Reasoning about Causal Models with Infinitely Many Variables

Joseph Y. Halpern, Spencer Peters
Cornell University
halpern@cs.cornell.edu, sp2473@cornell.edu

Abstract

Generalized structural equations models (GSEMs) (Peters and Halpern 2021), are, as the name suggests, a generalization of structural equations models (SEMs). They can deal with (among other things) infinitely many variables with infinite ranges, which is critical for capturing dynamical systems. We provide a sound and complete axiomatization of causal reasoning in GSEMs that is an extension of the sound and complete axiomatization provided by Halpern (2000) for SEMs. Considering GSEMs helps clarify what properties Halpern's axioms capture.

1 Introduction

Systems that evolve in continuous time are ubiquitous in all areas of science and engineering. A number of approaches have been used to model causality in such systems, ranging from dynamical systems involving differential equations to *rule-based models* (Laurent, Yang, and Fontana 2018) for capturing complex interactions in molecular biology and *hybrid automata* (Alur et al. 1992) for describing mixed discrete-continuous systems. The standard approach to modeling causality, *structural-equations models* (SEMs), introduced by Pearl (2000), cannot handle such systems, since it allows only finitely many variables, which each have finite ranges. But continuous systems typically have real-valued variables indexed by time, which ranges over the reals (e.g., the temperature at time t).

An extension of SEMs, *generalized structural-equations models (GSEMs)*, which can capture such systems was recently proposed (Peters and Halpern 2021).The goal of this paper is to provide a sound and complete axiomatization of GSEMs, in the spirit of that provided for SEMs by Halpern (2000). There are a number of features of GSEMs that make reasoning about them subtle. We briefly discuss some of them here.

Like SEMs, GSEMs are defined with respect to a signature $\mathcal{S}$ that describes the variables in the model, their possible values, and the allowed interventions. The language $\mathcal{L}(\mathcal{S})$ of causal formulas that we consider (like that of Halpern (2000)) is parameterized by $\mathcal{S}$. A GSEM is a mapping that, given an intervention (and a *context*, see Section 1 for details), produces a set of assignments to the variables given in $\mathcal{S}$, called *outcomes*; intuitively, they correspond to possible outcomes after the intervention is performed. If the signature $\mathcal{S}$ is finite (i.e., there are finitely many variables, each of which can take on only finitely many values), there can be only finitely many outcomes. This is, in particular, the case with SEMs. But in general, in a GSEM, there may be infinitely many outcomes. This complicates reasoning about them, as we shall see.

Another complication involves *allowed interventions*. In SEMs, all possible interventions are allowed; that is, we can intervene by setting any subset of the variables to any of the values in their ranges. In GSEMs, we have more expressive power: we can specify which interventions are allowed. The idea of limiting the set of interventions has already appeared in earlier work (Beckers and Halpern 2019; Rubenstein et al. 2017). Intuitively, allowed interventions are the ones that are feasible or meaningful. The set of allowed interventions is part of the signature; it also has an impact on the language. In the language $\mathcal{L}(\mathcal{S})$, we allow a formula of the form $[\vec{X} \leftarrow \vec{x}]\varphi$ (which can be read "after intervening by setting the variables in $\vec{X}$ to $\vec{x}$, φ holds") only if $\vec{X} \leftarrow \vec{x}$ is an allowed intervention: if an intervention is not allowed, we cannot talk about it in the language. As shown by the shell game example in (Peters and Halpern 2021), restricting to allowed interventions is useful even when the signature is finite; we can describe interesting situations that are inconsistent with all interventions being allowed.

Besides creating the possibility for infinitely many outcomes, the infinitary signatures required for continuous-time systems pose certain technical problems. If, for example, we have variables ranging over the reals, and we can refer to all possible real numbers in the language, then the language must be uncountable. Although we believe that all our results continue to hold for uncountable languages, having uncountably many formulas makes soundness and completeness arguments much more complicated. We thus restrict the language so that it can refer explicitly to only countably many values and countably many interventions. This still leaves us with an extremely rich language, which easily suffices to characterize systems that occur in practice.

It is shown in (Peters and Halpern 2021, Theorem 2.1) that $\mathcal{L}(\mathcal{S})$ is rich enough to completely characterize SEMs, as well as GSEMs over infinitary signatures where there are only finitely many outcomes to each intervention (so, in particular, GSEMs with finite signatures); specifically, it is shown that if each of $\mathcal{M}$ and $\mathcal{M}'$ is either a SEM or a GSEM for which there are only finitely many outcomes to each intervention, then $\mathcal{M}$ and $\mathcal{M}'$ are $\mathcal{L}(\mathcal{S})$-*equivalent*, that is, they agree on all formulas in $\mathcal{L}(\mathcal{S})$, iff they are *equivalent*, that is, iff they have the same outcomes under all allowed interventions. (We remark that this is no longer the case if we consider GSEMs for which there may be infinitely many outcomes for a given context and intervention, which can certainly be the case in dynamical systems; see Example 3.3.)

Halpern (2000) provided axiom systems $AX^+(\mathcal{S})$ and $AX^+_{rec}(\mathcal{S})$ that he showed were sound and complete for general SEMs and acyclic SEMs, respectively. In this paper, we extend $AX^+(\mathcal{S})$ to arbitrary GSEMs, and several interesting subclasses of GSEMs (such as GSEMs with unique outcomes). First, we show that $AX^+(\mathcal{S})$ is sound and complete for the class of GSEMs satisfying $AX^+(\mathcal{S})$, if $\mathcal{S}$ is finite and $\mathcal{S}$ is *universal*; that is, if all interventions are allowed. This is an easy corollary of (Peters and Halpern 2021, Theorem 3.4), which states that if $\mathcal{S}$ is finite and universal, then every SEM with signature $\mathcal{S}$ is equivalent to a GSEM satisfying $AX^+(\mathcal{S})$, and vice versa. The assumption that $\mathcal{S}$ is universal is critical here. Example 3.3 (from (Peters and Halpern 2021)) gives a GSEM over a finite signature $\mathcal{S}$ that satisfies all the axioms of $AX^+(\mathcal{S})$ but is not equivalent to any SEM. This implies that $AX^+(\mathcal{S})$ is no longer complete for SEMs when $\mathcal{S}$ is not universal (Theorem 4.6).

We then show that a subsystem of $AX^+(\mathcal{S})$ that we call $AX^+_{basic}(\mathcal{S})$ is sound and complete for arbitrary GSEMs over a finite signature $\mathcal{S}$. We also show that, as in SEMs, extending $AX^+(\mathcal{S})$ with one more axiom gives a sound and complete system for *acyclic* GSEMs. Proving these results for arbitrary (possibly infinite) signatures $\mathcal{S}$ is nontrivial, because one of the axioms of $AX^+_{basic}(\mathcal{S})$ is no longer in the language $\mathcal{L}(\mathcal{S})$ when $\mathcal{S}$ is and the axiom corresponding to acyclicity must be strengthened. We show that this axiom can be replaced with a new inference rule that gives an equivalent system when $\mathcal{S}$ is finite. Moreover, the resulting axiom system, $AX^*_{basic}(\mathcal{S})$, is sound and complete for arbitrary GSEMs (Theorem 5.2). We further show that several properties of SEMs (such as having unique outcomes for all interventions in all contexts) can be enforced in GSEMs by adding axioms from $AX^+(\mathcal{S})$ to $AX^*_{basic}(\mathcal{S})$ (Theorem 5.2). Doing so helps clarify what properties the added axioms capture.

SEMs: a Review

Formally, a *structural-equations 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 *modifiable structural equations*, relating the values of variables. We extend the signature to include a set of *allowed interventions*, as was done in earlier work (Beckers and Halpern 2019; Rubenstein et al. 2017). Intuitively, allowed interventions are the ones that are feasible or meaningful. A signature $\mathcal{S}$ is a tuple $(\mathcal{U}, \mathcal{V}, \mathcal{R}, \mathcal{I})$, 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, finite set $\mathcal{R}(Y)$ of possible values for Y (i.e., the set of values over which Y *ranges*). We assume (as is typical for SEMs) that $\mathcal{U}$ and $\mathcal{V}$ are finite sets, and adopt the convention that for $\vec{Y} \subseteq \mathcal{U} \cup \mathcal{V}$, $\mathcal{R}(\vec{Y})$ denotes the product of the ranges of the variables appearing in $\vec{Y}$; that is, $\mathcal{R}(\vec{Y}) := \times_{Y \in \vec{Y}} \mathcal{R}(Y)$. Finally, an intervention $I \in \mathcal{I}$ is a set of pairs (X, x), where $X \in \mathcal{V}$ and $x \in \mathcal{R}(X)$. We abbreviate an intervention I by $\vec{X} \leftarrow \vec{x}$, where $\vec{X} \subseteq \mathcal{V}$. We allow $\vec{X}$ to be empty (which amounts to not intervening at all).

$\mathcal{F}$ associates with each endogenous variable $X \in \mathcal{V}$ a function denoted F_X such that $F_X : \mathcal{R}(\mathcal{U} \cup \mathcal{V} - \{X\}) \to \mathcal{R}(X)$. This mathematical notation just makes precise the fact that F_X determines the value of X, given the values of all the other variables in $\mathcal{U} \cup \mathcal{V}$. If there is one exogenous variable U and three endogenous variables, X, Y, and Z, then F_X defines the values of X in terms of the values of Y, Z, and U. For example, we might have $F_X(u, y, z) = u + y$, which is usually written as $X = U + Y$. Thus, if $Y = 3$ and $U = 2$, then $X = 5$, regardless of how Z is set.

The structural equations define what happens in the presence of external interventions. Setting the value of some variable X to x in a SEM $M = (\mathcal{S}, \mathcal{F})$ results in a new SEM, denoted $M_{X \leftarrow x}$, which is identical to M, except that the equation for X in $\mathcal{F}$ is replaced by $X = x$. Interventions on subsets $\vec{X}$ of $\mathcal{V}$ are defined similarly. Notice that $M_{\vec{X} \leftarrow \vec{x}}$ is always well defined, even if $(\vec{X} \leftarrow \vec{x}) \notin \mathcal{I}$. In earlier work, the reason that the model included allowed interventions was that, for example, relationships between two models were required to hold only for allowed interventions (i.e., the interventions that were meaningful). Here, the set of allowed interventions plays a different role, influencing the language (what we are allowed to talk about).

Given a *context* $\mathbf{u} \in \mathcal{R}(\mathcal{U})$, the *outcomes* of a SEM M under intervention $\vec{X} \leftarrow \vec{x}$ are all assignments of values $\mathbf{v} \in \mathcal{R}(\mathcal{V})$ such that the assignments $\mathbf{u}$ and $\mathbf{v}$ together satisfy the structural equations of $M_{\vec{X} \leftarrow \vec{x}}$. This set of outcomes is denoted $M(\mathbf{u}, \vec{X} \leftarrow \vec{x})$. Given an outcome $\mathbf{v}$, denote by $\mathbf{v}[X]$ and $\mathbf{v}[\vec{X}]$ the value (resp., tuple of values) that $\mathbf{v}$ assigns to X and the variables in $\vec{X}$, respectively.

An *acyclic* SEM is one for which, for every context $\mathbf{u} \in \mathcal{R}(U)$, there is some total ordering $\prec_{\mathbf{u}}$ of the endogenous variables (the ones in $\mathcal{V}$) such that if $X \prec_{\mathbf{u}} Y$, then X is independent of Y, that is, $F_X(\mathbf{u}, \ldots, y, \ldots) = F_X(\mathbf{u}, \ldots, y', \ldots)$ for all $y, y' \in \mathcal{R}(Y)$.

2 Axiomatizing SEMs

In order to talk about SEMs and the information they represent more precisely, we use the formal language $\mathcal{L}(\mathcal{S})$ for SEMs having signature $\mathcal{S}$, introduced by Halpern (2000).

We restrict the language used by Halpern (2000) to formulas containing only allowed interventions. Fix a signature $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R}, \mathcal{I})$. A *primitive event* (over signature $\mathcal{S}$) has the form $X = x$, where $X \in \mathcal{V}$ and $x \in \mathcal{R}(\mathcal{V})$. An *event* is a Boolean combination of primitive events. An *atomic formula* (over $\mathcal{S}$) has the form $[\vec{Y} \leftarrow \vec{y}]\varphi$, where $\vec{Y} \leftarrow \vec{y} \in \mathcal{I}$ (i.e., it is an allowed intervention), and φ is an event. A *causal formula* (over $\mathcal{S}$) is a Boolean combination of atomic formulas. The language $\mathcal{L}(\mathcal{S})$ consists of all causal formulas over $\mathcal{S}$. There are a number of minor differences between the language considered here and that considered by Halpern (2000). First, since Halpern implicitly assumed that all interventions were allowed, he did not have the restriction to allowed interventions. Second, Halpern considered a slightly richer language, where the context $\mathbf{u}$ was part of the formula, not on the left-hand side of the $\models$ (see below). Specifically, a primitive event had the form $X(\mathbf{u}) = x$. It has become standard not to include the context $\mathbf{u}$ in the formula (see, e.g., (Halpern and Pearl 2005; Halpern 2016)).

Next we define the semantics of $\mathcal{L}(\mathcal{S})$. An assignment $\mathbf{v} \in \mathcal{R}(\mathcal{V})$ *satisfies* the primitive event $X = x$, written $\mathbf{v} \models (X = x)$ if $\mathbf{v}[X] = x$. We extend this definition to Boolean combinations of primitive events by structural induction in the obvious way, that is, say that $\mathbf{v} \models e_1 \wedge e_2$ iff $\mathbf{v} \models e_1$ and $\mathbf{v} \models e_2$, and similarly for the other Boolean connectives $\vee$ and $\neg$. Fix a SEM M with signature $\mathcal{S}$. Given a context $\mathbf{u} \in \mathcal{R}(U)$, we say that M satisfies the atomic formula $[\vec{Y} \leftarrow \vec{y}]\varphi$ in context $\mathbf{u}$, written $(M, \mathbf{u}) \models [\vec{Y} \leftarrow \vec{y}]\varphi$, if all outcomes $\mathbf{v} \in M(\mathbf{u}, \vec{Y} \leftarrow \vec{y})$ satisfy φ. Finally, we extend this definition to causal formulas by structural induction as above. That is, $M, \mathbf{u} \models [\vec{Y} \leftarrow \vec{y}]\varphi \wedge [\vec{Z} \leftarrow \vec{z}]\psi$ iff $(M, \mathbf{u}) \models [\vec{Y} \leftarrow \vec{y}]\varphi$ and $(M, \mathbf{u}) \models [\vec{Z} \leftarrow \vec{z}]\psi$, and similarly for $\vee$ and $\neg$. As usual, $\langle \vec{Y} \leftarrow \vec{y} \rangle \varphi$ is taken to be an abbreviation for $\neg[\vec{Y} \leftarrow \vec{y}](\neg\varphi)$. It is easy to check that $(M, \mathbf{u}) \models \langle \vec{Y} \leftarrow \vec{y} \rangle \varphi$ iff φ is true of at least one outcome $\mathbf{v} \in M(\mathbf{u}, \vec{Y} \leftarrow \vec{y})$. The causal formula ψ is *valid in* M, written $M \models \psi$, if $M, \mathbf{u} \models \psi$ for all $\mathbf{u} \in \mathcal{R}(\mathcal{U})$; ψ is *satisfied in* M if $(M, \mathbf{u}) \models \psi$ for some context $\mathbf{u} \in \mathcal{R}(\mathcal{U})$.

We now review Halpern's axiomatization of SEMs where all interventions are allowed (which is based on that of Galles and Pearl (1998)). Below, for disjoint sets of variables $\vec{X}, \vec{Y}$ and interventions $\vec{X} \leftarrow \vec{x}, \vec{Y} \leftarrow \vec{y}$, the notation $\vec{X} \leftarrow \vec{x}, \vec{Y} \leftarrow \vec{y}$ stands for the combined intervention $(\vec{X} \leftarrow \vec{x}) \cup (\vec{Y} \leftarrow \vec{y})$. Notice that the order in which the interventions are combined does not matter; that is, $\vec{X} \leftarrow \vec{x}, \vec{Y} \leftarrow \vec{y} = \vec{Y} \leftarrow \vec{y}, \vec{X} \leftarrow \vec{x}$. To axiomatize acyclic SEMs, following Halpern, we define $Y \rightsquigarrow Z$, read "Y affects Z", as an abbreviation for the formula

$$\vee_{\vec{X} \subseteq \mathcal{V}, Y \notin \vec{X}, \vec{x} \in \mathcal{R}(\vec{X}), y \in \mathcal{R}(y), z \neq z' \in \mathcal{R}(Z)} ([\vec{X} \leftarrow \vec{x}](Z = z) \wedge [\vec{X} \leftarrow \vec{x}, Y \leftarrow y](Z = z'));$$

that is, Y affects Z if there is some setting of some endogenous variables $\vec{X}$ for which changing the value of Y changes the value of Z. This definition is used in axiom D6 below, which characterizes acyclicity.

Consider the following axioms:

D0. All instances of propositional tautologies.

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

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

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

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

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

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

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

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

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

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

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

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

Let AX^+ consist of axiom schema D0-D5 and D7-D9, and inference rule MP; let AX^+_{rec} be the result of adding D6 and D10 to AX^+, and removing D5 and D9.

These are not quite the same axioms that Halpern (2000) used, although they are equivalent for SEMs. In more detail:

- Instead of an arbitrary formula φ in D3, Halpern had just formulas of the form $\vec{Y} = \vec{y}$. But since in the case of SEMs, every propositional formula φ is equivalent to a disjunction of formulas of the form $\vec{Y} = \vec{y}$, and $\langle \vec{X} \leftarrow \vec{x} \rangle (\varphi \vee \psi) \Rightarrow \langle \vec{X} \leftarrow \vec{x} \rangle \varphi \vee \langle \vec{X} \leftarrow \vec{x} \rangle \psi$ is provable from the axioms (see the full paper), our version of D3 is easily seen to be equivalent to the original version for SEMs, but is stronger in the case of GSEMs.
- D5 follows from D2, D3, D6, D7, D8, D10, and MP, so it is not needed for AX^+_{rec}. (This was already essentially observed by Galles and Pearl (1998).) Indeed, as we show in the full paper, in the presence of these other axioms, D5 holds even without the requirement that $\vec{Z} = \mathcal{V} - \{\vec{X}, \vec{Y}\}$.
- Halpern's version of D4 said $[\vec{W} \leftarrow \vec{w}, X \leftarrow x](X = x)$. Using D0, D7, and D8 (and some standard modal logic reasoning), it is easy to see that the two versions are equivalent.
- Halpern had slightly different versions of D9 and D10. Specifically, the second conjunct is Halpern's version of D9 is $\vee_{x \in \mathcal{R}(X)} [\vec{Y} \leftarrow \vec{y}](X = x)$. For finite signatures, our version of D9 is equivalent to Halpern's in the presence of the other axioms, as we prove in the full paper (Halpern and Peters 2021).
- Finally, Halpern also had an additional axiom D11; we discuss this below.

As mentioned before, the language $\mathcal{L}(\mathcal{S})$ considered here differs from the language considered by Halpern (2000), which we denote $\mathcal{L}_H(\mathcal{S})$, in two ways. First, Halpern implicitly assumed that all interventions were allowed, so he did not have the restriction to allowed interventions. That is, all formulas of the form $[\vec{Y} \leftarrow \vec{y}]\varphi$ were included in $\mathcal{L}_H(\mathcal{S})$, where $\vec{Y} \subseteq \mathcal{V}$ and $\vec{y} \subseteq \mathcal{R}(\vec{Y})$. Second, the causal formulas in $\mathcal{L}_H(\mathcal{S})$ were built from atomic events of the form $X(\mathbf{u}) = x$ as opposed to the form $X = x$. Halpern (2000) gave semantics to formulas with respect to models M, not with respect to pairs $(M, \mathbf{u})$. In Halpern's semantics, $M \models [\vec{Y} \leftarrow \vec{y}](X(\mathbf{u}) = x)$ if all outcomes $\mathbf{v} \in M(\mathbf{u}, \vec{Y} \leftarrow \vec{y})$ satisfy $X = x$. It is easy to see that $M \models [\vec{Y} \leftarrow \vec{y}](X(\mathbf{u}) = x)$ iff, in the semantics of this paper, $(M, \mathbf{u}) \models [\vec{Y} \leftarrow \vec{y}](X = x)$. To deal with the richer language, Halpern (2000) had an additional axiom:

D11. $\langle \vec{Y} \leftarrow \vec{y} \rangle (\varphi_1(\mathbf{u}_1) \wedge \ldots \wedge \varphi_k(\mathbf{u}_k)) \Leftrightarrow (\langle \vec{Y} \leftarrow \vec{y} \rangle \varphi_1(\mathbf{u}_1) \wedge \ldots \wedge \langle \vec{Y} \leftarrow \vec{y} \rangle \varphi_k(\mathbf{u}_k)$, if $\varphi_i(\mathbf{u}_i)$ is a Boolean combination of formulas of the form $X(\mathbf{u}_i) = x$ and $\mathbf{u}_i \neq \mathbf{u}_j$ for $i \neq j$.

D11 is used in Halpern's completeness proof only to reduce consideration from formulas that mention multiple contexts to formulas that mention only one context, which are easily seen to be equivalent to formulas in $\mathcal{L}(\mathcal{S})$. We can show that the axioms without D11 are sound and complete for $\mathcal{L}(\mathcal{S})$ using exactly the same proof as used by Halpern to show that the axioms with D11 are sound and complete for $\mathcal{L}_H(\mathcal{S})$, just skipping the step that uses D11 to reduce to formulas involving just one context. This is formalized in the following theorem, where a signature $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R}, \mathcal{I})$ is *universal* if $\mathcal{I} = \mathcal{I}_{univ}$, the set of all interventions.

Theorem 2.1: (Halpern 2000) *If $\mathcal{S}$ is a universal signature, then AX^+ (resp., AX^+_{rec}) is a sound and complete axiomatization for the language $\mathcal{L}(\mathcal{S})$ for SEMs (resp., acyclic SEMs) with a universal signature $\mathcal{S}$.*

As we shall see (Theorem 4.6), the assumption that $\mathcal{S}$ is universal is critical here; Theorem 2.1 is not true in general without it.

3 GSEMs

In this section, we briefly review the definition of GSEMs (Peters and Halpern 2021); we encourage the reader to consult (Peters and Halpern 2021) for more details and intuition. We also prove some results regarding the extent to which the language $\mathcal{L}(\mathcal{S})$ characterizes GSEMs, and introduce the class of *acyclic* GSEMs. We include a few results from (Peters and Halpern 2021) to help set the scene.

The main purpose of causal modeling is to reason about a system's behavior under intervention. A SEM can be viewed as a function that takes a context $\mathbf{u}$ and an intervention $\vec{Y} \leftarrow \vec{y}$ and returns a set of assignments to the endogenous variables (i.e., a set of outcomes), namely, the set of all solutions to the structural equations after replacing the equations for the variables in $\vec{Y}$ with $\vec{Y} = \vec{y}$. Viewed in this way, generalized structural-equations models (GSEMs) are a generalization of SEMs. In a GSEM, there is a function that takes a context $\mathbf{u}$ and an intervention $\vec{Y} \leftarrow \vec{y}$ and returns a set of outcomes. However, the outcomes need not be obtained by solving a set of suitably modified equations as they are for SEMs–they may be specified arbitrarily. This relaxation gives GSEMs the ability to concisely represent dynamical systems and other systems with infinitely many variables, and the flexibility to handle situations involving finitely many variables that cannot be modeled by SEMs.

Formally, a *generalized structural-equations model (GSEM)* M is a pair $(\mathcal{S}, \mathbf{F})$, where $\mathcal{S}$ is a signature, and $\mathbf{F}$ is a mapping from contexts and interventions to sets of outcomes. As before, a signature $\mathcal{S}$ is a quadruple $(\mathcal{U}, \mathcal{V}, \mathcal{R}, \mathcal{I})$, except that we no longer require $\mathcal{U}$ and $\mathcal{V}$ to be finite, nor $\mathcal{R}(Y)$ to be finite for all $Y \in \mathcal{U} \cup \mathcal{V}$. The big difference is that $\mathbf{F}$ is a function $\mathbf{F} : \mathcal{I} \times \mathcal{R}(\mathcal{U}) \rightarrow \mathcal{P}(\mathcal{R}(\mathcal{V}))$, where $\mathcal{P}$ denotes the powerset operation. That is, it maps a context $\mathbf{u} \in \mathcal{R}(U)$ and an allowed intervention $I \in \mathcal{I}$ to a set of *outcomes* $\mathbf{F}(\mathbf{u}, I) \in \mathcal{P}(\mathcal{R}(\mathcal{V}))$. As with SEMs, we denote these outcomes by $M(\mathbf{u}, I)$. We require that each outcome $\mathbf{v} \in \mathbf{F}(\mathbf{u}, \vec{X} \leftarrow \vec{x})$ satisfy $\mathbf{v}[\vec{X}] = \vec{x}$, since at a minimum, after intervening to set $\vec{X}$ to $\vec{x}$, the variables $\vec{X}$ should actually have the values $\vec{x}$. Since the semantics of $\models$ as we have given it is defined in terms of $M(\mathbf{u}, I)$, we can define $\models$ for GSEMs in the identical way.

It is shown by Peters and Halpern (2021) that GSEMs generalize SEMs in the following sense: Two causal models M and M', which may either be SEMs or GSEMs, are *equivalent*, denoted $M \equiv M'$, if they have the same signature, and they have the same outcomes; that is, if for all $\vec{X} \subseteq \mathcal{V}$, all values $\vec{x} \in \mathcal{R}(\vec{X})$ such that $\vec{X} \leftarrow \vec{x} \in \mathcal{I}$, and all contexts $\mathbf{u} \in \mathcal{R}(\mathcal{U})$, we have $M(\mathbf{u}, \vec{X} \leftarrow \vec{x}) = M'(\mathbf{u}, \vec{X} \leftarrow \vec{x})$.

Theorem 3.1: *(Peters and Halpern 2021, Theorem 3.1) For all SEMs M, there is a GSEM M' such that $M \equiv M'$.*

Recall that in the introduction we defined two models with signature $\mathcal{S}$ to be $\mathcal{L}(\mathcal{S})$-*equivalent* if they agree on all formulas in $\mathcal{L}(\mathcal{S})$. Call a GSEM M *finitary* if, for all contexts and interventions, the set of outcomes is finite. A GSEM with a finite signature (a finite GSEM) is bound to be finitary, but even infinite GSEMs may be finitary. As shown by Peters and Halpern (2021), equivalence and $\mathcal{L}(\mathcal{S})$-equivalence coincide in SEMs and finitary GSEMs.

Theorem 3.2: *(Peters and Halpern 2021, Theorem 2.1) If M and M' are finitary causal models over the same signature $\mathcal{S}$, then $M \equiv M'$ iff M and M' are $\mathcal{L}(\mathcal{S})$-equivalent.*

As the following example shows, the assumption that M and M' are finitary is critical.

Example 3.3: Consider two GSEMs M, M' with the same signature $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R}, \mathcal{I})$. $\mathcal{V}$ consists of countably many binary endogenous variables, that is, $\mathcal{V} = \{X_1, X_2, \ldots\}$ and $\mathcal{R}(X_i) = \{0, 1\}$ for all i. The models have only one context $\mathbf{u}$ (i.e., $\mathcal{U}$ consists of one exogenous variable with a single value). There is only one allowed intervention, the null intervention $\emptyset$. The outcomes of this intervention are as follows. $\mathbf{F}(\mathbf{u}, \emptyset)$ consists of all assignments to the variables X_i where only finitely many of the X_i take the value 0. $\mathbf{F}'(\mathbf{u}, \emptyset)$ consists of all assignments to the X_i where only finitely many X_i take the value 1. Note that for a finite subset of the variables, restricting the outcomes of either model to that subset yields all assignments to that subset. Hence, a formula of the form $\langle\emptyset\rangle\varphi$ is false in both models if $\neg\varphi$ is valid, and true in both models otherwise, because φ is a finite formula and as such can depend only on finitely many variables. Hence, the distinct models M and M' satisfy the same set of causal formulas over $\mathcal{L}(\mathcal{S})$. ∎

3.1 Acyclic GSEMs

In this subsection, we introduce a class of GSEMs analogous to acyclic SEMs. Just as many SEMs used in practice are acyclic, we expect that many GSEMs of practical interest will also be acyclic. For example, the GSEMs constructed in (Peters and Halpern 2021) to model dynamical systems are acyclic according to our definition.

In SEMs, acyclicity is defined using the notion of *independence*. Recall from Section 1 that given a SEM M and endogenous variables X and Y, we say that Y *is independent of* X (in context $\mathbf{u}$) if the structural equation $F_Y(\mathbf{u}, \dots)$ for Y does not depend on X. An acyclic SEM is a SEM whose endogenous variables $\mathcal{V}$ can be totally ordered (for all contexts $\mathbf{u}$) such that if $X \preceq_{\mathbf{u}} Y$, then X is independent of Y in context $\mathbf{u}$.

We cannot use this definition for GSEMs, since there are no equations. But we can generalize an alternate characterization of acyclicity. In acyclic SEMs, intervening on a variable X does not affect variables preceding X. More precisely, let $\mathcal{V}_{\prec_{\mathbf{u}} X} = \{Y \in \mathcal{V} : Y \prec_{\mathbf{u}} X\}$. Then the (unique) outcome $\mathbf{v}$ of doing $I, X \leftarrow x$ in context $\mathbf{u}$ and the outcome $\mathbf{v}'$ of doing $I, X \leftarrow x'$ in context $\mathbf{u}$ agree on $\mathcal{V}_{\prec_{\mathbf{u}} X}$ ($\mathbf{v}[\mathcal{V}_{\prec_{\mathbf{u}} X}] = \mathbf{v}'[\mathcal{V}_{\prec_{\mathbf{u}} X}]$). In fact, a SEM is acyclic if and only if there are orderings $\preceq_{\mathbf{u}}$ such that this condition holds.

This gives a natural way to extend the definition of acyclicity to GSEMs. Since the condition $\mathbf{v}[\mathcal{V}_{\prec_{\mathbf{u}} X}] = \mathbf{v}'[\mathcal{V}_{\prec_{\mathbf{u}} X}]$ is a condition on outcomes, it makes sense for GSEMs. Acyclic GSEMs may have multiple solutions, so we need to strengthen the condition slightly. Given a set S of outcomes and a subset $\vec{Y}$ of $\mathcal{V}$, define the *restriction of* S *to* $\vec{Y}$, denoted $S[\vec{Y}]$, as $S[\vec{Y}] = \{\mathbf{v}[\vec{Y}] \mid \mathbf{v} \in S\}$.

Definition 3.4: A GSEM M is acyclic if, for all contexts $\mathbf{u}$, there is a total ordering $\prec_{\mathbf{u}}$ of $\mathcal{V}$ such that:

Acyc1. For all $X \in \mathcal{V}$, all $x, x' \in \mathcal{R}(X)$, and all $\vec{Y} \leftarrow \vec{y} \in \mathcal{I}$ with $X \notin \vec{Y}$, we have $M(\mathbf{u}, (\vec{Y} \leftarrow \vec{y}, X \leftarrow x))[\mathcal{V}_{\prec_{\mathbf{u}} X}] = M(\mathbf{u}, (\vec{Y} \leftarrow \vec{y}, X \leftarrow x'))[\mathcal{V}_{\prec_{\mathbf{u}} X}]$.

It is natural to wonder whether this condition needs to involve all variables preceding X. After all, in SEMs, acyclicity is defined in terms of independence, and independence is defined pairwise. Indeed, the pairwise version of this condition is sufficient for SEMs; a SEM M is acyclic if and only if for all contexts $\mathbf{u}$, there is a total ordering $\prec_{\mathbf{u}}$ such that the following holds.

Acyc2. If $Y \prec_{\mathbf{u}} X$, then for all $\vec{Y} \leftarrow \vec{y}$ with $X \notin \vec{Y}$ and x, x', we have $M(\mathbf{u}, (\vec{Y} \leftarrow \vec{y}, X \leftarrow x))[Y] = M(\mathbf{u}, (\vec{Y} \leftarrow \vec{y}, X \leftarrow x'))[Y]$.

Clearly Acyc1 implies Acyc2. In SEMs, they are equivalent.

Proposition 3.5: *If M is a SEM, then M satisfies Acyc1 iff M satisfies Acyc2 (for a fixed context $\mathbf{u}$).*

However, in GSEMs, the two conditions are not equivalent; we claim that the stronger condition Acyc1 is more appropriate for characterizing acyclicity. The following example illustrates why.

Example 3.6: Define a GSEM M with binary variables A, B, C, a single context $\mathbf{u}$, allowed interventions $\mathcal{I} = \{A \leftarrow 0, A \leftarrow 1, B \leftarrow 0, B \leftarrow 1, C \leftarrow 0, C \leftarrow 1\}$, and the outcomes

$$M(\mathbf{u}, C \leftarrow 0) = \{(0,0,0), (1,1,0)\} \text{ and}$$
$$M(\mathbf{u}, C \leftarrow 1) = \{(0,1,1), (1,0,1)\},$$

where (a, b, c) is short for $(A = a, B = b, C = c)$. The outcomes for $A \leftarrow a$ and $B \leftarrow b$ are similar: for example, after the intervention $A \leftarrow a$, $A = a$ and $B = C \oplus a$.

M is not acyclic when acyclicity is defined using Acyc1. To see this, fix an ordering of the variables; since the model is symmetric, we take the ordering A, B, C without loss of generality. Then intervening on C, the last variable in the ordering, changes the outcomes for the other two; $M(\mathbf{u}, C \leftarrow 0)[\{A, B\}] = \{(0,0), (1,1)\}$, but $M(\mathbf{u}, C \leftarrow 1)[\{A, B\}] = \{(0,1), (1,0)\}$, violating Acyc1. This seems to us the correct classification: M should not be acyclic. The fact that intervening on C changes the possible values for (A, B), but both A and B precede C in $\prec_{\mathbf{u}}$, cannot occur in acyclic SEMs. However, M is acyclic when acyclicity is defined using Acyc2. This is because intervening on C does not affect the possible values for A ($A = 0$ and $A = 1$ in the two outcomes for each intervention) or for B ($B = 0$ and $B = 1$ in the two outcomes for each intervention). ∎

As we said above, all the GSEMs introduced in (Peters and Halpern 2021) for modeling of dynamical systems, namely, GSEMs for systems of ordinary differential equations, GSEMs for hybrid automata, and GSEMs for rule-based models are acyclic. The order $\prec_{\mathbf{u}}$ in each case corresponds to the natural notion of time in the dynamical system; intervening on variables at a given time cannot affect variables earlier in time (or at the same time).

In SEMs, acyclicity corresponds to the axiom D6, which captures the weaker Acyc2. To get an axiom for acyclicity for GSEMs, we need a modification of D6 that captures the stronger Acyc1. But we cannot express the full Acyc1, because variables may have infinite ranges, the set $\mathcal{V}_{\prec_{\mathbf{u}}} X$ may be infinite, and the set of interventions may be infinite. Thus, we consider a finitary version of Acyc1.

Given a finite set $\{X_1, \dots, X_k\}$ of variables, finite sets $U_i \subseteq \mathcal{R}(X_i)$, for $i = 1, \dots, k$, and a finite set $\mathcal{I}'$ of interventions, let $X_i \rightsquigarrow_{\mathcal{I}', U_1, \dots, U_k, X_1, \dots, X_k} \vec{X}_{-i}$ describe the following conditions: after performing some intervention in $\mathcal{I}'$, intervening on X_i affects the joint values of the variables $\vec{X}_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k)$. Moreover, for some intervention on X_i, the joint value is in $U_{-i} = U_1 \times \cdots \times U_{i-1} \times U_{i+1} \times \cdots \times U_k$. That is, $X_i \rightsquigarrow_{\mathcal{I}', U_1, \ldots, U_k, X_1, \ldots, X_k} \vec{X}_{-i}$ is an abbreviation of

$$\vee_{x, x' \in U_i, \vec{y} \in U_{-i}, \vec{Z} \leftarrow \vec{z} \in \mathcal{I}', X \notin \vec{Z}} [\vec{Z} \leftarrow \vec{z}, X \leftarrow x](\vec{X}_{i-1} = \vec{y}) \wedge [\vec{Z} \leftarrow \vec{z}, X \leftarrow x'](\vec{X}_{i-1} \neq \vec{y}).$$

Consider the following axiom:

D6$^+$. $\bigvee_{i=1}^{k} \neg(X_i \rightsquigarrow_{\mathcal{I}', U_1, \ldots, U_k, X_1, \ldots, X_k} X_{-i})$.

(There is an instance of this axiom for all choices of $\mathcal{I}', U_1, \ldots, U_k$, and $X_1, \ldots, X_k$.) D6$^+$ is clearly sound in acyclic GSEMs, since in an acyclic GSEM, every finite subset of variables has a maximum element under $\preceq_{\mathbf{u}}$, and this element does not affect the others. Moreover, in SEMs, D6$^+$ implies D6, which Halpern used for acyclic SEMs. This is because D6$^+$ holds iff there is an ordering $\prec_{\mathbf{u}}$ of the endogenous variables such that Acyc1 holds, and D6 holds iff there is an ordering such that the weaker condition Acyc2 holds. As Theorem 5.2 shows, D6$^+$ captures acyclicity.

4 Axiomatizing Finite GSEMs

Our goal is to provide a sound and complete axiomatization of GSEMs. We start with finite GSEMs, that is, GSEMs over a finite signature; in the next section, we consider arbitrary GSEMs. Note that the language $\mathcal{L}(\mathcal{S})$ given above for SEMs makes perfectly good sense for finite GSEMs; the semantics of the language for GSEMs is identical to the semantics for SEMs. Because GSEMs are more flexible than SEMs, they do not satisfy all the axioms in AX^+. As we now show, a strict subset of AX^+ provides a sound and complete axiomatization of finite GSEMs.

Definition 4.1: AX^+_{basic} consists of axiom schema D0, D1, D2, D4, D7, D8, and inference rule MP.

Theorem 4.2: *AX^+_{basic} is sound and complete for finite GSEMs.*

The proof of this and all other results can be found in the full paper (Halpern and Peters 2021). Let $AX^+_{basic,rec}$ consist of the axioms in AX^+_{basic} along with axiom schema D6$^+$. Then

Theorem 4.3: *$AX^+_{basic,rec}$ is sound and complete for acyclic finite GSEMs.*

As shown in (Peters and Halpern 2021), if $\mathcal{S}$ is a universal signature, then SEMs over $\mathcal{S}$ are equivalent in expressive power to finite GSEMs where the axioms in AX^+ hold.

Theorem 4.4: *(Peters and Halpern 2021) If $\mathcal{S}$ is a universal signature, then for every finite GSEM over $\mathcal{S}$ that satisfies the axioms of $AX^+(\mathcal{S})$ there is an equivalent SEM over $\mathcal{S}$, and for every SEM over $\mathcal{S}$ there is an equivalent finite GSEM over $\mathcal{S}$ that satisfies the axioms of $AX^+(\mathcal{S})$.*

Since equivalence is the same as $\mathcal{L}(\mathcal{S})$-equivalence for finitary models (Theorem 3.2), this immediately implies the following.

Corollary 4.5: *If $\mathcal{S}$ is a universal signature, then $AX^+(\mathcal{S})$ is a sound and complete axiomatization for $\mathcal{L}(\mathcal{S})$ for finite GSEMs over $\mathcal{S}$ satisfying $AX^+(\mathcal{S})$.*

Although it may seem trivial, Corollary 4.5 does not hold in general for non-universal signatures, as (Peters and Halpern 2021, Example 3.6) shows. This example is a GSEM over a finite signature $\mathcal{S}$ that satisfies the axioms of $AX^+(S)$ but is not equivalent to a SEM.

The existence of a finite GSEM satisfying AX^+ that is not equivalent to an SEM has a significant implication.

Theorem 4.6: *There is a (non-universal) signature $\mathcal{S}$ for which $AX^+(\mathcal{S})$, although sound, is not complete for SEMs of signature $\mathcal{S}$.*

5 Axiomatizing Infinite GSEMs

Things change significantly in infinite GSEMs. To see just one of the problems, note that if X is a variable with infinite range, then instances of D2 corresponding to X, namely $[\vec{Y} \leftarrow \vec{y}](\bigvee_{x \in \mathcal{R}(X)} X = x)$, are no longer in the language, since the disjunction is infinitary. Moreover, if $\mathcal{R}(X)$ is uncountable and the language includes all formulas of the form $X = x$ for $x \in \mathcal{R}(X)$, then the language will be uncountable. While there is no difficulty giving semantics to this uncountable language, there seem to be nontrivial technical problems when it comes to axiomatizations.

On the other hand, suppose that, for example, the range of X is the real numbers. In practice, we do not want to make statements like $X = \pi^3 - e$. It should certainly suffice in practice to be able to mention explicitly only countably many real numbers. (Indeed, we expect that, in practice, it will suffice to talk explicitly about only finitely many real numbers.) Similarly, it should suffice to talk explicitly about only countably many variables and interventions. To get a countable language, we thus proceed as follows.

Given a signature $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R}, \mathcal{I})$, let $\vec{W}$ be a countable subset of $\mathcal{V}$; we call the elements of $\vec{W}$ *named variables*. For each named variable X, let $\mathcal{R}'(X)$ be a countable subset of $\mathcal{R}(X)$, except that we require that (a) if $\mathcal{R}(X)$ is finite, then $\mathcal{R}'(X) = \mathcal{R}(X)$ and (b) if $\mathcal{R}(X)$ is infinite, then so is $\mathcal{R}'(X)$. The elements of $\mathcal{R}'(X)$ are called *named values*. Finally, let $\mathcal{I}'$ be an arbitrary countable subset of $\mathcal{I}$, except that we require that if $\vec{X} \leftarrow \vec{x} \in \mathcal{I}'$, then $\vec{X} \subseteq \vec{W}$ and $\vec{x} \subseteq \mathcal{R}'(X)$ and we assume that $\mathcal{I}'$ *is closed under finite differences with* $\mathcal{I}$, so that if $I_1 \in \mathcal{I}'$, $I_2 \in \mathcal{I}$, $(I_1 - I_2) \cup (I_2 - I_1)$ is finite, and $I_2 = \vec{X} \leftarrow \vec{x}$, where $\vec{X} \subseteq \vec{W}$ and $\vec{x} \in \mathcal{R}'(\vec{X})$, then $I_2 \in \mathcal{I}'$. That is, if we are willing to talk about the intervention I_1, and I_2 is an allowable intervention that differs from I_1 only in how it sets a finite number of variables, all of which we are willing to talk about, as well as the values that they are set to, then we should be willing to talk about I_2 as well. The language $\mathcal{L}_{\vec{W}, \mathcal{R}', \mathcal{I}'}(\mathcal{S})$ consists of Boolean combinations of basic causal formulas $[\vec{Y} \leftarrow \vec{y}]\varphi$ where $\vec{Y} \leftarrow \vec{y} \in \mathcal{I}'$ and φ is a Boolean combination of events of the form $X = x$, where $X \in \vec{W}$ and $x \in \mathcal{R}'(X)$. $\mathcal{L}_{\vec{W}, \mathcal{R}', \mathcal{I}'}(\mathcal{S})$ is clearly a sublanguage of $\mathcal{L}(\mathcal{S})$. Intuitively, it consists only of entities (variables, values, and interventions) that can be named. Since there are only countably many entities that can be named, it easily follows that $\mathcal{L}_{\vec{W},\mathcal{R}',\mathcal{I}'}(\mathcal{S})$ is countable. $\mathcal{L}_{\vec{W},\mathcal{R}',\mathcal{I}'}(\mathcal{S})$ is quite expressive. For example, if the exogenous variables are X_t for t ranging over the real numbers, we could choose $\vec{W}$ to be the subset of $\{X_t \mid t \in \mathbb{R}\}$ for which t is rational. Likewise, if each variable X_t ranges over the real numbers, we could choose $\mathcal{R}(X_t)$ to be the rationals.

We are interested in axiomatizing classes of GSEMs essentially using subsets of the axioms in AX^+, but it seems that we need one new inference rule. While we keep axiom D2, it applies only to variables X such that $\mathcal{R}(X)$ is finite. However, even if $\mathcal{R}(X)$ is infinite, we still want to be able to conclude something like $[\vec{Y} \leftarrow \vec{y}](\exists x(X = x))$: after setting $\vec{Y}$ to $\vec{y}$, X takes on *some* value. Of course, we cannot say this, since we have no existential quantification in the language. Although it is far from obvious, the following rule of inference plays the same role as D2 for variables X with infinite ranges.

D2$^+$. Suppose that $S \subseteq \mathcal{R}'(X)$ is finite and contains all the values of X mentioned in the formula $\varphi \Rightarrow [\vec{Y} \leftarrow \vec{y}]\psi$, and some value in $\mathcal{R}'(X)$ not in the formula if there is such a value. Then from $\varphi \Rightarrow \wedge_{x \in S}[\vec{Y} \leftarrow \vec{y}](\psi \Rightarrow (X \neq x))$ infer $\varphi \Rightarrow [\vec{Y} \leftarrow \vec{y}]\neg\psi$.

Note that if $\mathcal{R}(X)$ is infinite, since we have assumed that $\mathcal{R}'(X)$ is infinite if $\mathcal{R}(X)$ is, there will always be an element in $\mathcal{R}'(X)$ that is not mentioned in φ or ψ.

While D2$^+$ may not look anything like D2, we can show that in the case of variables X with finite range, it is equivalent to D2 in the following precise sense:

Proposition 5.1: *If AX^*_{basic} is the result of replacing D2 with D2$^+$ in AX^+_{basic}, then we can derive D2 for variables with finite ranges in AX^*_{basic}. Moreover, D2$^+$ is derivable in AX^+_{basic} for variables X with finite range, in the sense that if $AX^+_{basic} \vdash \varphi \Rightarrow \wedge_{x\in S}[\vec{Y} \leftarrow \vec{y}](\psi \Rightarrow (X \neq x))$ and $\mathcal{R}(X)$ is finite, then $AX^+_{basic} \vdash \varphi \Rightarrow [\vec{Y} \leftarrow \vec{y}]\neg\psi$.*

While D2$^+$ is unnecessary for finite GSEMs, it is necessary for infinite GSEMs. Let $AX^*_{basic}(\mathcal{S}, \vec{W}, \mathcal{R}', \mathcal{I}')$ consist of all the axioms and inference rules in $AX^+_{basic}(\mathcal{S})$ together with D2$^+$, restricted to formulas in $\mathcal{L}_{\vec{W},\mathcal{R}',\mathcal{I}'}(\mathcal{S})$. Then $AX^*_{basic}(\mathcal{S}, \vec{W}, \mathcal{R}', \mathcal{I}')$ is sound and complete for GSEMs over $\mathcal{S}$ (see Theorem 5.2 below).[1]

Considering GSEMs also helps explain the role of some of the other axioms. A GSEM $\mathcal{S}$ is *coherent* if for all interventions $\vec{X} \leftarrow \vec{x}, \vec{Y} \leftarrow \vec{y}$ in $\mathcal{I}$ (with $\vec{X}$ and $\vec{Y}$ disjoint), if $\vec{Y}$ is finite, $\mathbf{v} \in \mathbf{F}(\mathbf{u}, \vec{X} \leftarrow \vec{x})$, and $\mathbf{v}[\vec{Y}] = \vec{y}$, then $\mathbf{v} \in \mathbf{F}(\mathbf{u}, \vec{X} \leftarrow \vec{x}, \vec{Y} \leftarrow \vec{y})$. The intuition for coherence is straightforward: setting the variables in $\vec{Y}$ to values $\vec{y}$ they already have (in some outcome $\mathbf{v} \in \mathbf{F}(\mathbf{u}, \vec{X}, \leftarrow \vec{x})$ resulting from setting $\vec{X}$ to $\vec{x}$) does not affect the outcome $\mathbf{v}$ (so it is also in $\mathbf{F}(\mathbf{u}, \vec{X} \leftarrow \vec{x}, \vec{Y} \leftarrow \vec{y})$). As we show, D3 corresponds to coherence and D6$^+$ corresponds to acyclicity. D10(a) corresponds to each intervention having at least one outcome (in any given context), and D10(b) corresponds to each intervention having at most one outcome, so D10 (i.e., the combination of 10(a) and 10(b)) corresponds to each intervention having a unique outcome. This is made precise in Theorem 5.2 below.

On the other hand, D5 and D9 do not seem meaningful in GSEMs. They do not have analogues if we have infinitely many variables, since we cannot express $\vec{Z} = \vec{z}$, and there are uncountably many complete interventions (interventions of the form $\vec{Y} \leftarrow \vec{y}$ for $\vec{Y} = \mathcal{V} - \{X\}$).

Let $\mathcal{G}^{\emptyset}(\mathcal{S})$ denote the class of GSEMs over $\mathcal{S}$. Let $\mathcal{G}^{\geq 1}(\mathcal{S})$ and $\mathcal{G}^{\leq 1}(\mathcal{S})$ denote the class of GSEMs over $\mathcal{S}$ where each intervention has at least one and at most one outcome, respectively; let $\mathcal{G}^{coh}$ denote the class of coherent GSEMs over $\mathcal{S}$; let $\mathcal{G}^{acyc}$ denote the class of acyclic GSEMs over $\mathcal{S}$. Given a subset A of $\{$D3, D6$^+$, D10(a), D10(b)$\}$, let $\mathcal{A}$ be the corresponding subset of $\{coh, acyc, \geq 1, \leq 1\}$. Let $AX^*_{basic,A}(\mathcal{S}, \vec{W}, \mathcal{R}', \mathcal{I}')$ be the axiom system consisting of the axioms and rules of inference of AX^*_{basic} together with the axioms in A, restricted to the language $\mathcal{L}_{\vec{W},\mathcal{R}',\mathcal{I}'}(\mathcal{S})$. Let $\mathcal{G}^{\mathcal{A}}$ be the class of GSEMs satisfying the properties in $\mathcal{A}$; that is, $\mathcal{G}^{\mathcal{A}} = \bigcap_{P\in\mathcal{A}} \mathcal{G}^P$. Then

Theorem 5.2: *$AX^*_{basic,A}(\mathcal{S}, \vec{W}, \mathcal{R}', \mathcal{I}')$ is sound and complete for the class $\mathcal{G}^{\mathcal{A}}$ of GSEMs with signature $\mathcal{S}$ over language $\mathcal{L}_{\vec{W},\mathcal{R}',\mathcal{I}'}(\mathcal{S})$.*

We remark that the completeness proof requires several nontrivial ideas beyond what is needed for the analogous results for SEMs; see the full paper for details.

Theorem 5.2 shows that each of the axioms D3, D6$^+$, D10(a), and D10(b) independently enforces a corresponding property in GSEMs; namely, coherence, acyclicity, having at most one outcome, and having at least one outcome. Since in finite GSEMs, D6$^+$ is equivalent to D6, and acyclicity is equivalent to the usual acyclicity in SEMs, Theorem 5.2 also implies that each of Halpern's axioms D3, D6, and D10 independently enforce coherence, acyclicity, and unique outcomes in SEMs. In (Peters and Halpern 2021), we make the case that GSEMs are the most general class of causal models that have the same input and output as SEMs (and satisfy effectiveness). Putting the pieces together gives a full picture of how each of Halpern's original axioms relates to the properties of SEMs. The axioms of AX^+_{basic} are just enough to prove statements that hold in *all* causal models with the same input and output as SEMs (and satisfying effectiveness). Each of the remaining axioms simply independently enforces a natural property of SEMs. This may be of interest completely independently of GSEMs and their applications.

[1]We remark that the soundness of D2$^+$ depends on the fact that we have assumed no structure on the domain, so the only way we have of comparing variable values is by equality. If we assumed an ordering on the domain, so that, for example, we could write $X \geq x$ in addition to $X = x$ and $X \neq x$, then D2$^+$ would no longer be sound. For example, taking $\varphi = true$, $\psi = (X > 2)$, $S = \{1, 2\}$, and $\vec{Y} = \emptyset$, from $X > 2 \Rightarrow (X \neq 2 \wedge X \neq 1)$, we would not want to infer $X \leq 2$! While we can extend D2$^+$ to deal with $>$ and other "nice" relations, pursuing this topic would take us too far afield here.

Acknowledgments

Work supported in part by NSF grants IIS-178108 and IIS-1703846, a grant from the Open Philanthropy Foundation, ARO grant W911NF-17-1-0592, and MURI grant W911NF-19-1-0217.

References

Alur, R.; Courcoubetis, C.; Henzinger, T. A.; and Ho, P.-H. 1992. Hybrid automata: An algorithmic approach to the specification and verification of hybrid systems. In *Hybrid Systems*, 209–229. Springer.

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

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

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

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

Halpern, J. Y.; and Pearl, J. 2005. Causes and explanations: a structural-model approach. Part II: Explanations. *British Journal for Philosophy of Science*, 56(4): 889–911.

Halpern, J. Y.; and Peters, S. 2021. Reasoning about causal models with infinitely many variables. *aaai*. Available at https://arxiv.org/abs/2112.11362.

Laurent, J.; Yang, J.; and Fontana, W. 2018. Counterfactual resimulation for causal analysis of rule-based models. In *Proc. Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI '18)*, 1882–1890.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.

Peters, S.; and Halpern, J. Y. 2021. Causal modeling with infinitely many variables. Available at https://arxiv.org/abs/2112.09171.

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