The Logic of Causal Models

Holger Andreas[1] and Mario Günther[2]

[1]University of British Columbia
[2]LMU Munich and CMU Pittsburgh

Abstract

While causal models are introduced very much like a formal logical system, they have not yet been taken to the level of a proper logic of causal reasoning with structural equations. In this paper, we furnish causal models with a distinct deductive system and a corresponding model-theoretic semantics. Interventionist conditionals will be defined in terms of inferential relations in this logic of causal models.

Keywords. Causal Models, Causal Reasoning, Interventions, Structural Equations.

Contents

1 Introduction

Causal models have become a powerful framework in formal epistemology and knowledge representation. They are used to study actual causation, causal explanation, discovery of causal structures, and conditionals (see, e.g., Pearl (2009), Woodward (2003), Halpern (2016), Briggs (2012), and Schulz (2011)). We focus on deterministic causal models with structural equations in this paper.

From a logical point of view, it is striking that deterministic causal models are introduced very much like a formal logical system. Yet, there are some noteworthy differences. While the semantics of interventionist conditionals uses model-theoretic concepts, we do not have a model-theoretic semantics for inferences in a causal model. Nor do we have a system of natural deduction that allows us to capture such inferences. Halpern (2000) and Briggs (2012) devised axiomatizations of interventionist conditionals using causal models. These axiomatizations, however, do not give us a logic of causal reasoning for drawing inferences from a set of structural equations. They define a logic of conditionals, but not a logic of causal reasoning with structural equations as premises.[1] The notion of a structural equation itself has a syntactic flavor, but its “official definition” in Halpern (2000) and Halpern and Pearl (2005) is a semantic one.

[1] For an investigation of the relation between conditionals based on causal models and conditionals based on possible worlds, see Halpern (2013) and Huber (2013).

Let us briefly explain and exemplify causal reasoning with structural equations. Suppose T stands for the proposition that a rock is thrown against the window, while B says that the window breaks. Take then the following structural equation:

$$B = T.$$

This equation tells us that throwing a rock may cause the window to break. If T is given, we can infer B from it, and this inference has a causal meaning. Potential causes of an event are on the right-hand side of a structural equation, while effects are on the left-hand side. The notion of a structural equation stands for a nonsymmetric determination of a variable by the values of certain other variables. And this nonsymmetry is supposed to mirror the nonsymmetry of causal relations. If C is a cause of E, then we cannot infer from this that E is also a cause of C. This contrasts with the symmetry of the biconditional $\leftrightarrow$ and the identity predicate $=$ in classical logic. In the absence of causal loops, a structural equation represents an asymmetric determination.

In light of a structural equation being nonsymmetric, we can distinguish between two types of inferences with structural equations. First, inferences from causes to effects, and second, inferences from effects to causes. The latter type of inference is commonly called *abductive*. Our logic of causal reasoning is *forward-directed* in the sense that it is only about inferences from causes to effects.

The logic of causal reasoning in this paper centers on a syntactic notion of a structural equation. We consider the equality symbol of such an equation a distinct logical symbol, and introduce natural deduction rules for it. These rules are supplemented by a model-theoretic semantics. We introduce the notion of a causal model as a set of structural equations, and describe the interpretation of such a model. The present logic of causal reasoning thus allows us to explain the notions of a structural equation and a causal model in a standard logical format. Once we have introduced the inference rules and the semantics of the new logical symbol of nonsymmetric determination, knowledge of causal models requires little more than knowledge of classical logic.

Why devise a logic of causal reasoning – from causes to effects – using structural equations? First, from a logical point of view, it seems worth exploring whether or not the framework of causal models may be represented in standard logical format. Second, propositional causal models become simpler and easier accessible. Third, some ambiguities in the representation of deterministic causal models in Pearl (2009) and Halpern (2000) are resolved. Finally, a syntactic account of structural equations seems more in line with the representation of such equations in automated systems and human cognition.

One word on the arity of variables in a causal model is in order. We begin with propositional causal models, which are restricted to binary variables. This restriction will be lifted in Section 10 using concepts from many-sorted first-order logic.

2 A Simple Example

Let us illustrate the motivation for a logic of causal reasoning with a simple and well-known example (Halpern and Pearl 2005, p. 861):

Example 1. **Suzy & Billy Throw Rocks at a Bottle**
"Suzy and Billy both pick up rocks and throw them at a bottle. Suzy's rock gets there first, shattering the bottle. Since both throws are perfectly accurate, Billy's would have shattered the bottle had it not been preempted by Suzy's throw."

Halpern and Pearl (2005, pp. 861-864) suggest modeling the example using five endogenous variables:

- ST: Suzy throws her rock.
- BT: Billy throws his rock.
- SH: Suzy's rock hits the bottle.
- BH: Billy's rock hits the bottle.
- BS: the bottle shatters.

The following causal network represents the dependences among these variables:

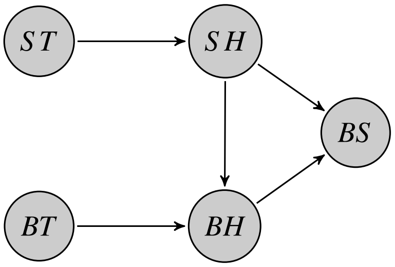

Figure 1: Causal network for the Suzy & Billy Throw Rocks at a Bottle Example.

We have the following structural equations:

- $SH = ST$.
- $BH = BT \wedge \neg SH$.
- $BS = SH \vee BH$.

All variables are understood as binary. The two classical truth values form the range of all variables. A structural equation of the form $A = \phi$ means that the truth value of A is determined by the truth value of ϕ. If ϕ is true, then A must be so. If ϕ is false, then A must be false. These determinations have a causal meaning: ϕ is a sentences about the causes of A. The other direction of determination, from A to ϕ, has a different meaning: if we infer ϕ from A, we go from a given effect to potential causes of this effect. As is standard, we call such an inference *abductive*. Inferences from causes to effects, by contrast, are called *forward-directed* here. Recall that our logic of causal reasoning aims to capture only the latter type of inferences.

Abductive inferences play a major role in Pearl (2009), but are less important when we analyse actual causation in a deterministic setting (see, e.g., Halpern (2016)). Suppose we know the above structural equations, but have no information as to which events actually occur. These equations then enable us to make a few straightforward observations: if Suzy throws her rock, the bottle will shatter. Likewise if Billy throws his. If Suzy or Billy throws a rock, the bottle will shatter as well. If the bottle shatters, we may infer that Suzy or Billy have thrown a rock, but this inference is of the abductive type. As just explained, our logic of causal reasoning aims to respect the nonsymmetry of a structural equation in the sense that it is confined to causal inferences from causes to effects.

Understanding the structural equations above requires little more than propositional logic. When writing a research paper on actual causation, it would therefore be helpful to instruct the reader along the following lines. We can understand the equations in the sense of classical, propositional logic, with one important qualification: whenever we use a structural equation for an inference, we infer the left-hand side from the right-hand side, but not the other way around. Even the consequences of an intervention on a variable can be represented by corresponding inferences. Instead of the structural equations of the original causal model, we use the equations of the submodel induced by the intervention. We will explain the notions of an intervention and a submodel in Section 8.

3 Structural Equations

What is a structural equation? Let us recall the original definition of a causal model in Pearl (2009, p. 203):

> A causal model is a triple $M = \langle U, V, F \rangle$, where
>
> (i) U is a set of background variables, (also called exogenous), that are determined by factors outside the model;
>
> (ii) V is a set $\{V_1, V_2, \dots, V_n\}$ of variables, called endogenous, that are determined by variables in the model – that is, variables in $U \cup V$; and
>
> (iii) F is a set of functions $\{f_1, f_2, \dots, f_n\}$ such that each f_i is a mapping from (the respective domains of) $U_i \cup PA_i$ to V_i, where $U_i \subseteq U$ and $PA_i \subseteq V \setminus V_i$ and the entire set F forms a mapping from U to V. In other words, each f_i in
>
> $$v_i = f_i(pa_i, u_i), i = 1, \dots, n,$$
>
> assigns a value to V_i that depends on (the values of) a select set of variables in $V \cup U$, and the entire set F has a unique solution $V(u)$.

This notion of a causal model contains syntactic and semantic elements. U and V are sets of variables, while F is a set of functions. A function is a semantic entity in the sense of formal logic since functions serve as interpretations of function symbols. Variables, by contrast, are syntactic entities since they can be interpreted.

Condition (ii) of the above definition suggests that the functions in F define structural equations of the form $v_i = f_i(pa_i, u_i)$. By using the phrase ‘in other words’, Pearl seems to say that we can switch back and forth between functions in F and structural equations. But this is not entirely correct for the following reasons. If F is a set of functions – understood as mappings from a domain to a codomain – then we do not know which of these functions is supposed to determine which variable. So, it seems more accurate to say that F is a sequence of functions such that the i-th element of this sequence determines the value of variable V_i. In fact, the notation $\{f_1, \dots, f_n\}$ suggests that F is understood as a totally ordered set.

However, some ambiguities remain, even if it is made explicit which function in F determines which variable. Take, for example, the structural equation for the variable SH in the above causal scenario:

$$SH = ST.$$

Suzy's rock hits the bottle if Suzy throws her rock. The function f_{SH} – understood as a mapping from the domain of ST to the domain of SH – is as follows:

$$\{\langle T, T\rangle, \langle F, F\rangle\}.$$

This notation assumes that functions are represented by binary many-to-one relations, as is standard (Halmos 1974). Such a relation assigns every object of its domain a unique object of its codomain. The problem is that f_{SH} does not tell us what the parent variable of SH is. Suppose, for example, we change the causal model such that Suzy's rock hits the bottle if Billy throws his rock, while the other equations remain unchanged. Instead of $SH = ST$, we have:

$$SH = BT.$$

Then, BT is the parent variable of SH. Such an alternate causal model yields different causal judgments since it does not verify the same interventionist conditionals as the original model. However, we still obtain $f_{SH} = \{\langle T, T\rangle, \langle F, F\rangle\}$. The function f_{SH} remains unchanged. It is therefore not entirely correct to say that the set F of functions contain the same information as a set of structural equations of the form $v_i = f_i(pa_i, u_i)$ do. Nor is it entirely correct to say that a causal model $M = \langle U, V, F\rangle$ has a unique directed graph that represents the causal structure of M.[2] While these observations may seem pedantic, they are made for a reason. We take a closer look at the relationship between syntactic and semantic elements of causal models in order to devise a deductive system for structural equations.

The account of causal models in Halpern (2000) and Halpern and Pearl (2005) is more explicit about the relation between syntactic and semantic components of a causal model. Let $\mathcal{U}$ be the set of exogenous variables and $\mathcal{V}$ the set of endogenous variables. $\mathcal{R}(X)$ is the range of variable X. For each endogenous variable X of the causal model, there is a function that maps the Cartesian product of the ranges of all variables (but X) onto the range of the variable X (Halpern 2000, p. 318n):

[2] Subsequent to defining the notion of a causal model $M = \langle U, V, F\rangle$, Pearl (2009, p. 203) seems to make this claim.

> A causal model over signature $\mathcal{S}$ is a tuple $T = \langle \mathcal{S}, \mathcal{F} \rangle$ where $\mathcal{F}$ associates with each variable $X \in \mathcal{V}$ a function denoted F_X such that $F_X : (\times_{U \in \mathcal{U}} \mathcal{R}(U)) \times (\times_{Y \in \mathcal{V} \setminus \{X\}} \mathcal{R}(Y)) \mapsto \mathcal{R}(X)$. F_X tells us the value of X given the values of all the other variables in $\mathcal{U} \cup \mathcal{V}$. We think of the functions F_X as defining a set of (modifiable) structural equations, relating the values of the variables.

For this notion of a causal model, it is well defined which function is supposed to agree with which variable. The function F_X must have the same value as the variable X. However, there remains an ambiguity concerning the parent variables of a given variable. Suppose the cardinality of the set $\mathcal{U} \cup \mathcal{V}$ is n. Then, F_X assigns every $(n-1)$-tuple of $(\times_{U \in \mathcal{U}} \mathcal{R}(U)) \times (\times_{Y \in \mathcal{V} \setminus \{X\}} \mathcal{R}(Y))$ a unique object in $\mathcal{R}(X)$. Each element of such an $(n-1)$-tuple stands for a possible value of a certain variable in $\mathcal{U} \cup \mathcal{V} \setminus \{X\}$. But we do not know which element of a given $(n-1)$-tuple interprets which variable. In our rock-throwing example, we do not know whether the first element of the tuple $\langle T, F, T, F, T \rangle$ stands for the truth value of ST or for the truth value of BT, or for the truth value of any of the other variables. So, $(n-1)$-tuples of $(\times_{U \in \mathcal{U}} \mathcal{R}(U)) \times (\times_{Y \in \mathcal{V} \setminus \{X\}} \mathcal{R}(Y))$ must be understood against the background of some total order of the variables in $\mathcal{U} \cup \mathcal{V}$. Then, such tuples can be understood as possible interpretations of the variables in $\mathcal{U} \cup \mathcal{V} \setminus \{X\}$. And the notion of a solution of a set of structural equations (defined by functions F_X) is well defined. In fact, this seems to be the intended meaning of a set of structural equations – defined by functions F_X – in Halpern (2000) and Halpern and Pearl (2005).

With some minor clarifications added, Halpern (2000) and Halpern and Pearl (2005) thus give us a well-defined notion of a structural equation. But this notion seems unnecessarily complex. It also seems unnecessarily semantic, at least if we view the general notion of an equation from the vantage point of logic. The semantic notion of a structural equation lacks a well-defined syntactic counterpart, even though concrete causal models are commonly represented with syntactic structural equations. From a logical point of view, it seems worth exploring if we can furnish causal models with a distinct deductive system and a distinct model-theoretic semantics. This is the objective of the present paper.

Notice, finally, that it is often computationally not feasible to represent a structural equation in terms of a mapping of the Cartesian product of the ranges of variables onto the range of another variable if we were to literally represent this mapping as a set of ordered pairs. A syntactic representation of a structural equation in terms of some method of determining its values seems more in line with the working of automated system and human cognition. To give a simple example, if a structural equation uses the mathematical function of addition, we do not literally represent addition as a set of ordered pairs.

4 Propositional Causal Models

Let us now develop a general account of causal models that will serve as foundation for the present logic of causal reasoning. In spirit, we very much follow Halpern (2000) and Halpern and Pearl (2005). The main difference to these accounts is that we define the notion of a structural equation in a syntactic manner. Two further differences are noteworthy. First, our formalism does not require any distinction between exogenous and endogenous variables. Second, interventions are defined for arbitrary Boolean formulas, not only for conjunctions of atomic formulas.[3]

Causal models represent law-like relations between events. Some events occur and so are actual, other events do not occur and so are non-actual. Any event occurs or does not. We represent events by propositional variables. The truth value of the propositional variables denotes whether or not an event occurs. A being true means that the corresponding event occurs, while A being false means that the event in question does not occur.

Let P be a set of propositional variables such that every member of P represents a distinct event. $\mathcal{L}_P$ is a propositional language whose logical symbols are the Boolean connectives. It is defined recursively in the standard way: (i) Any $A \in P$ is a formula. (ii) If ϕ is a formula, then so is $\neg\phi$. (iii) If ϕ and ψ are formulas, then so are $\phi \vee \psi$ and $\phi \wedge \psi$. (iv) Nothing else is a formula.

We explain the notion of a structural equation in a syntactic fashion. Let A be a propositional variable of $\mathcal{L}_P$, and ϕ be a formula of that language. Then,

$$A = \phi$$

is a structural equation based on $\mathcal{L}_P$. Nothing else is a structural equation. The intended meaning of such an equation is that the truth value of ϕ determines that of A. This determination has a causal meaning: it goes from causes to an effect. We shall come to see at a later stage how this directedness of a structural equation is

[3] The latter generalization has already been realized in Briggs (2012). We agree with Briggs (2012) on the general strategy for defining such interventions and corresponding conditionals.

expressed in the formalism. Since $\mathcal{L}_P$ is a propositional language, $A = \phi$ is not a formula of $\mathcal{L}_P$.

Definition 1. Causal Model (binary variables)
Let M be a set of structural equations, based on the language $\mathcal{L}_P$. For an $\mathcal{L}_P$ sentence ϕ, $Var(\phi)$ is the set of propositional variables that occur in ϕ. M is a causal model iff it satisfies two conditions:

(1) For any $A \in P$, there is at most one $\sigma \in M$ such that σ has the logical form $A = \phi$.

(2) If $A = \phi$ is a member of M, then there is no ϕ' such that (i) ϕ and ϕ' are (classically) logically equivalent, and (ii) $Var(\phi') \subset Var(\phi)$.

In brief, a causal model M is a set of structural equations such that every propositional variable in $\mathcal{L}_P$ has at most one occurrence on the left-hand side of a structural equation in M. Further, no structural equation in M must have vacuous occurrences of a propositional variable. We call the occurrence of a variable in an equation $A = \phi$ *vacuous* iff the value of this variable never makes a difference to the value of ϕ and A. For example, in the equation $SH = ST \wedge (BT \vee \neg BT)$, the variable BT occurs vacuously.

There remains to define the notions of a parent variable and a causal graph of M. We say that the propositional variable A is a parent variable of the propositional variable B – and write $(A, B) \in PA_M$ – iff there is $\sigma \in M$ such that σ has the logical form $B = \phi$ and A occurs in ϕ. A being a parent variable of B means that the value of A determines the value of B. Note that we have excluded vacuous determinations by condition (2) of the above definition of a causal model. To give an example, $B = A \vee \neg A$ cannot be a member of a causal model M according to this condition.

The causal graph G_M of M is the ordered pair $\langle P, PA_M \rangle$. So, the directed edges of G_M are given by the ordered pairs (A, B) of variables such that A is a parent of B. The causal model M is called *acyclic* iff G_M is a directed acyclic graph. (In such a graph, there is no directed path that starts and ends at one and the same node.) Note that the above definition of a causal model does not exclude cycles. All subsequent definitions and theorems apply to both cyclic and acyclic causal models.

5 Natural Deduction

In this section, we set forth a system of natural deduction for reasoning with structural equations. Suppose Γ is a set of Boolean formulas in the language $\mathcal{L}_P$. Further, suppose M is a set of structural equations based on the language $\mathcal{L}_P$. Which formulas can be derived from Γ and M? We aim to answer this question by devising a system of natural deduction that defines a relation $\Gamma \vdash_M \phi$.

Obviously, our system needs to include the inference rules of the Boolean connectives: negation, disjunction, and conjunction, including rules for the introduction and elimination of a contradiction $\bot$. We assume the reader is familiar with some formulation of these rules.

So, there remains the logical symbol = of nonsymmetric determination. This is the genuinely novel logical symbol of causal models. Fortunately, reasoning with structural equations is static in the following sense: we normally do not infer a new structural equation from a given set of structural equations. In light of this, there is no need for an introduction rule of the logical symbol =. Another reason for not having such a rule will be given at the end of this section.

Which elimination rules best capture our inferences from a set of structural equations? Let us start with a simple proposal:

$$\frac{A = \phi \qquad \phi}{A}.$$

In words, if $A = \phi$ is a member of M and ϕ can be derived, then A. In addition to this rule, we need an inference rule for deriving $\neg A$:

$$\frac{A = \phi \qquad \neg\phi}{\neg A}.$$

However, this pair of inference rules fails to break the symmetry between the left-hand and the right-hand side of a structural equation. We could still draw inferences from effects to causes, and so go against the direction of causation. To give an example, let M be the structural equations of the rock-throwing example from Section 2. Our premise set Γ is given by $\{BS\}$ (which says that the bottle shatters). Then, lets start a subderivation with the assumption $\neg ST \wedge \neg BT$ (which says that Suzy does not throw her rock and Billy does not throw his either). From this assumption, we can derive $\neg BS$ (which means that the bottle does not shatter). This conclusion contradicts the single member of our premise set $\Gamma = \{BS\}$. By Negation Introduction, we can therefore infer $\neg(\neg ST \wedge \neg BT)$. From this, we can derive

$ST \vee BT$, which says that Suzy or Billy throws a rock. Clearly, we have inferred a sentence about potential causes from a sentence about an effect. The above rules fail to express the nonsymmetry of structural equations.

Therefore, we need to impose further constraints on the application of the above rules. Let $Var(\Gamma)$ be the set of propositional variables that occur in at least one formula of Γ, our set of premises. Then, we require that a structural equation $A = \phi$ can only be used if A has no occurrences in any formula of Γ:

$$\frac{A = \phi \qquad \phi}{A} \qquad [A \notin Var(\Gamma)] \qquad (= \text{Elim}_1)$$

$$\frac{A = \phi \qquad \neg\phi}{\neg A} \qquad [A \notin Var(\Gamma)]. \qquad (= \text{Elim}_2)$$

This simple constraint on the application of the inference rules for = does the trick. It blocks inferences from effects to causes, but allows causal reasoning from causes to effects. To resume our running example, suppose once more $\Gamma = \{BS\}$ (which means that the bottle shatters). Then, the constraint $[A \notin Var(\Gamma)]$ disallows using the structural equation $BS = SH \vee BH$ in whatever derivation from the premise set $\{BS\}$. For $Var(\Gamma) = \{BS\}$ and BS occurs on the left-hand side of the structural equation in question. Hence, there is no way to infer $ST \vee BT$ from BS. Abductive inferences – which go from effects to causes – are blocked, as desired.

It is worth noting that the constraint $[A \notin Var(\Gamma)]$ translates the operation of an intervention into the language of natural deduction. Recall that an intervention on a contextualized causal model $\langle M, \vec{u} \rangle$ by $\vec{X} = \vec{x}$ assigns a specific value to the variable X – for all X that are a member of $\vec{X}$ – such that any assignment by the function F_X is overruled.[4] Put simply, the structural equation defined by F_X becomes irrelevant once we intervene on X. In a similar vein, the structural equation $A = \phi$ becomes irrelevant if we make an assumption about A in the premise set Γ. This set thus expresses an intervention on certain variables in P, which may well be logically complex in the sense of containing disjunctions and conjunctions. We will define interventionist conditionals using the notion of an interpreted causal model in Section 8.

Let LC be the logic whose inference rules are given by the standard natural deduction rules for Boolean connectives, extended by the rules $(= \text{Elim}_1)$ and $(= \text{Elim}_2)$.

[4] See Halpern and Pearl (2005, Sec. 2) or Pearl (2009, Ch. 7).

LC simply stands for the *logic of causal models*. The definition of a derivation in *LC* is now straightforward:

Definition 2. $\Gamma \vdash_M \phi$
Let Γ be a set of $\mathcal{L}_P$ sentences, and let ϕ be such a sentence. M is a causal model, based on the language $\mathcal{L}_P$. Let $\mathcal{L}_{PM}$ be given by the set of $\mathcal{L}_P$ sentences united with M. We say that ϕ is derivable from Γ and M – and write $\Gamma \vdash_M \phi$ – iff there is a tree of $\mathcal{L}_{PM}$ sentences that satisfies the following conditions:

(1) The topmost sentences are either in $\Gamma \cup M$ or discharged by an inference in the tree.

(2) The bottommost sentence is ϕ.

(3) Every sentence in the tree, except ϕ, is a premise of an application of an inference rule of *LC* such that the conclusion of this application stands directly below that sentence.

We write $\Gamma \vdash_M \phi$ instead of $M, \Gamma \vdash \phi$ in order to emphasize that M and Γ contain different types of premises. Alternatively, we could also write $\langle M, \Gamma \rangle \vdash \phi$. The next step is to introduce the semantics of *LC*.

Let us finally resume the discussion of an introduction rule for the equality symbol. If such a rule was available, we could derive new structural equations from a given set of such equations. Why is this not desirable? Following Halpern and Pearl (2005, p. 847), we take a structural equation to represent a "distinct mechanism (or law) in the world". We further assume that such mechanisms are elementary in the sense that the causal model does not provide us with any information about any submechanisms. The structural equations of a causal model are thus comparable to atomic sentences in truth-value semantics: as atomic sentences are used to explain the semantics of logically complex sentences, so are structural equations used to analyse complex causal inferences and judgements.

6 Semantics

Recall that a structural equation $A = \phi$ in M simply pairs a propositional variable A with a propositional formula. We can therefore define the semantics of structural equations in terms of the semantics of propositional logic. As is well known, the semantics of a propositional language centers on the notion of an assignment of truth values to the propositional variables. A value assignment $v : P \mapsto \{T, F\}$ maps the set P of propositional variables to the set of truth values. This in mind, we define what it is for a valuation v to satisfy a structural equation:

$$v \models A = \phi \text{ iff, } v \models_{Cl} A \text{ iff } v \models_{Cl} \phi. \qquad \text{(Def } v \models A = \phi\text{)}$$

In simpler terms, the valuation v satisfies the structural equation $A = \phi$ iff both sides of the equation have the same truth value on v. Notably, at this stage, the semantics of = does not differ from the semantics of the standard biconditional ↔ of classical logic. $\models_{Cl}$ stands for the satisfaction relation in classical propositional logic.

The satisfaction relation for sets that contain propositional formulas and structural equations can now be defined in the standard way. Let Δ be a set of formulas such that any $\delta \in \Delta$ is either an $\mathcal{L}_P$ formula or a structural equation based on $\mathcal{L}_P$.

$$v \models \Delta \text{ iff, for all } \delta \in \Delta, v \models \delta. \qquad \text{(Def } v \models \Delta\text{)}$$

It seems as if we could define the relation of logical entailment in a straightforward manner as well:

$$\Delta \models \phi \text{ iff, for all } v \text{ s.t. } v \models \Delta, v \models \phi.$$

However, this relation of logical entailment fails to capture the nonsymmetry of the equality symbol in a structural equation. For a valuation v satisfies a structural equation $A = \phi$ iff v satisfies $A \leftrightarrow \phi$. The two formulas have the same truth conditions.

How can we express the nonsymmetric determination of the equality symbol = within a relation of entailment? Recall that we view the premise set Γ – with regard to the relation $\Gamma \vdash_M \phi$ – as expressing an intervention on the propositional variables that have occurrences in Γ. Once we intervene on a variable A, the structural equation determining A – if there is one – becomes irrelevant for the determination of A. Hence, we can say that a possibly complex intervention Γ on the causal model M turns M into a set M_Γ:

$$M_\Gamma = \{\sigma \mid \sigma \in M, \sigma \equiv A = \phi, \text{ and } A \notin Var(\Gamma)\}. \qquad \text{(Def } M_\Gamma\text{)}$$

M_Γ is the subset of M such that $A = \phi$ is in M_Γ iff A does not occur in any formula of Γ. Using this operation on a set M of structural equations, we can define the relation of entailment for our logic LC:

Definition 3. $\Gamma \models_M \phi$
Let Γ be a set of $\mathcal{L}_P$ sentences and let ϕ be such a sentence. M is a set of structural equations, based on the language $\mathcal{L}_P$. We say that ϕ is entailed by Γ and M – and write $\Gamma \models_M \phi$ – iff, for all valuations v such that $v \models \Gamma \cup M_\Gamma$, $v \models \phi$.

Notice that the nonsymmetry of = comes into play through an intervention on M, which is expressed by the operation of turning a set M of structural equations into a set M_Γ of such equations. M_Γ is obtained from M by eliminating all structural equations that determine a variable that occurs in a formula of Γ. The union of Γ and M_Γ plays a role that is analogous to the notion of a submodel in Pearl (2009, p. 204). This will become more obvious when we define interventions on interpreted causal models in Section 8.

7 Soundness and Completeness

The relation $\Gamma \models_M \phi$ of entailment is aimed to capture the relation $\Gamma \vdash_M \phi$ of derivability. Or shall we say that the derivability relation for causal models is aimed to capture the entailment relation? More important than this question of priority is that semantics and proof theory are in harmony with one another. We show soundness and completeness in this section. Let Γ be a set of $\mathcal{L}_P$ sentences, M be as introduced above, and let ϕ be an $\mathcal{L}_P$ sentence.

Theorem 1. **Soundness**
If $\Gamma \vdash_M \phi$, then $\Gamma \models_M \phi$.

Soundness can be proven by induction on the number of inferences in a derivation, as is standard. The inductive step for the inference rules (= Elim_1) and (= Elim_1) is analogous to the inductive step for (→ Elim) in proofs of soundness in classical logic. Details are spelled out in the appendix.

Theorem 2. **Completeness**
If $\Gamma \models_M \phi$, then $\Gamma \vdash_M \phi$.

The proof in the appendix exploits two facts. First, a structural equation $A = \phi$ is satisfied by a truth value interpretation v iff $A \leftrightarrow \phi$ is satisfied by v, where $\leftrightarrow$ has its classical meaning. Second, propositional classical logic whose logical symbols are the Boolean connectives is complete.

8 Interventionist Conditionals

So far, we have dealt with uninterpreted causal models, which are given by a set M of structural equations. Let us now extend this analysis to interpreted causal models. Such models are needed to study causal relations in concrete causal scenarios. A given causal scenario is about specific events some of which occur, while others may not occur. Obviously, we can represent information as to which events occur by a set of literals. This gives rise to the notion of an interpreted causal model:

Definition 4. Interpreted Causal Model $\langle M, V\rangle$
$\langle M, V\rangle$ is an interpreted causal model iff M is a causal model and V a classically consistent set of literals, both of which are given in the language $\mathcal{L}_P$.

Recall that a formula is called a *literal* iff it is an atom or the negation of an atomic formula. We say that a set V of literals is complete – relative to $\mathcal{L}_P$ – iff every propositional variable of P occurs in exactly one formula of V. If $\langle M, V\rangle$ is an interpreted causal model, V may or may not be complete. All subsequent definitions, therefore, apply to completely and partially interpreted causal models. While the premises of the inference relation $\vdash_M$ behave like interventions, the set V represents what is often referred to as *observations*.

We thus define the interpretation of a causal model M in terms of a set of literals, thereby assigning a set of syntactic entities a semantic role. This strategy is inspired by Carnap's notion of a state description in Carnap (1947) and the notion of a Hintikka set, first expounded in Hintikka (1955). The main benefit of this strategy is simplicity in the subsequent definitions. Moreover, a set of literals proves more handy and notationally simpler than an interpretation function when we study concrete causal scenarios (see, e.g., Andreas and Günther (2021)).

What does it mean that an interpreted causal model $\langle M, V\rangle$ satisfies a Boolean formula ϕ? We define this relation in terms of an entailment relation that encompasses both forward-directed and abductive causal reasoning:

$$\langle M, V\rangle \models \phi \text{ iff for all } v \text{ s.t. } v \models M \cup V,\ v \models \phi. \qquad (\text{Def } \langle M, V\rangle \models \phi)$$

v stands for a classical propositional valuation, as explained in Section 6. For the expert reader, it may be worth noting that the notion of an interpreted causal model $\langle M, V\rangle$ is a syntactic generalization of a causal model M in a context $\vec{u}$, abbreviated by $\langle M, \vec{u}\rangle$ in Halpern and Pearl (2005). The former notion is more general than the latter since there is no restriction on which variables are given a direct interpretation.

The next step is to define the crucial notion of an interventionist conditional. Let α and γ be two Boolean formulas of $\mathcal{L}_P$. What does it mean to say that γ would be true if α were for an interpreted causal model $\langle M, V\rangle$? To get started, let us first deal with uninterpreted causal models $\langle M, V\rangle$ where V is empty. (We simply take an uninterpreted causal model as a limiting case of an interpreted causal model.) Then, we can define the truth conditions of the conditional 'if α were true, γ would be' in terms of an entailment relation:

$$\langle M, \emptyset\rangle \models [\alpha]\gamma \text{ iff } \alpha \models_M \gamma.$$

To generalize this account of conditionals for interpreted causal models $\langle M, V\rangle$, we need to understand how an intervention by α changes the interpretation V. Roughly speaking, intervening by α means to change the values of some variables such that α becomes true, and then see what happens. The operations of changing values and seeing what happens can be understood logically and experimentally. If there is a mismatch between experimental outcomes and logical predictions, we may have to change the model. The intervention by a disjunction α may be understood via a set of conjunctive interventions, each of which verifies α.[5]

The crucial question is which variables determined by V may change their values as a consequence of an intervention by α? Which other variables need to be kept fixed? The notion of a causal graph G_M of M (explained and defined in Section 4) helps answer this question. If a variable is a descendant of a variable in α or occurrent in α, then its value may change due to the intervention by α. If, by contrast, a variable does not occur in α and is not a descendant of any variable in α, we need to keep its value fixed. In symbols:

$$V_\alpha = \{l \in V \mid Var(l) \cap (Des(Var(\alpha)) \cup Var(\alpha)) = \emptyset\}. \qquad \text{(Def } V_\alpha\text{)}$$

V_α is the core of the interpretation V that remains unmodified in an intervention by α. For the intervention by α does not causally affect any variable interpreted by V_α. As is obvious, $Des(Var(\alpha))$ designates the set of descendants of the variables that occur in α.[6] We can now define an interventionist conditional in terms of the entailment relation $\models_M$:

[5] This approach is pursued in Briggs (2012).

[6] The definition of $Des(Var(\alpha))$ is inspired by Brigg's (2012) account of interventionist conditionals. This is not to say that the present account of interventionist conditionals itself coincides with the one given in Briggs (2012).

Definition 5. $\langle M, V\rangle \models [\alpha]\gamma$
Let $\langle M, V\rangle$ be an interpreted causal model in the language $\mathcal{L}_P$. α and γ are formulas of this language. We say that the conditional 'if α were true, γ would be' is true in $\langle M, V\rangle$ – and write $\langle M, V\rangle \models [\alpha]\gamma$ – iff $V_\alpha, \alpha \models_M \gamma$.

Or, equivalently:

$$\langle M, V\rangle \models [\alpha]\gamma \;\text{ iff }\; V_\alpha, \alpha \vdash_M \gamma.$$

Equivalence between the two definitions follows from soundness and completeness concerning the relations $\vdash_M$ and $\models_M$.

We have thus defined interventionist conditionals in terms of a relatively simple inferential relation. On the side of the premises, we have a set V_α of literals (which represents the interpretation of non-descendant variables) and the antecedent α itself. The conditional 'if α were true, γ would be' holds true iff we can infer the consequent γ from these premises. This inference relation is defined semantically and syntactically.

Let us briefly apply the present account of interventionist conditionals to our running example. $V = \{ST, BT, SH, \neg BH, BS\}$. What happens if Suzy does not throw her rock? Let $\alpha \equiv \neg ST$. Then, $V_\alpha = \{BT\}$. And for the premise set $\Gamma = V_\alpha \cup \{\alpha\}$, $M_\Gamma = \{SH = ST, BH = BT \land \neg SH, BS = SH \lor BH\}$. Obviously, $V_\alpha, \alpha \models_M BS$ and $V_\alpha, \alpha \vdash_M BS$. So, the bottle still shatters, even if Suzy does not throw her rock.

9 Comparison with the Standard Account

The expert reader familiar with the standard account of causal models will be interested in whether or not the present account of interventionist conditionals remains truthful to the definition of such conditionals in Halpern (2000) and related work. There, the semantics of interventionist conditionals is roughly explained as follows (p. 320):

> $[\vec{Y} \leftarrow \vec{y}]X(\vec{u}) = x$ can be interpreted as "in all possible solutions to the structural equations obtained after setting Y_i to $y_i, i = 1, \ldots, k$ and the exogenous variables to $\vec{u}$, random variable X has value x".

While this explanation is confined to conditionals with a consequent of the logical form $X(\vec{u}) = x$, its generalization to arbitrary Boolean formulas is straightforward.

As for the relation between interventionist conditionals in the standard account and the present account, we can prove the following proposition:

Proposition 1. Let $T = (\mathcal{S}, \mathcal{F})$ be a causal model over signature $\mathcal{S}$, where $\mathcal{F}$ is a set of structural equations, as defined in Halpern (2000). Let all variables of the causal model be binary ones. $\vec{u}$ is a vector that represents an assignment to the exogenous variables. Let M be a set of syntactic structural equations such that each equation has a unique semantic representation by a member in $\mathcal{F}$. V is a set of literals that syntactically represents the assignment given by vector $\vec{u}$. α is a conjunction of literals such that this conjunction represents the assignment $\vec{Y} \leftarrow \vec{y}$. γ is an arbitrary Boolean formula. Then, if $T \models [\vec{Y} \leftarrow \vec{y}]\gamma$, then $\langle M, V \rangle \models [\alpha]\gamma$.

We prove this proposition in the appendix. Does the other direction of the proposition also hold? That is, can we show that, if $\langle M, V \rangle \models [\alpha]\gamma$, then the "corresponding" semantic causal model T verifies the "corresponding" conditional $[\vec{Y} \leftarrow \vec{y}]\gamma$? For this to be shown, further assumptions need to be made. First, we need to assume that the antecedent α is a literal or a conjunction of literals. Second, V represents the assignment of the root variables, but does not represent any further assignments. (A root variable is one that does not have ancestors.) In requiring that V is an assignment to the root variables, we assume that the set of exogenous variables simply equals the set of root variables in a causal model. (A variable X is called *exogenous* in the standard account iff there is no semantic structural equation F_x such that this equation determines its value.) On these assumptions, we can show that $T \models [\vec{Y} \leftarrow \vec{y}]\gamma$ follows from $\langle M, V \rangle \models [\alpha]\gamma$, given the obvious translations from the syntactic account of causal models into the semantic one. The proof is analogous to the proof of Proposition 1. Notice, however, that the other direction of Proposition 1 cannot be shown for conditionals with a disjunction in the antecedent since the semantics in Halpern (2000) does not allow for such conditionals.

Conditionals with a disjunction in the antecedent are available in the account by Briggs (2012), though. Allowing for such conditionals comes at a price in her account: substituting a given antecedent α with a logically equivalent antecedent α' may change the truth value of the respective conditional. For instance, in our running example, we have $[BT]BS$, but not $[BT \wedge (BS \vee \neg BS)]BS$. Briggs is willing to pay this price, and so are we. If one thinks this price is too high, one should simply disallow disjunctions in the antecedent.

A similar problem arises with our inference relation $\Gamma \vdash_M \phi$: we may have $\Gamma \vdash_M \phi$, but not $\Gamma \cup \{A \vee \neg A\} \vdash_M \phi$ since the premises of the inference relation $\vdash_M$ behave like interventions. Again, if one thinks this is counterintuitive, one should restrict the logical form of the premises admitted in the relation $\vdash_M$ such that disjunctions are excluded. The problem could also be addressed by requiring that, for a given premise set Γ, there is no Γ' such that Γ and Γ' are classically logically equivalent and $Var(\Gamma') \subset Var(\Gamma)$.

10 Non-Binary Variables

So far, we have studied propositional causal models. In such a model, all variables are binary. Let us now lift this restriction, and define causal models with non-binary variables. For this to be achieved, some elements of first-order logic are needed. What is a non-binary variable in a first-order language? Which causal scenarios require non-binary variables? Let us study a simple and well-known example. A tower of a certain height casts a shadow. The length of the shadow causally depends on the height of the tower and the angle of the sun rays. These quantities may be represented by first-order functions. More precisely, they can be represented by the values of unary functions for certain arguments:

- a: tower
- b: sun rays
- c: shadow
- $h(a)$: height of the tower
- $n(b)$: angle between sun rays and surface of the earth
- $l(c)$: length of the shadow.

The values of these functions are governed by the following equation (in which *cot* stands for the trigonometric function of cotangent):

$$l(c) = cot(n(b)) \cdot h(a).$$

This equation can be read as a structural equation. For we think that the length of the shadow causally depends on the height of the tower and the angle of the sun rays, but not vice versa. So, let us take the equation as a structural equation in the technical sense of causal models. Also, let M be the causal model that contains only this equation.

On this reading, $l(c), n(b)$, and $h(a)$ are the variables of M. It is important to note that the variables of a causal model in a first-order language are not variables in the sense of first-order logic at all. The variables of a deterministic causal model are rather *ground terms* with occurrences of a function symbol. That is, they are terms (in the sense of first-order logic) which do not contain any variables (in the sense of first-order logic) but at least one function symbol. For clarification, we shall also speak of *causal variables* when referring to the variables of a causal model.

It is important to note that not all ground terms in a first-order theory about a causal scenario have a causal role. Take the ground terms for natural numbers. These terms do neither causally determine other variables nor are they causally determined. Likewise, we do not want to understand the constant symbols for tower, shadow, and sun rays as causal variables. Is the tower a cause of its height? This does not seem correct. Nor is it correct to say that all ground terms with an occurrence of a function symbol are causal variables. The value of 2 + 2, for example, is not causally determined, and so $2+2$ should not be considered a causal variable. Hence, choices are to be made as to which ground terms of a first-order theory are considered variables in the sense of the envisioned causal model. Causal modeling is an art (Halpern and Hitchcock 2010).

When working with a concrete causal model, we may want to say that a certain causal variable has a certain value. For propositional causal models, we can simply assert A if we want to say that the variable A has the Boolean value T. For non-binary causal variables, a direct statement about its value has the logical form $f(c) = c'$, where c and c' are individual constants. Note that the equation symbol in such a sentence does not have a causal meaning. If we say that the tower has a certain height, expressed by a rational number and a unit of length, we are thereby not implying that a rational number causally determines the height of the tower. Our logical account of causal models with non-binary variables must therefore distinguish between two equality symbols, one with a causal meaning and another without such a meaning. Let us adopt := for the equality symbol with a causal meaning. The above structural equation must then be rewritten as follows:

$$l(c) := cot(n(b)) \cdot h(a).$$

If the causal model contains non-causal mathematical equations, these need to be written with the standard equality symbol =, not with :=.

Each causal variable in a causal model has a well-defined range of values. We can take this into account by working with many-sorted first-order logic. The distinctive feature of this logic is that we have several domains of interpretation instead of a single domain. The different domains correspond to different *sorts*. Any constant symbol must be of a certain sort.

The formation of atomic formulas is constrained by sorts: if R is a predicate of type $\langle\sigma_1, \ldots, \sigma_n\rangle$, then $R(t_1, \ldots, t_n)$ is a formula iff, for all i $(1 \leq i \leq n)$, t_i is a term of sort σ_i. In what follows, let $D(\sigma_i)$ be the domain of interpretation of sort σ_i. The type of a function f is of the form $\langle\sigma_1, \ldots, \sigma_n\rangle \mapsto \sigma_j$. That is, such a function is a mapping of the set $D(\sigma_1) \times \ldots \times D(\sigma_n)$ onto the set $D(\sigma_j)$. Obviously, if f is of the form $\langle\sigma_1, \ldots, \sigma_n\rangle \mapsto \sigma_j$, then $f(t_1, \ldots, t_n)$ is a term iff, for all i $(1 \leq i \leq n)$, t_i is a term of sort σ_i. The term $f(t_1, \ldots, t_n)$ itself is of sort σ_j.

The semantics of many-sorted first-order logic generalizes the semantics of first-order logic in a straightforward manner. An interpretation of a many-sorted language must respect that each constant c is of a certain sort σ_i such that c is interpreted in the domain $D(\sigma_i)$. Likewise, for predicates, function symbols, and variables.[7]

Many-sorted first-order logic has been argued to be reducible to standard first-order logic. The precise meaning of this reduction is not obvious, though.[8] In any case, it seems obvious that many-sorted logic allows us to define the range of non-binary causal variables in a relatively straightforward manner. Suppose $f(t_1, \ldots t_n)$ is a ground term and a causal variable of a causal model. Let f be of the sort $\langle\sigma_1, \ldots, \sigma_n\rangle \mapsto \sigma_j$. Then, $D(\sigma_j)$ is the range of the causal variable $f(t_1, \ldots t_n)$.

We are now in a position to generalize our account of propositional causal models to causal models with non-binary variables. Let $\mathcal{L}$ be a many-sorted language of first-order logic. Further, let $\mathcal{V}$ be a set of ground terms of $\mathcal{L}$. The members of $\mathcal{V}$ are considered variables of the respective causal model. A structural equation is a sentence of the logical form

$$t := t'$$

where t and t' are ground terms. Moreover, $t \in \mathcal{V}$ and t' must have at least one occurrence of a ground term in $\mathcal{V}$. No other formulas are structural equations of $\mathcal{L}$ with the set $\mathcal{V}$ of causal variables. Note that a structural equation thus defined has no occurrences of quantifiers or first-order variables.

Definition 6. Causal Model (non-binary variables)
Let M be a set of structural equations, based on the language $\mathcal{L}$ with the set $\mathcal{V}$ of

[7] See Enderton (2001, Sect. 4.3) for a textbook account of many-sorted first-order logic.

[8] See Barrett and Halvorson (2017) for a detailed discussion.

causal variables. For an $\mathcal{L}$ term t, $Var(t)$ is the set of causal variables that occur in t. M is a causal model iff it satisfies two conditions:

(1) For any $t \in \mathcal{V}$, there is at most one $\sigma \in M$ such that σ has the logical form $t := t'$.

(2) If $t := t'$ is a member of M, then there is no t'' such that (i) $t' = t''$ on all interpretations of the language $\mathcal{L}$ and (ii) $Var(t'') \subset Var(t')$.

In brief, a causal model based on $\mathcal{L}$ is a set of structural equations such that any term of $\mathcal{V}$ occurs at most once on the left-hand side of an equation in M. Further, no structural equation in M must have vacuous occurrences of a causal variable.

Note that a structural equation in a causal model thus defined has no occurrences of quantifiers or first-order variables. Causal models with non-binary may or may not be embedded into full first-order reasoning. In this paper, we merely describe the core of causal reasoning with non-binary variables which is based on a fragment of many-sorted first-order logic. Inference rules for quantifiers do not belong to this fragment.

As for the elimination rule of :=, we adopt:

$$\frac{t := t' \quad t' = t''}{t = t''} \qquad [t \notin Var(\Gamma)] \qquad \text{(Elim :=)}$$

where Γ is the respective set of premises. This inference rule captures the interplay of reasoning about non-causal equality statements and structural equations. There is no need for an introduction rule of := for reasons explained in Section 5. $Var(\Gamma)$ is the set of causal variables of M that occur in at least one premise in Γ.

This is the only genuinely causal inference rule needed in our account of causal models with non-binary variables. All the other infer rules are adopted from classical logic. At the very least, we need the introduction and elimination rules for the non-causal equality symbol =. Once the set of classical inference rules is specified, the definition of $\Gamma \vdash_M \phi$ in Section 5 can be generalized to causal models with non-binary variables in a straightforward manner. For lack of space, we leave this to the reader.

It remains to specify the semantics of causal models with non-binary variables. Recall that we defined the semantics of propositional structural equations $A = \phi$ in terms of classical interpretations of a propositional language. Likewise, we can define the semantics of a structural equation $t := t'$ in terms of classical interpretations. Let $\mathcal{I}$ be a classical, model-theoretic interpretation of $\mathcal{L}$. Then, equation $t := t'$ is true on $\mathcal{I}$ iff t and t' designate the same object on the interpretation $\mathcal{I}$.

Note, furthermore, that the set M_Γ (used above in the definition of $\models_M$) remains well defined in the present setting of non-binary variables. M_Γ is simply the set of structural equations of M such that no causal variable of M occurs in any sentence of Γ. Since $\mathcal{V}$ is a set of ground terms, we can even represent an interpretation of M by a set V of literals in a manner analogous to the propositional case. Then, the set V_α – which represents the core of V that remains unchanged in an intervention by α – remains well defined too.

Since M_Γ remains well defined, we can easily generalize the definition of $\Gamma \models_M \phi$ in Section 6 to causal models with non-binary variables. We merely need to replace classical Boolean interpretations of $\mathcal{L}_P$ with model-theoretic interpretations of $\mathcal{L}$. Likewise, the proofs of soundness and completeness for causal reasoning with binary variables require only minor modifications to be generalized to the non-binary case. We leave this as an exercise to the reader.

11 Conclusion

The meaning of the equality symbol in a structural equation is nonsymmetric. The right-hand side of such an equation is thought to causally determine the left-hand side, but not vice versa. We have shown how this nonsymmetry can be captured by inference rules in a system of natural deduction. Thereby, we have furnished causal models with a distinct deductive system. This system has been supplemented by a model-theoretic semantics. Finally, we have defined interventionist conditionals in terms of inferential relations using the deduction system for causal models.

The present logic of causal models has been worked out completely for structural equations with binary variables. Then, we have shown how this logic can be taken to the level of causal models with non-binary variables. Accounts of explanation and causation in science may benefit from our logical investigation of causal models.

Appendix A Proofs

Theorem 1. **Soundness**
If $\Gamma \vdash_M \phi$, then $\Gamma \models_M \phi$.

Proof. Soundness can be proven by induction on the number of inferences in a derivation, as is standard. Suppose $\Gamma \vdash_M \phi$.

Induction basis: suppose the number of inferences is zero. Since $\phi \in \mathcal{L}_P$, this implies that $\phi \in \Gamma$. So, the derivation consists of a single formula which is a member of Γ. By the definition of $\models_M$, $\Gamma \models_M \phi$ iff, for all truth value interpretations v, if $v \models \Gamma \cup M_\Gamma$, then $v \models \phi$. Since $\phi \in \Gamma$, it holds that, if $v \models \Gamma \cup M_\Gamma$, then $v \models \phi$. Hence, $\Gamma \models_M \phi$.

Induction step. Suppose we have a derivation of n inference steps. Let Γ' be the union of Γ and the set of assumptions that are not in Γ and so far undischarged. By the induction hypothesis, we know that, for all so far derived sentences ψ, it holds that $\Gamma' \models_M \psi$. We need to show that, for any application of an inference rule of LC, if the next inference consists in inferring δ, then we have $\Gamma' \models_M \delta$. For the inference rules of the Boolean connectives (including $\bot$), this demonstration does not differ from the corresponding inductive step in the soundness proof for a natural deduction system of classical propositional logic. We can therefore focus on the inductive step for the inference rules $(= Elim_1)$ and $(= Elim_2)$.

Suppose the next inference step (following the n-th step in the derivation) has the form

$$\frac{A = \psi \qquad \psi}{A} \qquad\qquad [A \notin Var(\Gamma)].$$

We need to show that, for all valuations v, if $v \models \Gamma' \cup M_\Gamma$, then $v \models A$. Since there are no inference rules for the derivation of a formula of the type $A = \psi$, the equation $A = \psi$ is a member of M. Because of the condition $A \notin Var(\Gamma)$, it must even hold that $A = \psi$ is a member in M_Γ. By the induction hypothesis, (i) $\Gamma' \models_M \psi$. Suppose v is a valuation such that $v \models \Gamma' \cup M_\Gamma$. Since the equation $A = \psi$ is a member in M_Γ, this implies that $v \models A = \psi$. Hence, by the semantics of $=$, (ii) A and ψ have the same truth value on the valuation v. Further, we can infer from (i) that (iii) $v \models \psi$. So, ψ is true on v. Obviously, (ii) and (iii) imply that A is true on v. In symbols, $v \models A$. Thus, we have shown that, for all valuations v, if $v \models \Gamma' \cup M_\Gamma$, then $v \models A$. This concludes the inductive step for the inference rule $(= Elim_1)$. The demonstration of the inductive step for $(= Elim_2)$ is analogous.

Note finally that at the end of the derivation, when the last inference step has been completed, $\Gamma' = \Gamma$. All assumptions that are not in Γ must have been discharged. Therefore, by complete induction on the number of inference steps, $\Gamma \models_M \phi$. □

Theorem 2. **Completeness**
If $\Gamma \models_M \phi$, then $\Gamma \vdash_M \phi$.

Proof. Suppose $\Gamma \models_M \phi$. Hence, by definition of the entailment relation $\models_M$, for all valuations v such that $v \models \Gamma \cup M_\Gamma$, $v \models \phi$. Let M'_Γ be the set that we obtain from M_Γ by replacing every structural equation $A = \psi$ in M_Γ by the classical biconditional $A \leftrightarrow \psi$. Since the truth conditions of $=$ do not differ from the truth conditions of $\leftrightarrow$, $\Gamma \models_M \phi$ implies (i) $\Gamma \cup M'_\Gamma \models_{Cl} \phi$. Further, let M''_Γ be set the set that we obtain from M'_Γ by replacing every biconditional $A \leftrightarrow \psi$ in M'_Γ by $(A \vee \neg\psi) \wedge (\neg A \vee \psi)$. In symbols, $(A \vee \neg\psi) \vee (\neg A \vee \psi) \in M''_\Gamma$ iff $A \leftrightarrow \psi \in M'$. Since $(A \vee \neg\psi) \wedge (\neg A \vee \psi)$ and $A \leftrightarrow \psi$ are satisfied by the same classical valuations v, (i) implies (ii) $\Gamma \cup M''_\Gamma \models_{Cl} \phi$. By completeness of classical propositional logic, this implies that $\Gamma \cup M''_\Gamma \vdash_{Cl} \phi$. Since classical propositional logic with just the logical symbols $\vee$, $\wedge$, and $\neg$ is complete, $\Gamma \cup M''_\Gamma \vdash_{Cl} \psi$ holds, even if $\vdash_{Cl}$ is defined in terms of the inference rules of the Boolean connectives (including $\bot$). So, (iii) there is a derivation of ϕ from $\Gamma \cup M''_\Gamma$ using only the classical inference rules of the Boolean connectives (including $\bot$).

Now, we can show that $\Gamma \vdash_M (A \vee \neg\psi) \wedge (\neg A \vee \psi)$ for all $A = \psi \in M_\Gamma$. Let us first show that (iv) $\Gamma \vdash_M A \vee \neg\psi$ if $A = \psi \in M_\Gamma$. This can be done by the following derivation:

$$1\ \dfrac{\begin{array}{c}\vdots\\ \psi \vee \neg\psi\end{array} \qquad \dfrac{\dfrac{[\psi]^1 \qquad A = \psi}{A}\,(= \text{Elim}_1)}{A \vee \neg\psi}\,(\vee \text{ Intro}) \qquad \dfrac{[\neg\psi]^1}{A \vee \neg\psi}\,(\vee \text{ Intro})}{A \vee \neg\psi}\,(\vee \text{ Elim})$$

Note that $\psi \vee \neg\psi$ can be derived from the empty premise set using only inference rules of *LC*, which is indicated by the vertical dots on the left-hand side. Analogously, we can show (v) $\vdash_M \neg A \vee \psi$ if $A = \psi \in M$. From (iv) and (v) we can infer $\vdash_M (A \vee \neg\psi) \wedge (\neg A \vee \psi)$ if $A = \psi \in M$. Thus, we have shown that (vi) $\Gamma \vdash_M (A \vee \neg\psi) \wedge (\neg A \vee \psi)$ for all $A = \psi \in M_\Gamma$.

Recall that we have shown that (iii) there is a derivation of ϕ from $\Gamma \cup M''_\Gamma$ using only the inference rules of the Boolean connectives (including $\bot$). By (vi), we know that we can transform this derivation into a derivation of ϕ from Γ and M in the logic LC. The transformation goes as follows: instead of taking the sentences $(A \vee \neg\psi) \wedge (\neg A \vee \psi)$ in M''_{Γ} as premises, we derive these sentences from the structural equations in M using the inference rules of LC, as just demonstrated. Hence, $\Gamma \vdash_M \phi$. □

Proposition 1. Let $T = (\mathcal{S}, \mathcal{F})$ be a causal model over signature $\mathcal{S}$, where $\mathcal{F}$ is a set of structural equations, as defined in Halpern (2000). Let all variables of the causal model be binary ones. $\vec{u}$ is a vector that represents an assignment to the exogenous variables. Let M be a set of syntactic structural equations such that each equation has a unique semantic representation by a member in $\mathcal{F}$. V is a set of literals that syntactically represents the assignment given by vector $\vec{u}$. α is a conjunction of literals such that this conjunction represents the assignment $\vec{Y} \leftarrow \vec{y}$. γ is an arbitrary Boolean formula. Then, if $T \models [\vec{Y} \leftarrow \vec{y}]\gamma$, then $\langle M, V \rangle \models [\alpha]\gamma$.

Proof. Suppose $T \models [\vec{Y} \leftarrow \vec{y}]\gamma$. Since $\vec{u}$ is a vector assigned to exogenous variables and since V translates this assignment, all propositional variables that occur in the literals in V are root variables. (A root variable is one that does not have ancestors.) Now, if we set Y_i to y_i ($i = 1, \ldots, k$), this means replacing the structural equations for the variables in $\vec{Y}$ by certain value assignments. The other structural equations of the original causal model remain in place. Let M' be the set of structural equations $X = \phi$ such that X is not an element in $\vec{Y}$, which means that there is no intervention on X. It holds that $T \models [\vec{Y} \leftarrow \vec{y}]\gamma$ iff all solutions of the submodel of T after the intervention by $\vec{Y} \leftarrow \vec{y}$ in the context $\vec{u}$ satisfy γ. Since we assume $T \models [\vec{Y} \leftarrow \vec{y}]\gamma$, we can infer from this that (i) all solutions of the submodel of T after the intervention by $\vec{Y} \leftarrow \vec{y}$ in the context $\vec{u}$ satisfy γ.

Since all members of V concern exogenous variables and since all occurrences of a variable in α are of the endogenous type, it holds that all variables that occur in a literal in V are non-descendants of all variables that occur in α. Hence, (ii) $V_\alpha = V$. Since all members of V concern exogenous variables (for which there is no structural equation in $\mathcal{F}$) and since α represents the intervention $\vec{Y} \leftarrow \vec{y}$, (iii) $M_{\Gamma'} = M'$ for $\Gamma' = V_\alpha \cup \{\alpha\}$. Because of (ii) and (iii), (i) translates – in the propositional setting of our syntactic account – to the claim that all valuations of P that satisfy $M' \cup V_\alpha \cup \{\alpha\}$ are such that they verify γ. Using definitions 3 and 5, we can infer from this that $\langle M, V \rangle \models [\alpha]\gamma$. □

References

Andreas, H. and Günther, M. (2021). Difference-Making Causation. *The Journal of Philosophy* **118**(12): 680–701.

Barrett, T. W. and Halvorson, H. (2017). Quine's Conjecture on Many-Sorted Logic. *Synthese* **194**(9): 3563–3582.

Briggs, R. (2012). Interventionist Counterfactuals. *Philosophical Studies* **160**(1): 139–166.

Carnap, R. (1947). *Meaning and Necessity: A Study in Semantics and Modal Logic*. Chicago: University of Chicago Press.

Enderton, H. (2001). *A Mathematical Introduction to Logic*. San Diego: Harcourt Academic Press.

Halmos, P. R. (1974). *Naive Set Theory*. New York: Springer.

Halpern, J. Y. (2000). Axiomatizing Causal Reasoning. *Journal of Artificial Intelligence Research* **12**(1): 317–337.

Halpern, J. Y. (2013). From Causal Models to Counterfactual Structures. *Review of Symbolic Logic* **6**(2): 305–322.

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

Halpern, J. Y. and Hitchcock, C. (2010). Actual Causation and the Art of Modeling. In *Heuristics, Probability, and Causality: a Tribute to Judea Pearl*, edited by R. Dechter, H. Geffner, and J. Y. Halpern, London: College Publications. 383–406.

Halpern, J. Y. and Pearl, J. (2005). Causes and Explanations: A Structural-Model Approach. Part I: Causes. *British Journal for the Philosophy of Science* **56**(4): 843–887.

Hintikka, J. (1955). *Two Papers on Symbolic Logic Form and Content in Quantification Theory and Reductions in the Theory of Types*. Acta Philosophica Fennica, Helsinki: Edidit Societas Philosophica.

Huber, F. (2013). Structural Equations and Beyond. *Review of Symbolic Logic* **6**(4): 709–732.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. New York, NY, USA: Cambridge University Press, 2nd edn.

Schulz, K. (2011). “If you’d wiggled A, then B would’ve changed”: Causality and counterfactual conditionals. *Synthese* **179**(2): 239–251.

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