff $\psi \in C$. Using standard modal reasoning as in (Halpern, 2000), it suffices to consider primitive events and formulas of the form $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (\boldsymbol{X} = \boldsymbol{x})$. Using D4, we can further restrict to the case where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are disjoint. We proceed by induction on $|\mathcal{V} - \boldsymbol{Y}|$. If $|\mathcal{V} - \boldsymbol{Y}| = 0$, then $\boldsymbol{Y} = \mathcal{V}$ and we can take $\boldsymbol{X} = \boldsymbol{x}$ to be the formula $true$ and take $\boldsymbol{Y} = \boldsymbol{y}$ to be $\mathcal{V} = \boldsymbol{v}$ for some state $\boldsymbol{v}$. Note that $\langle \mathcal{V} \leftarrow \boldsymbol{v} \rangle true \in C$ iff $\boldsymbol{v} \in \mathcal{C}$ iff $(M, \boldsymbol{u}) \models \langle \mathcal{V} \leftarrow \boldsymbol{v} \rangle true$, as desired.

If $|\mathcal{V} - \boldsymbol{Y}| = 1$, then $\mathcal{V} - \boldsymbol{Y} = \{X\}$ for some variable $X \in \mathcal{V}$. Suppose that $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (X = x) \in C$. Then by D3, we must have $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y}, X \leftarrow x \rangle true \in C$. There are two cases: If for some $\boldsymbol{y}^* \in \mathcal{R}(\boldsymbol{Y})$ there exist two values x' and x'' in $\mathcal{R}(X)$ such that both $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y}^* \rangle (X = x') \in C$ and $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y}^* \rangle (X = x'') \in C$, then F_X is undefined. It easily follows that $(M, \boldsymbol{u}) \models \langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (X = x)$. Otherwise, for all $\boldsymbol{y}^* \in \mathcal{R}(\boldsymbol{Y})$, there is at most one x' such that $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y}^* \rangle (X = x') \in C$, so x has to be the unique value $x' \in \mathcal{R}(X)$ such that $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (X = x') \in C$; therefore, by construction, $F_X(\boldsymbol{u}, \boldsymbol{y}) = x$. It again follows that $(M, \boldsymbol{u}) \models \langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (X = x)$.

For the opposite direction, suppose that $(M, \boldsymbol{u}) \models \langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (X = x)$. Then $(M, \boldsymbol{u}) \models \langle \boldsymbol{Y} \leftarrow \boldsymbol{y}, X \leftarrow x \rangle true$, so $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y}, X \leftarrow x \rangle true \in C$ by the induction hypothesis, and either (1) $F_X(\boldsymbol{u}, \boldsymbol{y}) = x$ or (2) F_X is undefined. In case (1), by construction, $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (X = x) \in C$. In case (2), there must be two values x' and x'' in $\mathcal{R}(X)$ and some value $\boldsymbol{y}^* \in \mathcal{R}(\boldsymbol{Y})$ such that such that $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y}^* \rangle (X = x') \in C$ and $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y}^* \rangle (X = x'') \in C$. Since $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y}, X \leftarrow x \rangle true \in C$, by D9$'$, $\langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (X = x) \in C$, as desired.

The inductive step proceeds just as in (Halpern, 2000), using D3 and D5; we omit the details here.

4.2. Soundness

We now prove the soundness of $AX^{+,d}$. Given $M = (\mathcal{S}, \mathcal{F}, \mathcal{C}) \in \mathcal{M}_c^{\mathcal{S}}$, we want to show that all the axioms are valid in M. The argument for D0-D5, D7, and D8 is much like that given in (Galles and Pearl, 1998; Halpern, 2000); we leave the details to the reader.

For D9$'$, observe that without constraints, D9 is sound because for each $\boldsymbol{u}$, there is a unique solution $\boldsymbol{v}$ to the equations in $\mathcal{F}_{\boldsymbol{Y} \leftarrow \boldsymbol{y}}$ if $\boldsymbol{Y}$ consists of all but one endogenous variable. With constraints, there may not be a solution at all (so the first conjunct of D9 is not sound), and there may be many solutions if F_X is undefined. If $(M, \boldsymbol{u}, \boldsymbol{v}) \models (\langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (X = x) \wedge \langle \boldsymbol{Y} \leftarrow \boldsymbol{y} \rangle (X = x') \wedge \langle \boldsymbol{Y} \leftarrow \boldsymbol{y}, X \leftarrow x'' \rangle true)$, (with $x \neq x'$), then there are at least two solutions to the equations

(x and x'), so F_X must be undefined. That means that if $(\boldsymbol{u}, \boldsymbol{y}^*, x'') \in \mathcal{C}$, which must be the case if $(M, \boldsymbol{u}) \models \langle \boldsymbol{Y} \leftarrow \boldsymbol{y}^*, X \leftarrow x''\rangle true$, then $(M, \boldsymbol{u}, \boldsymbol{y}) \models \langle \boldsymbol{Y} \leftarrow \boldsymbol{y}^*\rangle(X = x'')$, as desired.

For D9$''$, suppose that $(M, \boldsymbol{u}) \models \wedge_{x \in \mathcal{R}(X)} \langle \boldsymbol{Y} \leftarrow \boldsymbol{y}, X \leftarrow x\rangle true$. We want to show that $(M, \boldsymbol{u}) \models \langle \boldsymbol{Y} \leftarrow \boldsymbol{y}\rangle true$. Suppose that F_X is defined and $F_X(\boldsymbol{u}, \boldsymbol{y}) = x$. Let $\boldsymbol{v}$ be such that $\boldsymbol{v} \models \boldsymbol{Y} = \boldsymbol{y} \wedge X = x$. Since $(M, \boldsymbol{u}) \models \langle \boldsymbol{Y} \leftarrow \boldsymbol{y}, X \leftarrow x\rangle true$, and $\boldsymbol{v}$ is the unique state such that $(\boldsymbol{u}, \boldsymbol{v})$ satisfies the equations in $\mathcal{F}_{\boldsymbol{Y} \leftarrow \boldsymbol{y}, X \leftarrow x}$, it must be the case that $(\boldsymbol{u}, \boldsymbol{v}) \in \mathcal{C}$ and satisfies the equations in $\mathcal{F}_{\boldsymbol{Y} \leftarrow \boldsymbol{y}}$. Thus, $(M, \boldsymbol{u}) \models \langle \boldsymbol{Y} \leftarrow \boldsymbol{y}\rangle true$, as desired. On the other hand, if F_X is undefined, since $(M, \boldsymbol{u}) \models \wedge_{x \in \mathcal{R}(X)} \langle \boldsymbol{Y} \leftarrow \boldsymbol{y}, X \leftarrow x\rangle true$, it must be the case that $(\boldsymbol{u}, \boldsymbol{y}, x) \in \mathcal{C}$ for all $x \in \mathcal{R}(X)$, and $(\boldsymbol{u}, \boldsymbol{y}, x)$ satisfies all the equations in $\mathcal{F}_{\boldsymbol{Y} \leftarrow \boldsymbol{y}}$, so $(M, \boldsymbol{u}) \models \bigwedge_{x \in \mathcal{R}(X)} \langle \boldsymbol{Y} \leftarrow \boldsymbol{y}\rangle(X = x)$, and hence $(M, \boldsymbol{u}) \models \langle \boldsymbol{Y} \leftarrow \boldsymbol{y}\rangle true$,

Finally, for DSC, suppose that $(M, \boldsymbol{u}) \models [disc(\boldsymbol{X}), \boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$. Then $(M_{-\boldsymbol{X}}, \boldsymbol{u}) \models [\boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$. So for all $(\boldsymbol{u}, \boldsymbol{v}) \in \mathcal{C}$ such that $(\boldsymbol{u}, \boldsymbol{v})$ satisfies the equations in $\mathcal{F}_{-\boldsymbol{X}, \boldsymbol{Y} \leftarrow \boldsymbol{y}}$, we have that $\boldsymbol{v} \models \varphi$. We claim that, for all $\boldsymbol{x} \in \mathcal{R}(\boldsymbol{X})$, we have that $(M, \boldsymbol{u}) \models [\boldsymbol{X} \leftarrow \boldsymbol{x}, \boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$. For suppose that $(\boldsymbol{u}, \boldsymbol{v}) \in \mathcal{C}$ and $(\boldsymbol{u}, \boldsymbol{v})$ satisfies the equations in $\mathcal{F}_{\boldsymbol{X} \leftarrow \boldsymbol{x}, \boldsymbol{Y} \leftarrow \boldsymbol{y}}$. Then $(\boldsymbol{u}, \boldsymbol{v})$ clearly satisfies the equations in $\mathcal{F}_{-\boldsymbol{X}, \boldsymbol{Y} \leftarrow \boldsymbol{y}}$, so $\boldsymbol{v} \models \varphi$. The result follows.

Conversely, suppose that $(M, \boldsymbol{u}) \models \bigwedge_{\boldsymbol{x} \in \mathcal{R}(\boldsymbol{X})} [\boldsymbol{X} \leftarrow \boldsymbol{x}, \boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$. We want to show that $(M, \boldsymbol{u}) \models [disc(\boldsymbol{X}), \boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$. Suppose that $(\boldsymbol{u}, \boldsymbol{v}) \in \mathcal{C}$ and $(\boldsymbol{u}, \boldsymbol{v})$ satisfies the equations in $\mathcal{F}_{-\boldsymbol{X}, \boldsymbol{Y} \leftarrow \boldsymbol{y}}$. There must be some $\boldsymbol{x} \in \mathcal{R}(\boldsymbol{X})$ such that $\boldsymbol{v} \models \boldsymbol{X} = \boldsymbol{x}$. It follows that $(\boldsymbol{u}, \boldsymbol{v})$ satisfies the equations in $\mathcal{F}_{\boldsymbol{X} \leftarrow \boldsymbol{x}, \boldsymbol{Y} \leftarrow \boldsymbol{y}}$. Since $(M, \boldsymbol{u}) \models [\boldsymbol{X} \leftarrow \boldsymbol{x}, \boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$, we must have that $\boldsymbol{v} \models \varphi$. Since this is the case for all $(\boldsymbol{u}, \boldsymbol{v}) \in \mathcal{C}$ such that $(\boldsymbol{u}, \boldsymbol{v})$ satisfies the equations in $\mathcal{F}_{-\boldsymbol{X}, \boldsymbol{Y} \leftarrow \boldsymbol{y}}$, it follows that $(M, \boldsymbol{u}) \models [disc(\boldsymbol{X}), \boldsymbol{Y} \leftarrow \boldsymbol{y}]\varphi$, as desired. ■

5. Discussion and Related Work

We have introduced an approach for allowing non-causal constraints in causal models. We believe that our approach will have applications well beyond those that we have discussed. We mention just some of them here that we hope to address in future work.

First, there has been recent work on representing causal models at different levels of abstraction (Beckers and Halpern, 2019; Rubenstein et al., 2017). Representing that a (standard) causal model M_H (intuitively, the high-level model) is an abstraction of M_L (the low-level model) is done using an *abstraction function* that relates the values of variables in M_L to those in M_H. Models with constraints can easily capture abstraction. Concretely, given a function $\tau : \mathcal{R}(\mathcal{V}_L) \to \mathcal{R}(\mathcal{V}_H)$ such that the causal model M_H is a τ-abstraction of causal model M_L, we can construct a model with constraints M that simply combines M_L and M_H (both the signatures and the equations), and let the constraints $\mathcal{C}$ consist of all extended states $(\boldsymbol{u}_L, \tau_{\mathcal{U}}(\boldsymbol{u}_L), \boldsymbol{v}_L, \tau(\boldsymbol{v}_L))$.

The work on abstraction has two features that are not directly captured by this map. First, they include a set of *allowed interventions*. Intuitively, disallowed interventions are not meaningful or cannot be performed. Disallowed interventions in a causal model with constraints can be viewed as ones that do not have a solution. However, it seems useful to have a more systematic understanding of the set of interventions that are meaningful and will give rise to solutions. Second, abstractions have been generalized to the *approximate* case, so that the solutions to the equations in both causal models may deviate slightly from the abstraction relation (Beckers et al., 2019). As such a situation seems more realistic in practice, it would be good to generalize causal models with constraints in a similar manner. One way of doing so would be to consider a metric $d_{\mathcal{V}}(\cdot, \cdot)$ on the range of endogenous variables $\mathcal{R}(\mathcal{V})$ and consider as solutions of the model all extended states $(\boldsymbol{u}, \boldsymbol{v})$ that are within α of some $(\boldsymbol{u}', \boldsymbol{v}') \in \mathcal{C}$. This could then be combined with a probability distribution over $\mathcal{R}(\mathcal{U})$, allowing the tools for approximate abstraction to be carried over to models with constraints.

Second, causal discovery algorithms are usually limited to learning a causal model using just a single dataset. There has been interesting work on generalizing causal discovery algorithms to overcome this limitation, meaning they can take advantage of various datasets using different variables, greatly improving accuracy (Tillman and Eberhardt, 2014; Huang et al., 2020). This work has not yet considered non-causal relationships between variables. A natural step to take is to modify these algorithms so that they can exploit the constraints between variables appearing in different datasets, and learn a causal model with constraints.

Third, it is worth examining the relative expressive power of our approach and that of Blom et al. (2019). As we said, they also allow non-causally related variables. They in fact allow a more general class of constraints, ones that are active only under certain interventions. However, we allow disconnection (i.e., the $disc()$ operation), which allows us to remove causal constraints. As we saw in our examples, disconnection plays a critical role; in particular, as Example 3 shows, it allows us to specify how we want to implement an intervention on a particular variable in a way that we believe is quite useful in practice. There is no analog of this in the framework of Blom et al. It would be useful to get a deeper understanding of the connection between the two approaches.

We conclude with a brief comparison of causal models with constraints to the GSEMs (generalized structural equations models) of Peters and Halpern (2021). GSEMs are more expressive than causal models with constraints (at least, if all variables have finite range); they can simply express the effect of an intervention in a given context directly, by having a function $\boldsymbol{F}$ that takes as input a context $\boldsymbol{u}$ and an intervention I, and returns a set of states (intuitively, the set of states that might result by performing intervention I in context $\boldsymbol{u}$). Thus, given a causal model with constraints M, we can define a GSEM M' that agrees with M on all formulas in $\mathcal{L}(\mathcal{S})$ (which suffices, given that we can replace all occurrences of the $disc$ operator using the DSC axiom if all variables have finite range). However, causal models with constraints allow us to describe constraints directly, which makes them more practical for many applications.

Acknowledgments

Sander Beckers was supported by the German Research Foundation (DFG) under Germany's Excellence Strategy – EXC number 2064/1 – Project number 390727645, and by the Alexander von Humboldt Foundation. Halpern was supported in part by NSF grants IIS-178108 and IIS-1703846, MURI grant W911NF-19-1-0217, and ARO grant W911NF-22-1-0061.

References

S. Beckers and J. Y. Halpern. Abstracting causal models. In *Proc. Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)*, 2019. The full version appears at arxiv.org/abs/1812.03789.

S. Beckers, F. Eberhardt, and J. Y. Halpern. Approximate causal models. In *Proc. 35th Conference on Uncertainty in Artificial Intelligence (UAI 2019)*, 2019. The full version appears at arxiv.org/abs/1906.11583.

T. Blom, S. Bongers, and J. M. Mooij. Beyond structural causal models: causal constraints models. In *Proc. 35th Conference on Uncertainty in Artificial Intelligence (UAI 2019)*, 2019.

D. Galles and J. Pearl. An axiomatic characterization of causal counterfactuals. *Foundation of Science*, 3(1):151–182, 1998.

J. Y. Halpern. Axiomatizing causal reasoning. *Journal of A.I. Research*, 12:317–337, 2000.

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

J. Y. Halpern and J. Pearl. Causes and explanations: a structural-model approach. Part I: Causes. *British Journal for Philosophy of Science*, 56(4):843–887, 2005.

J. Y. Halpern and S. Peters. Reasoning about causal models with infinitely many variables. In *Proc. Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI-20)*, 2022. To appear.

B. Huang, K. Zhang, M. Gong, and C. Glymour. Causal discovery from multiple data sets with non-identical variable sets. In *Proc. Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)*, 2020.

S. Peters and J. Y. Halpern. Causal modeling with infinitely many variables. Available on arxiv., 2021.

P. K. Rubenstein, S. Weichwald, S. Bongers, J. M. Mooij, D. Janzing, M. Grosse-Wentrup, and B. Schölkopf. Causal consistency of structural equation models. In *Proc. 33rd Conference on Uncertainty in Artificial Intelligence (UAI 2017)*, 2017.

P. Spirtes and R. Scheines. Causal inference of ambiguous manipulations. *Philosophy of Science*, 71:833–845, 2004.

R. E. Tillman and F. Eberhardt. Learning causal structure from multiple datasets with similar variable sets. *Behaviormetrika*, 41:41–64, 2014.

J. Woodward. Interventionism and causal exclusion. *Philosophy and Phenomenological Research*, 91(2), 2015.