Journal of Artificial Intelligence Research 12 (2000) 317–337 Submitted 6/99; published 5/00

Axiomatizing Causal Reasoning

Joseph Y. Halpern HALPERN@CS.CORNELL.EDU
Cornell University, Computer Science Department
Ithaca, NY 14853
http://www.cs.cornell.edu/home/halpern

Abstract

Causal models defined in terms of a collection of equations, as defined by Pearl, are axiomatized here. Axiomatizations are provided for three successively more general classes of causal models: (1) the class of recursive theories (those without feedback), (2) the class of theories where the solutions to the equations are unique, (3) arbitrary theories (where the equations may not have solutions and, if they do, they are not necessarily unique). It is shown that to reason about causality in the most general third class, we must extend the language used by Galles and Pearl (1997, 1998). In addition, the complexity of the decision procedures is characterized for all the languages and classes of models considered.

1. Introduction

The important role of causal reasoning—in prediction, explanation, and counterfactual reasoning—has been argued eloquently in a number of recent papers and books (Chajewska & Halpern, 1997; Heckerman & Shachter, 1995; Henrion & Druzdzel, 1990; Druzdzel & Simon, 1993; Pearl, 1995; Pearl & Verma, 1991; Spirtes, Glymour, & Scheines, 1993). If we are to reason about causality, then it is certainly useful to find axioms that characterize such reasoning. The way we go about axiomatizing causal reasoning depends on two critical factors:

- how we model causality, and
- the language that we use to reason about it.

In this paper, I consider one approach to modeling causality, using *structural equations*. The use of structural equations as a model for causality is standard in the social sciences, and seems to go back to the work of Sewall Wright in the 1920s (see (Goldberger, 1972) for a discussion); the particular framework that I use here is due to Pearl (1995). Galles and Pearl (1997) introduce some axioms for causal reasoning in this framework; they also provide a complete axiomatic characterization of reasoning about causality in this framework, under the strong assumption that there is a fixed, given *causal ordering* $\prec$ of the equations (Galles & Pearl, 1998). Roughly speaking, this means there is a way of ordering the variables that appear in the equations and we have explicit axioms that say X_j has no influence of X_i if $X_i \prec X_j$ in this causal ordering.

In this paper, I extend the results of Galles and Pearl by providing a complete axiomatic characterization for three increasingly general classes of causal models (defined by structural equations):

1. the class of recursive theories (those without feedback—this generalizes the situation considered by Galles and Pearl (1998), since every fixed causal ordering of the variables gives rise to a recursive theory),
2. the class of theories where the solutions to the equations are unique,
3. arbitrary theories (where the equations may not have solutions and, if they do, they are not necessarily unique).

In the process, I clarify some problems in the Galles-Pearl completeness proof that arise from the lack of propositional connectives (particularly disjunction) in the language they consider and, more generally, highlight the role of the language in reasoning about causality. I also characterize the complexity of the decision problem for all these languages and classes of models.

The rest of the paper is organized as follows. In Section 2, I give the syntax and semantics of the languages I will be considering and review the definition of modifiable causal models. In Section 3, I present the complete axiomatizations. In Section 4 I consider the complexity of the decision procedure. I conclude in Section 5.

2. Syntax and Semantics

An axiomatization is given with respect to a particular language and a class of models, so we need to make both precise. Both the language and models I use are based on those considered by Galles and Pearl (1997, 1998). To make comparisons easier, I use their notation as much as possible. I start with the semantic model, since it motivates some of the choices in the syntax, then give the syntax, and finally define the semantics of formulas.

2.1 Causal Models

The basic picture here is that we are interested in the values of random variables, some of which have a causal effect on others. This effect is modeled by a set of *structural equations*.

In practice, it seems useful to split the random variables into two sets, the *exogenous* variables, whose values are determined by factors outside the model, and the *endogenous* variables. It is these endogenous variables whose values are described by the structural equations.

More formally, a *signature* $\mathcal{S}$ is a tuple $(\mathcal{U}, \mathcal{V}, \mathcal{R}\}$, where $\mathcal{U}$ is a finite set of exogenous variables, $\mathcal{V}$ is a finite set of endogenous variables, and $\mathcal{R}$ associates with every variable $Y \in \mathcal{U} \cup \mathcal{V}$ a nonempty set $\mathcal{R}(Y)$ of possible values for Y (the *range* of possible values of Y). Unless explicitly noted otherwise, I assume that $\mathcal{R}(Y)$ is a *finite* set for each $Y \in \mathcal{U} \cup \mathcal{V}$ and $|\mathcal{R}(Y)| \geq 2$. The assumption that $\mathcal{U}$ and $\mathcal{V}$ are finite is relatively innocuous; as we shall see, the assumption that $\mathcal{R}(Y)$ is finite has more of an impact on the axioms. The assumption that $|\mathcal{R}(Y)| \geq 2$ allows us to ignore the trivial situation where $|\mathcal{R}(Y)| = 1$. If $|\mathcal{R}(Y)| = 1$, we can just remove the variable Y from the signature without loss of expressiveness.

A *causal model* over signature $\mathcal{S}$ is a tuple $T = (\mathcal{S}, \mathcal{F})$ where $\mathcal{F}$ associates with each variable $X \in \mathcal{V}$ a function denoted F_X such that $F_X : (\times_{U \in \mathcal{U}} \mathcal{R}(U)) \times (\times_{Y \in \mathcal{V} - \{X\}} \mathcal{R}(Y)) \rightarrow \mathcal{R}(X)$. 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. Because F_X is a function, there is a unique value of X once we have set all the other variables. Notice we have such functions only for the endogenous variables. The exogenous variables are taken as given; it is their effect on the endogenous variables (and the effect of the endogenous variables on each other) that we are modeling with the structural equations.

Given a causal model $T = (\mathcal{S}, \mathcal{F})$ over signature $\mathcal{S}$, a (possibly empty) vector $\vec{X}$ of variables in $\mathcal{V}$, and vectors $\vec{x}$ and $\vec{u}$ of values for the variables in $\vec{X}$ and $\mathcal{U}$, respectively, we can define a new causal model denoted $T_{\vec{X}\leftarrow\vec{x}}(\vec{u})$ over the signature $\mathcal{S}_{\vec{X}} = (\emptyset, \mathcal{V}-\vec{X}, \mathcal{R}|_{\mathcal{V}-\vec{X}})$.[1] Intuitively, this is the causal model that results when the variables in $\vec{X}$ are set to $\vec{x}$ and the variables in $\mathcal{U}$ are set to $\vec{u}$. Formally, $T_{\vec{X}\leftarrow\vec{x}}(\vec{u}) = (\mathcal{S}_{\vec{X}}, \mathcal{F}^{\vec{X}\leftarrow\vec{x},\vec{u}}\})$, where $F_Y^{\vec{X}\leftarrow\vec{x},\vec{u}}$ is obtained from F_Y by setting the values of the variables in $\vec{X}$ to $\vec{x}$ and the values of the variables in $\mathcal{U}$ to $\vec{u}$. The causal model $T_{\vec{X}\leftarrow\vec{x}}(\vec{u})$ is called a *submodel* of T by Pearl (1999). It can describe a possible *counterfactual* situation; that is, even though, under normal circumstances, setting the exogenous variables to $\vec{u}$ may result in the variables $\vec{X}$ having values $\vec{x}' \neq \vec{x}$, this submodel describes what happens if they are set to $\vec{x}$ due to some "external action", the cause of which is not modeled explicitly. For example, to determine if the manufacturer is at fault in an accident that involved a poorly maintained car, we may want to consider what would have happened had the car been well maintained. If there is a random variable in the signature that describes how well maintained the car is, then this means examining the submodel where that random variable is set to 1 (the car is well maintained). It is this ability to examine counterfactual situations that makes causal structures a useful tool for reasoning about causality.

Notice that, in general, there may not be a unique vector of values that simultaneously satisfies the equations in $T_{\vec{X}\leftarrow\vec{x}}(\vec{u})$; indeed, there may not be a solution at all. One special case where there is guaranteed to be such a unique solution is if there is some total ordering $\prec$ of the variables in $\mathcal{V}$ such that if $X \prec Y$, then F_X is independent of the value of Y; i.e., $F_X(\ldots, y, \ldots) = F_X(\ldots, y', \ldots)$ for all $y, y' \in \mathcal{R}(Y)$. In this case, the causal model is said to be *recursive* or *acyclic*. Intuitively, if the theory is recursive, there is no feedback. If $X \prec Y$, then the value of X may affect the value of Y, but the value of Y has no effect on the value of X.

It should be clear that if T is a recursive theory, then there is always a unique solution to the equations in $T_{\vec{X}\leftarrow\vec{x}}(\vec{u})$, for all $\vec{X}$, $\vec{x}$, and $\vec{u}$. (We simply solve for the variables in the order given by $\prec$.) On the other hand, as the following example shows, it is not hard to construct nonrecursive theories for which there is always a unique solution to the equations that arise.

Example 2.1: Let $\mathcal{S} = (\emptyset, \{X, Y\}, \mathcal{R}\})$, where $\mathcal{R}(X) = \mathcal{R}(Y) = \{-1, 0, 1\}$, and let $T = (\mathcal{S}, \mathcal{F})$, where F_X is characterized by the equation $X = Y$ and F_Y is characterized by the equation $Y = -X$ (that is, $F_X(y) = y$ and $F_Y(x) = -x$). Clearly T is not recursive; the value of X depends on the value of Y and the value of Y depends on that of X. Nevertheless, it is easy to see that T has the unique solution $X = 0, Y = 0$, $T_{X\leftarrow x}$ has the unique solution $Y = -x$, and $T_{Y\leftarrow y}$ has the unique solution $X = y$. ∎

1. I am implicitly identifying the vector $\vec{X}$ with the subset of $\mathcal{V}$ consisting of the variables in $\vec{X}$. I do this throughout the paper. $\mathcal{R}|_{\mathcal{V}-\vec{X}}$ is the restriction of $\mathcal{R}$ to the variables in $\mathcal{V} - \vec{X}$.

In this paper, I consider three successively more general classes of causal models for a given signature $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R})$.

1. $\mathcal{T}_{\rm rec}(\mathcal{S})$: the class of recursive causal models over signature $\mathcal{S}$,
2. $\mathcal{T}_{\rm uniq}(\mathcal{S})$: the class of causal models T over $\mathcal{S}$ where for all $\vec{X} \subseteq \mathcal{V}$, $\vec{x}$, and $\vec{u}$, the equations in $T_{\vec{X}\leftarrow\vec{x}}(\vec{u})$ have a unique solution,
3. $\mathcal{T}(\mathcal{S})$: the class of all causal models over $\mathcal{S}$.

I often omit the signature $\mathcal{S}$ when it is clear from context or irrelevant, but the reader should bear in mind its important role.

Why should we be interested in causal models that do not possess unique solutions? Are there real causal systems that do not possess unique solutions? The issue of whether nonrecursive system can be given a causal interpretation is discussed at some length by Strotz and Wold (1960). They argue that there are reasonable ways of interpreting causal interpretations where the answer is yes. Under these interpretations, there may well be more than one solution to the equations. Perhaps the best way to view such equations is to think of the variables in $\mathcal{V}$ as being mutually interdependent; changing any one of them may cause a change in the others. (Think of demand and supply in economics or populations of rabbits and wolves.) The solutions to the equations then represent equilibrium situations. If there is more than one equilibrium, there will be more than one solution to the equations. Of course, if there are no equilibria, then there will be no solutions to the equations.

A related way of thinking about these equations is that they represent atemporal versions of temporal causal equations. That is, suppose that we replace every variable $Y \in \mathcal{U} \cup \mathcal{V}$ by a family of variables $Y_0, Y_1, Y_2, \ldots$, where, intuitively, Y_t represents the value of Y at time t. Each equation $f_X \in \mathcal{F}$ is then replaced by a family of equations f_{X_t}, where f_{X_t} depends only on exogenous variables $U_{t'}$ with $t' \le t$ and endogenous variables $Y_{t'}$ with $t' < t$. This gives us a recursive system. The values of X_t under some setting of the variables with subscript 0 represents the evolution of X under that setting of the variables. If X_t eventually stabilizes, then we might expect the equilibrium value to be the value of X in some solution to the original set of equations. If X_t stabilizes, then there would not be a solution to the original set of equations.

2.2 Syntax

I focus here on two languages. Both languages are parameterized by a signature $\mathcal{S}$. The first language, $\mathcal{L}^+(\mathcal{S})$, borrows ideas from dynamic logic (Harel, 1979). Again, I often write $\mathcal{L}^+$ rather than $\mathcal{L}^+(\mathcal{S})$ (and similarly for the other languages defined below) to simplify the notation. A *basic causal formula* is one of the form $[Y_1 \leftarrow y_1, \ldots, Y_k \leftarrow y_k]\varphi$, where φ is a Boolean combination of formulas of the form $X(\vec{u}) = x$, $Y_1, \ldots, Y_k, X$ are variables in $\mathcal{V}$, $Y_1, \ldots, Y_k$ are distinct, $x \in \mathcal{R}(X)$, and $\vec{u}$ is a vector of values for all the variables in $\mathcal{U}$. I typically abbreviate such a formula as $[\vec{Y} \leftarrow \vec{y}]\varphi$. The special case where $k = 0$ (which is allowed) is abbreviated as $[true]\varphi$. $[\vec{Y} \leftarrow \vec{y}]X(\vec{u}) = x$ 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". As we shall see, this formula is true in a causal model T if in all solutions to the equations in $T_{\vec{Y}\leftarrow\vec{y}}(\vec{u})$, the random variable X has value x. Note that this formula is trivially true if there are no solutions to the structural equations. A *causal formula* is a Boolean combination of basic causal formulas.

Just as with dynamic logic, we can also define the formula $\langle\vec{Y} \leftarrow \vec{y}\rangle(X(\vec{u}) = x)$ to be an abbreviation of $\neg[\vec{Y} \leftarrow \vec{y}]\neg(X(\vec{u}) = x)$. $\langle\vec{Y} \leftarrow \vec{y}\rangle(X(\vec{u}) = x)$ is the dual of $[\vec{Y} \leftarrow \vec{y}](X(\vec{u}) = x)$; it is true if, in some solution 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. Taking $true(\vec{u})$ to be an abbreviation for $X(\vec{u}) = x \vee X(\vec{u}) \neq x$ for some variable X and $x \in \mathcal{R}(X)$, and taking $false(\vec{u})$ to be an abbreviation for $\neg true(\vec{u})$, we have that $\langle\vec{Y} \leftarrow \vec{y}\rangle true(\vec{u})$ is true if there is some solution to the equations obtained by setting Y_i to y_i, $i = 1, \ldots, k$, and the variables in $\mathcal{U}$ to $\vec{u}$ (since $[\vec{Y} \leftarrow \vec{y}]false(\vec{u})$ says that in every solution to the equations obtained by setting Y_i to y_i and $\mathcal{U}$ to $\vec{u}$, the formula $false(\vec{u})$ is true, and thus holds exactly if the equations have no solution).

Let $\mathcal{L}_{\text{uniq}}(\mathcal{S})$ be the sublanguage of $\mathcal{L}^+(\mathcal{S})$ which consists of Boolean combinations of formulas of the form $[\vec{Y} \leftarrow \vec{y}]X(\vec{u}) = x$. Thus, the difference between $\mathcal{L}_{\text{uniq}}$ and $\mathcal{L}^+$ is that in $\mathcal{L}_{\text{uniq}}$, only $X(\vec{u}) = x$ is allowed after $[\vec{Y} \leftarrow \vec{y}]$, while in $\mathcal{L}^+$, arbitrary Boolean combinations of formulas of the form $X(\vec{u}) = x$ are allowed. As we shall see, for reasoning about causality in $\mathcal{T}_{\text{uniq}}$, the language $\mathcal{L}_{\text{uniq}}$ is adequate, since it is equivalent in expressive power to $\mathcal{L}^+$. However, this is no longer the case when reasoning about causality in $\mathcal{T}$.

Following Galles and Pearl's notation, I often write $[\vec{Y} \leftarrow \vec{y}]X(\vec{u}) = x$ as $X_{\vec{Y}\leftarrow\vec{y}}(\vec{u}) = x$. If $\vec{Y}$ is clear from context or irrelevant, I further abbreviate this as $X_{\vec{y}}(\vec{u}) = x$. (This is actually the notation used by Galles and Pearl.) Let $\mathcal{L}_{\text{GP}}(\mathcal{S})$ be the sublanguage of $\mathcal{L}_{\text{uniq}}(\mathcal{S})$ consisting of just conjunctions of formulas of the form $X_{\vec{y}}(\vec{u}) = x$. In particular, it does not contain disjunctions or negations of such formulas. Although Galles and Pearl (1998) are not explicit about the language they are using, it seems to be $\mathcal{L}_{\text{GP}}$.[2]

2.3 Semantics

A formula in $\mathcal{L}^+(\mathcal{S})$ is true or false in a causal model in $\mathcal{T}(\mathcal{S})$. As usual, we write $T \models \varphi$ if the causal formula φ is true in causal model T. For a basic causal formula, we have $T \models [\vec{Y} \leftarrow \vec{y}](X(\vec{u}) = x)$ if in all solutions to $T_{\vec{Y}\leftarrow\vec{y}}(\vec{u})$ (i.e., in all vectors of values for the variables in $\mathcal{V} - \vec{Y}$ that simultaneously satisfy all the equations $F_Z^{\vec{Y}\leftarrow\vec{y}}$, for $Z \in \mathcal{V} - \vec{Y}$), the variable X has value x. We define the truth value of arbitrary causal formulas, which are just Boolean combinations of basic causal formulas, in the obvious way:

- $T \models \varphi_1 \wedge \varphi_2$ if $T \models \varphi_1$ and $T \models \varphi_2$
- $T \models \neg\varphi$ if $T \not\models \varphi$.

As usual, we say that a formula φ is *valid* with respect to a class $\mathcal{T}'$ of causal models if $T \models \varphi$ for all $T \in \mathcal{T}'$.

I can now make precise the earlier claim that in $\mathcal{T}_{\text{uniq}}$ (and hence $\mathcal{T}_{\text{rec}}$), the language $\mathcal{L}_{\text{uniq}}$ is just as expressive as the full language $\mathcal{L}^+$.

Lemma 2.2: *The following formulas are valid in $\mathcal{T}_{\text{uniq}}$:*

2. This was confirmed by Judea Pearl [private communication, 1997].

(a) $\mathcal{T}_{\text{uniq}} \models [\vec{Y} \leftarrow \vec{y}](\varphi \vee \psi) \Leftrightarrow [\vec{Y} \leftarrow \vec{y}]\varphi \vee [\vec{Y} \leftarrow \vec{y}]\psi$*,*

(b) $\mathcal{T}_{\text{uniq}} \models [\vec{Y} \leftarrow \vec{y}](\varphi \wedge \psi) \Leftrightarrow [\vec{Y} \leftarrow \vec{y}]\varphi \wedge [\vec{Y} \leftarrow \vec{y}]\psi$*,*

(c) $\mathcal{T}_{\text{uniq}} \models [\vec{Y} \leftarrow \vec{y}]\neg\varphi \Leftrightarrow \neg[\vec{Y} \leftarrow \vec{y}]\varphi$*.*

Hence, in $\mathcal{T}_{\text{uniq}}$*, every formula in* $\mathcal{L}^+$ *is equivalent to a formula in* $\mathcal{L}_{\text{uniq}}$*.*

Proof: Straightforward; left to the reader. ∎

Note that it follows from these equivalences that in $\mathcal{T}_{\text{uniq}}$, $[\vec{Y} \leftarrow \vec{y}]\varphi$ is equivalent to $\langle\vec{Y} \leftarrow \vec{y}\rangle\varphi$. It is also worth noting that Lemma 2.2(b) holds in arbitrary causal models in $\mathcal{T}$, not just in $\mathcal{T}_{\text{uniq}}$. However, parts (a) and (c) do not, as the following example shows.

Example 2.3: Let $\mathcal{S} = (\emptyset, \{X, Y\}, \mathcal{R})$, where $\mathcal{R}(X) = \mathcal{R}(Y) = \{0,1\}$; let $T = (\mathcal{S}, \mathcal{F})$, where F_X is characterized by the equation $X = Y$ and F_Y is characterized by the equation $Y = X$. Clearly $T \notin \mathcal{T}_{\text{uniq}}$; both $(0,0)$ and $(1,1)$ are solutions to T. It is easy to see that $T \models [\mathit{true}](X = 0 \vee X = 1) \wedge \neg[\mathit{true}](X = 0) \wedge \neg[\mathit{true}](X = 1)$, showing that part (a) of Lemma 2.2 does not hold in T, and that $T \models \neg[\mathit{true}](X = 1) \wedge \neg[\mathit{true}]\neg(X = 1)$, showing that part (c) does not hold either. ∎

3. Complete Axiomatizations

I briefly recall some standard definitions from logic. An *axiom system* AX consists of a collection of *axioms* and *inference rules*. An axiom is a formula (in some predetermined language $\mathcal{L}$), and an inference rule has the form "from $\varphi_1, \ldots, \varphi_k$ infer ψ," where $\varphi_1, \ldots, \varphi_k, \psi$ are formulas in $\mathcal{L}$. A *proof* in AX consists of a sequence of formulas in $\mathcal{L}$, each of which is either an axiom in AX or follows by an application of an inference rule. A proof is said to be a *proof of the formula* φ if the last formula in the proof is φ. We say φ is *provable in AX*, and write $\text{AX} \vdash \varphi$, if there is a proof of φ in AX; similarly, we say that φ is *consistent with AX* if $\neg\varphi$ is not provable in AX.

An axiom system AX is said to be *sound* for a language $\mathcal{L}$ with respect to a class $\mathcal{T}'$ of causal models if every formula in $\mathcal{L}$ provable in AX is valid with respect to $\mathcal{T}'$. AX is *complete* for $\mathcal{L}$ with respect to $\mathcal{T}'$ if every formula in $\mathcal{L}$ that is valid with respect to $\mathcal{T}'$ is provable in AX.

We now want to find axioms that characterize the classes of causal models in which we are interested, namely $\mathcal{T}_{\text{rec}}$, $\mathcal{T}_{\text{uniq}}$, and $\mathcal{T}$. To deal with $\mathcal{T}_{\text{rec}}$, it is helpful to define $Y \rightsquigarrow Z$, read "Y affects Z", as an abbreviation for the formula

$$\vee_{\vec{X} \subseteq \mathcal{V}, \vec{x} \in \times_{X \in \mathcal{V}} \mathcal{R}(X), y \in \mathcal{R}(y), \vec{u} \in \times_{U \in \mathcal{U}} \mathcal{R}(U), z \neq z' \in \mathcal{R}(Z)} (Z_{\vec{x}y}(\vec{u}) = z' \wedge Z_{\vec{x}}(\vec{u}) = z).$$

Thus, Y affects Z if there is some setting of the exogenous variables and some other endogenous variables for which changing the value of Y changes the value of Z. This definition is used in axiom C6 below, which characterizes recursiveness.

Consider the following axioms:

C0. All instances of propositional tautologies.

C1. $X_{\vec{y}}(\vec{u}) = x \Rightarrow X_{\vec{y}}(\vec{u}) \neq x'$ if $x, x' \in \mathcal{R}(X), x \neq x'$ (equality)

C2. $\vee_{x\in\mathcal{R}(X)} X_{\vec{y}}(\vec{u}) = x$ (definiteness)

C3. $(W_{\vec{x}}(\vec{u}) = w \wedge Y_{\vec{x}}(\vec{u}) = y) \Rightarrow Y_{\vec{x}w}(\vec{u}) = y$ (composition)

C4. $X_{x\vec{w}}(\vec{u}) = x$ (effectiveness)

C5. $(Y_{\vec{x}w}(\vec{u}) = y \wedge W_{\vec{x}y}(\vec{u}) = w) \Rightarrow Y_{\vec{x}}(\vec{u}) = y$ (reversibility)

C6. $(X_0 \rightsquigarrow X_1 \wedge \ldots \wedge X_{k-1} \rightsquigarrow X_k) \Rightarrow \neg(X_k \rightsquigarrow X_0)$ (recursiveness)

We have one rule of inference:

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

C1 just states an obvious property of equality: if $X = x$ for every solution of the equations in $T_{\vec{x}}(\vec{u})$, then we cannot have $X = x'$, if $x' \neq x$.[3] In a richer language, this could have been expressed as $(X_{\vec{y}}(\vec{u}) = x \wedge X_{\vec{y}}(\vec{u}) = x') \Rightarrow (x = x')$, but this formula is not in $\mathcal{L}^+$ (since $\mathcal{L}^+$ does not include expressions such as $x' = x$). C2 states that there is some value $x \in \mathcal{R}(X)$ which is the value of X in all solutions to the equations in $T_{\vec{x}}(\vec{u})$. C2 is not valid in $\mathcal{T}$, but it is valid in $\mathcal{T}_{\text{uniq}}$. Note that in stating C2, I am making use of the fact that $\mathcal{R}(X)$ is finite (otherwise C2 would involve an infinite disjunction, and would no longer be a formula in $\mathcal{L}_{\text{uniq}}$). In fact, it can be shown that if we allow signatures where the sets $\mathcal{R}(X)$ are infinite, we include C2 only for those random variables X such that $\mathcal{R}(X)$ is finite.[4] C3–C5 were introduced by Galles and Pearl (1997, 1998), as were their names. Roughly speaking, C3 says that if the value of W is w in all solutions to the equations $T_{\vec{x}}(\vec{u})$, then all solutions to the equations in $T_{\vec{x}w}(\vec{u})$ are the same as the solutions to the equations in $T_{\vec{x}}(\vec{u})$. C3 is valid in $\mathcal{T}$ as well as $\mathcal{T}_{\text{uniq}}$. As we shall see, a variant of C3 (obtained by replacing "all" by "some") is also valid in $\mathcal{T}$. C4 simply says that in all solutions obtained after setting X to x, the value of X is x. C5 is perhaps the least obvious of these axioms; the proof of its soundness is not at all straightforward. It says that if setting $\vec{X}$ to $\vec{x}$ and W to w results in Y having value y and setting $\vec{X}$ to $\vec{x}$ and Y to y results in W having value w, then Y must already have value when we set $\vec{X}$ to x (and W must already have value w).

Finally, it is easy to see that C6 holds in recursive models. For if $Y \rightsquigarrow Z$, then Y must precede Z in the causal ordering. Thus, if $X_0 \rightsquigarrow X_1 \wedge \ldots \wedge X_{k-1} \rightsquigarrow X_k$, then X_0 must precede X_k in the causal ordering, so X_k cannot affect X_0. Thus, $\neg(X_k \rightsquigarrow X_0)$ holds. As we shall see, in a precise sense, C6 characterizes recursive models.

C6 can be viewed as a collection of axioms (actually, axiom schemes), one for each k. The case $k = 1$ already gives us $\neg(Y \rightsquigarrow Z) \vee \neg(Z \rightsquigarrow Y)$ for all variables Y and Z. That

3. In an earlier draft of this paper, where C1 and C2 were introduced, C1 was called "uniqueness". Galles and Pearl (1998) then adopted this name as well. In retrospect, this axiom really does not say anything about uniqueness. The axiom which does is D10, which will be discussed later.

4. The assumption that $\mathcal{R}(X)$ and $\mathcal{V}$ are finite is also necessary for the abbreviation $X \rightsquigarrow Y$ used in C6 to be in $\mathcal{L}_{\text{uniq}}$; however, we can replace C6 by the axiom scheme

$$\neg(\wedge_{i=0}^{k-1} (X_{i+1})_{\vec{y}_i x_i}(\vec{u}_i) = z_i \wedge (X_{i+1})_{\vec{y}_i} = z_i') \wedge (X_0)_{\vec{y}_k x_k}(\vec{u}_k) = z_k \wedge (X_0)_{\vec{y}_k} = z_k'),$$

where $x_i \in \mathcal{R}(X_i)$ for $i = 1, \ldots, k$. That is, we essentially replace C6 by all its instances. This axiom is equivalent to C6 (although not as transparent) and can be expressed even if $|\mathcal{V}|$ is infinite or $|\mathcal{R}(X)|$ is infinite for some $X \in \mathcal{V}$.

is, it tells us that, for any pair of variables, at most one affects the other. However, just restricting C6 to the case of $k = 1$ does not suffice to characterize $\mathcal{T}_{\text{rec}}$, as the following example shows.

Example 3.1: Let $\mathcal{S} = (\emptyset, \{X_0, X_1, X_2\}, \mathcal{R})$, where $\mathcal{R}(X_0) = \mathcal{R}(X_1) = \mathcal{R}(X_2) = \{0, 1, 2\}$, and let $T = (\mathcal{S}, \mathcal{F})$, where F_{X_i} is characterized by the equation

$$X_i = \begin{cases} 2 & \text{if } X_{i\oplus 1} = 1 \\ 0 & \text{otherwise} \end{cases}$$

and $\oplus$ is addition mod 3. It is easy to see that $T \in \mathcal{T}_{\text{uniq}}$: If any of the variables are set, the equations completely determine the values of all the other variables. On the other hand, if none of the variables are set, it is easy to see that $(0, 0, 0)$ is the only solution that satisfies all the equations. Moreover, in $T_{\vec{X} \leftarrow \vec{x}}$, the variable X_i is 0 unless it is set to a value other than 0 or $X_{i\oplus 1}$ is set to 1. It easily follows that X_i is affected only by $X_{i\oplus 1}$. A straightforward verification (or an appeal to Theorem 3.2 below) shows that T satisfies all the axioms other than C6. C6 does not hold in T, since $T \models X_0 \rightsquigarrow X_1 \wedge X_1 \rightsquigarrow X_2 \wedge X_2 \rightsquigarrow X_0$. This also shows that T is not recursive. However, the restricted version of C6 (where $k = 1$) does hold in T. A generalization of this example (with k random variables rather than just 2) can be used to show that we cannot bound k at all in C6; we need C6 to hold for all finite values of k. ∎

Let $\text{AX}_{\text{uniq}}(\mathcal{S})$ consist of C0–C5 and MP; let $\text{AX}_{\text{rec}}(\mathcal{S})$ consist of C0–C4, C6, and MP. We could include C5 in $\text{AX}_{\text{rec}}(\mathcal{S})$; I did not do so because, as Galles and Pearl (1998) point out, it follows from C3 and C6. Note that the signature $\mathcal{S}$ is a parameter of the axiom system, just as it is for the language and the set of models. This is because, for example, the set $\mathcal{R}(X)$ (which is determined by $\mathcal{S}$) appears explicitly in C1 and C2.

Theorem 3.2: *$AX_{\text{uniq}}(\mathcal{S})$ (resp., $AX_{\text{rec}}(\mathcal{S})$) is a sound and complete axiomatization for $\mathcal{L}_{\text{uniq}}(\mathcal{S})$ with respect to $\mathcal{T}_{\text{uniq}}(\mathcal{S})$ (resp., $\mathcal{T}_{\text{rec}}(\mathcal{S})$).*

Proof: See the appendix. ∎

As I said in the introduction, Galles and Pearl (1998) prove a similar completeness result for causal models whose variables satisfy a fixed causal ordering. Given a total ordering $\prec$ on the variables in $\mathcal{V}$, consider the following axiom:

Ord. $Y_{\vec{x}w}(\vec{u}) = Y_{\vec{x}}(\vec{u})$ if $Y \prec W$

Since $\vec{x}$, w, and $\vec{u}$ are implicitly universally quantified in Ord, this axiom says that $\neg(W \rightsquigarrow Y)$ holds if $Y \prec W$. It follows that if $W \rightsquigarrow Y$, then $W \prec Y$. From this and the fact that $\prec$ is a total order, it is easy to see that Ord implies C6.

Galles and Pearl show that C1–C4 and Ord is a sound and complete axiomatization with respect to the class of causal models satisfying Ord for $\mathcal{L}_{\text{GP}}$. More precisely, Galles and Pearl take A_C to consist of the axioms C1–C4 and Ord (but not C0 or MP), and show, in their notation, that $S \models \sigma$ implies $S \vdash_{A_C} \sigma$, where $S \cup \{\sigma\}$ is a set of formulas in $\mathcal{L}_{\text{GP}}$. There is an important subtle point worth stressing about their result: C1 and C2, which are axioms in A_C, are not expressible in $\mathcal{L}_{\mathrm{GP}}$ (since their statement involves disjunction and negation).

So what exactly is Galles and Pearl's result saying? They interpret $S \models \sigma$, as usual, as meaning that in all causal models satisfying S, σ is true.[5] They interpret $S \vdash_{A_C} \sigma$ as meaning that σ is provable from S and the axioms in the axioms of A_C "together with the rules of logic", which presumably means C0 and MP. It follows easily from Theorem 3.2 that their result is correct (see below), but it is unlike typical soundness and completeness proofs, since the proof of σ from S will in general involve formulas in $\mathcal{L}_{\mathrm{uniq}}$ that are not in $\mathcal{L}_{\mathrm{GP}}$. (In particular, this will happen whenever C1–C3 are used in the proof.)

To see that Galles and Pearl's result follows from Theorem 3.2, define S^* to be the formula in $\mathcal{L}_{\mathrm{uniq}}(\mathcal{S})$ which is the conjunction of the formulas in S (there can only be finitely many, since $\mathcal{L}_{\mathrm{GP}}(\mathcal{S})$ itself has only finitely many distinct formulas), together with the conjunction of all the instances of the axiom Ord (again, there are only finitely many). Note that $S \models \sigma$ holds iff $\mathcal{T}_{\mathrm{uniq}}(\mathcal{S}) \models S^* \Rightarrow \sigma$ (since the formulas in Ord guarantee that the only causal models that satisfy S^* are recursive, and hence are in $\mathcal{T}_{\mathrm{uniq}}(\mathcal{S})$). Thus, by Theorem 3.2, $S \models \sigma$ iff $\mathrm{AX}_{\mathrm{uniq}}(\mathcal{S}) \vdash S^* \Rightarrow \sigma$. The latter statement is equivalent to $S \vdash_{A_C} \sigma$, as defined by Galles and Pearl. In fact, Theorem 3.2 shows that $\mathrm{AX}_{\mathrm{uniq}}(\mathcal{S})$ + Ord gives a sound and complete axiomatization with respect to causal models satisfying Ord for the language $\mathcal{L}_{\mathrm{uniq}}(\mathcal{S})$, which allows Boolean connectives. (Of course, Theorem 3.2 shows more, since it extends Galles and Pearl's result to $\mathcal{T}_{\mathrm{rec}}(\mathcal{S})$ and $\mathcal{T}_{\mathrm{uniq}}(\mathcal{S})$.) This suggests that $\mathcal{L}_{\mathrm{uniq}}$ is a more appropriate language for reasoning about causality than $\mathcal{L}_{\mathrm{GP}}$, at least for causal models in $\mathcal{T}_{\mathrm{uniq}}$. $\mathcal{L}_{\mathrm{GP}}$ cannot express a number of properties of causal reasoning of interest (for example, the ones captured by axioms C1–C3). When we use $\mathcal{L}_{\mathrm{uniq}}$, not only is every formula in $\mathcal{L}_{\mathrm{uniq}}$ valid in $\mathcal{T}_{\mathrm{uniq}}$ provable from the axioms in $\mathrm{AX}_{\mathrm{uniq}}$, but the proof involves only formulas in $\mathcal{L}_{\mathrm{uniq}}$.

What about $\mathcal{T}$? I have not been able to find a complete axiomatization for the language $\mathcal{L}_{\mathrm{uniq}}$ with respect to $\mathcal{T}$. However, I do not think that finding a complete axiomatization for $\mathcal{L}_{\mathrm{uniq}}$ with respect to $\mathcal{T}$ is of great interest, because $\mathcal{L}_{\mathrm{uniq}}$ is simply not a language appropriate for reasoning about causality in $\mathcal{T}$. Because there is not necessarily a unique solution to the equations that arise in a causal model $T \in \mathcal{T}$, it is useful to be able to say both that there exists a solution with certain properties and that *all* solutions have certain properties. This is precisely what the language $\mathcal{L}^+$ lets us do.[6] As I now show, there is in fact an elegant sound and complete axiomatization for $\mathcal{L}^+$ with respect to $\mathcal{T}$.

Consider the following axioms:

D0. All instances of propositional tautologies.

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

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

D3. $\langle \vec{X} \leftarrow \vec{x} \rangle (W(\vec{u}) = w \wedge \vec{Y}(\vec{u}) = \vec{y}) \Rightarrow \langle \vec{X} \leftarrow \vec{x}; W \leftarrow w \rangle (\vec{Y}(\vec{u}) = \vec{y})$ (composition)

5. Although they do not say this explicitly, it is clear that they intend to further restrict to casual models satisfying *S and Ord*, for the fixed order $\prec$. Without this restriction, their result is not true.

6. Note that $\mathcal{L}^+$ allows us to say that there is a unique solution for a random variable X after setting some other variables. For example, $\langle \vec{Y} \leftarrow \vec{y} \rangle true(\vec{u}) \wedge [\vec{Y} \leftarrow \vec{y}](X(\vec{u}) = x)$ says that there are solutions to the equations when $\vec{Y}$ is set to $\vec{y}$ and $\mathcal{U}$ is set to $\vec{u}$ and, in all of them, X is uniquely determined to be x.

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

D5. $(\langle \vec{X} \leftarrow \vec{x}; Y \leftarrow y\rangle(W(\vec{u}) = w \wedge \vec{Z}(\vec{u}) = \vec{z}) \wedge \langle \vec{X} \leftarrow \vec{x}; W \leftarrow w\rangle(Y(\vec{u}) = y \wedge \vec{Z}(\vec{u}) = \vec{z}))$
$\Rightarrow \langle \vec{X} \leftarrow \vec{x}\rangle W(\vec{u}) = w \wedge Y(\vec{u}) = y \wedge \vec{Z}(\vec{u}) = \vec{z}))$, where $\vec{Z} = \mathcal{V} - (\vec{X} \cup \{W, Y\})$ (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. $\langle \vec{Y} \leftarrow \vec{y}\rangle \mathit{true}(\vec{u}) \wedge \vee_{x \in \mathcal{R}(X)}[\vec{Y} \leftarrow \vec{y}](X(\vec{u}) = x)$ if $Y = \mathcal{V} - \{X\}$ (unique solutions for $\mathcal{V} - \{X\}$)

D10. $\langle \vec{Y} \leftarrow \vec{y}\rangle \mathit{true}(\vec{u}) \wedge \vee_{x \in \mathcal{R}(X)}[\vec{Y} \leftarrow \vec{y}](X(\vec{u}) = x)$ (unique solutions)

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

D1–D6 are the analogues of C1–C6 in $\mathcal{L}^+$. D4 and D6 are just C4 and C6, with no changes at all. The other axioms are not quite the same though. For example, C1 is actually $[\vec{Y} \leftarrow \vec{y}](X(\vec{u}) = x) \Rightarrow \neg[\vec{Y} \leftarrow \vec{y}](X(\vec{u}) = x')$ if $x \neq x'$. By Lemma 2.2, this is equivalent to D1 in $\mathcal{T}_{\text{uniq}}$; however, the two formulas are not equivalent in general. Similarly, C2 is $\vee_{x \in \mathcal{R}(X)}[\vec{Y} \leftarrow \vec{y}](X(\vec{u}) = x)$, which is closer to D10 than D2 (since the disjunction is outside the scope of the $[\vec{Y} \leftarrow \vec{y}]$). Again, D10 and D2 are equivalent in $\mathcal{T}_{\text{uniq}}$ (both are equivalent to C2 in this case) but, in general, D10 is stronger than D2. Only D2 and D9, both weaker than D10, hold in $\mathcal{T}$. The exact analogue of C3 would use $[\,]$ instead of $\langle\,\rangle$ and say $Y(\vec{u}) = y$ instead of $\vec{Y}(\vec{u}) = \vec{y}$. For completeness, it is necessary to have a vector of variables here. Using $[\,]$ instead of $\langle\,\rangle$ also results in a valid formula (and would not require a vector $\vec{Y}$). While the two variants are equivalent in $\mathcal{T}_{\text{uniq}}$, they are different in general, and the one given here is the more useful. (More precisely, with it we get completeness, while the version with $[\,]$ does not suffice for completeness.) Similarly, in D5, we use $\langle\,\rangle$ instead of $[\,]$, and add the extra clause $\vec{Z}(\vec{u}) = \vec{z}$. Both turn out to be necessary for soundness. In some sense, we can think of D1–D6 as capturing the "true content" of C1–C6, once we drop the assumption that the structural equations have a unique solution. D7 and D8 are standard properties of modal operators. D10 is what we need to capture the fact that the structural equations have unique solutions. D11 essentially says that the solutions to the equations that arise when the exogenous variables are set to $\vec{u}$ are independent of the solutions that arise when the exogenous variables are set to $\vec{u}' \neq \vec{u}$.

Let AX^+ consist of D0–D5, D7–D9, D11, and MP (modus ponens); let $\mathrm{AX}^+_{\text{uniq}}$ be the result of adding D10 to AX^+; let $\mathrm{AX}^+_{\text{rec}}$ be the result of adding D6 to $\mathrm{AX}^+_{\text{uniq}}$.

Theorem 3.3: *$AX^+(\mathcal{S})$ (resp., $AX^+_{\text{uniq}}(\mathcal{S})$, $AX^+_{\text{rec}}(\mathcal{S})$) is a sound and complete axiomatization for $\mathcal{L}^+(\mathcal{S})$ with respect to $\mathcal{T}(\mathcal{S})$ (resp., $\mathcal{T}_{\text{uniq}}(\mathcal{S})$, $\mathcal{T}_{\text{rec}}(\mathcal{S})$).*

Proof: See the appendix. ∎

4. Decision Procedures

In this section I consider the complexity of deciding if a formula is satisfiable (or valid). This, of course, depends on the language ($\mathcal{L}^+$, $\mathcal{L}_{\rm uniq}$, or $\mathcal{L}_{\rm GP}$) and the class of models ($\mathcal{T}_{\rm rec}$, $\mathcal{T}_{\rm uniq}$, $\mathcal{T}$) we consider. It also depends on how we formulate the problem.

One version of the problem is to consider a fixed signature $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R})$, and ask how hard it is to decide if a formula $\varphi \in \mathcal{L}^+(\mathcal{S})$ (resp., $\mathcal{L}_{\rm uniq}(\mathcal{S})$, $\mathcal{L}_{\rm GP}(\mathcal{S})$) is satisfiable in $\mathcal{T}_{\rm rec}(\mathcal{S})$ (resp., $\mathcal{T}_{\rm uniq}(\mathcal{S})$, $\mathcal{T}(\mathcal{S})$). If $\mathcal{S}$ is finite (that is, if $\mathcal{V}$ and $\mathcal{U}$ are finite and $\mathcal{R}(Y)$ is finite for each $Y \in \mathcal{U} \in \mathcal{V}$), this turns out to be quite easy, for trivial reasons.

Theorem 4.1: *If $\mathcal{S}$ is a fixed finite signature, the problem of deciding if a formula $\varphi \in \mathcal{L}^+(\mathcal{S})$ (resp., $\mathcal{L}_{\rm uniq}(\mathcal{S})$, $\mathcal{L}_{\rm GP}(\mathcal{S})$) is satisfiable in $\mathcal{T}_{\rm rec}(\mathcal{S})$ (resp., $\mathcal{T}_{\rm uniq}(\mathcal{S})$, $\mathcal{T}(\mathcal{S})$) can be solved in time linear in $|\varphi|$ (the length of φ viewed as a string of symbols).*

Proof: If $\mathcal{S}$ is finite, there are only finitely many causal models in $\mathcal{T}(\mathcal{S})$, independent of φ. Given φ, we can explicitly check if φ is satisfied in any (or all) of them. This can be done in time linear in $|\varphi|$. Since $\mathcal{S}$ is not a parameter to the problem, the huge number of possible causal models that we have to check affects only the constant. ∎

We can do even better than Theorem 4.1 suggests if $\mathcal{S}$ is a fixed finite signature. Suppose that $\mathcal{V}$ consists of 100 variables and φ mentions only 3 of them. A causal model must specify the equations for all 100 variables. Is it really necessary to consider what happens to the 97 variables not mentioned in φ to decide if φ is satisfiable or valid? As the following result shows, if we restrict to models in $\mathcal{T}_{\rm uniq}$, then we need to check only the variables that appear in $\mathcal{S}$. Given a signature $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R})$, let $\mathcal{S}_\varphi = (\{U^*\}, \mathcal{V}_\varphi, \mathcal{R}_\varphi)$, where $\mathcal{V}_\varphi$ consists of the variables in $\mathcal{V}$ that appear in φ, U^* is a fresh exogenous variable, not mentioned in $\mathcal{V}$ or $\mathcal{U}$, $\mathcal{R}_\varphi(X) = \mathcal{R}(X)$ for $X \in \mathcal{V}_\varphi$, and $\mathcal{R}_\varphi(U^*)$ consists of all those tuples in $\times_{U \in \mathcal{U}} \mathcal{R}(U)$ that are mentioned in φ.

Theorem 4.2: *A formula $\varphi \in \mathcal{L}^+(\mathcal{S})$ is satisfiable in $\mathcal{T}_{\rm rec}(\mathcal{S})$ (resp., $\mathcal{T}_{\rm uniq}(\mathcal{S})$) iff it is satisfiable in $\mathcal{T}_{\rm rec}(\mathcal{S}_\varphi)$ (resp., $\mathcal{T}_{\rm uniq}(\mathcal{S}_\varphi)$).*

Proof: See the appendix. ∎

The analogue to Theorem 4.2 does not hold for $\mathcal{T}$. For example, suppose that $\mathcal{S} = (\emptyset, \{X, Y, Z\}, \mathcal{R})$, where $\mathcal{R}(X) = \mathcal{R}(Y) = \mathcal{R}(Z) = \{0, 1\}$, and φ is the formula $\langle X \leftarrow 0\rangle(Y = 0) \wedge \langle X \leftarrow 0\rangle(Y = 1)$. It is easy to see that there is a causal model in $\mathcal{T}(\mathcal{S})$ satisfying φ. For example, if $T = (\mathcal{S}, \mathcal{F})$, where $F_X(y, z) = y \oplus z$, $F_Y(x, z) = x \oplus z$ and $F_Z(x, y) = x \oplus y$, and $\oplus$ represents addition mod 2, then it is easy to check that $T \models \varphi$. On the other hand, there is no causal model $T' \in \mathcal{T}(\mathcal{S}_\varphi)$ such that $T' \models \varphi$. For suppose $T' \models \varphi$ and $T' = (\mathcal{S}_\varphi, \mathcal{F}')$. Since $T' \models \langle X \leftarrow 0\rangle(Y = 0)$, we must have $F'_Y(0) = 0$; since $T' \models \langle X \leftarrow 0\rangle(Y = 1)$, we must have $F'_Y(0) = 1$. But we cannot have both $F'_Y(0) = 0$ and $F'_Y(1) = 1$, since F'_Y is a function.

There is a variant of Theorem 4.2 that does hold for $\mathcal{T}$ that does give us a bound on the number of variables we need to consider. Given a signature $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R})$, define $||\mathcal{S}|| = \times_{X \in \mathcal{V}} |\mathcal{R}(X)|$ (where we take $||\mathcal{S}|| = \infty$ if either $\mathcal{V}$ is infinite or $|\mathcal{R}(X)| = \infty$ for some $X \in \mathcal{V}$). If $||\mathcal{S}|| > ||\mathcal{S}_\varphi||^2 + ||\mathcal{S}_\varphi||$, let $\mathcal{S}^+_\varphi = (\{U^*\}, \mathcal{V}^+_\varphi, \mathcal{R}^+_\varphi)$, where $\mathcal{V}^+_\varphi$ is $\mathcal{V}_\varphi$ as defined above together with one fresh endogenous variable X^*, $\mathcal{R}^+_\varphi(X^*) = \times_{X \in \mathcal{V}_\varphi} \mathcal{R}(X)$, and $\mathcal{R}^+_\varphi(U^*) = \mathcal{R}_\varphi(U^*)$. If $||\mathcal{S}|| \leq ||\mathcal{S}_\varphi||^2 + ||\mathcal{S}_\varphi||$, let $\mathcal{S}^+_\varphi = (\{U^*\}, \mathcal{V}, \mathcal{R}')$, where $\mathcal{R}'(X) = \mathcal{R}(X)$ for $X \in \mathcal{V}$ and $\mathcal{R}'(U^*) = \mathcal{R}_\varphi(U^*)$.

Theorem 4.3: *A formula $\varphi \in \mathcal{L}^+(\mathcal{S})$ is satisfiable in $\mathcal{T}(\mathcal{S})$ iff φ is satisfiable in $\mathcal{T}(\mathcal{S}^+_\varphi)$.*

Proof: See the appendix. ∎

Note that if $||\mathcal{S}|| \leq ||\mathcal{S}_\varphi||^2 + ||\mathcal{S}_\varphi||$, then, since we have assumed (without loss of generality) that $|\mathcal{R}(X)| \geq 2$ for each variable X, it must be the case that there are at most $2\log_2(||\mathcal{S}_\varphi||) + 1$ variables in signature $\mathcal{S}$.

Since Theorems 4.2 and 4.3 apply to all formulas in $\mathcal{L}^+(\mathcal{S})$, they apply *a fortiori* to formulas in $\mathcal{L}_{\text{uniq}}(\mathcal{S})$ and $\mathcal{L}_{\text{GP}}(\mathcal{S})$. Although stated only in terms of satisfiability, it is immediate that they also hold for validity. Thus, they tell us that, without loss of generality, when considering satisfiability or validity, we need to consider only finitely many variables (essentially, the ones that appear in φ, and perhaps a few more). In this sense, we can restrict to signatures with only finitely many variables without loss of generality. Note that these results do *not* tell us that we can restrict to finite sets of values for these variables without loss of generality.

Returning to the complexity of the decision problem, note that Theorem 4.1 is the analogue of the observation that for propositional logic, the satisfiability problem is in linear time if we restrict to a fixed set of primitive propositions. The proof that the satisfiability problem for propositional logic is NP-complete implicitly assumes that we have an unbounded number of primitive propositions at our disposal.

There are two ways to get an analogous result here. The first is to allow the signature $\mathcal{S}$ to be infinite and the second is to make the signature part of the input to the problem. The results in both cases are similar, so I just consider the case where the signature is part of the input here.

Theorem 4.4: *Given as input a pair $(\varphi, \mathcal{S})$, where $\varphi \in \mathcal{L}^+(\mathcal{S})$ (resp., $\mathcal{L}_{\text{uniq}}(\mathcal{S})$) and $\mathcal{S}$ is a finite signature, the problem of deciding if φ is satisfiable in $\mathcal{T}_{\text{rec}}(\mathcal{S})$ (resp., $\mathcal{T}_{\text{uniq}}(\mathcal{S})$, $\mathcal{T}(\mathcal{S})$) is NP-complete (resp., NP-hard) in $|\varphi|$; if $\varphi \in \mathcal{L}_{\text{GP}}(\mathcal{S})$, then the problem of deciding if φ is satisfiable in $\mathcal{T}_{\text{rec}}(\mathcal{S})$ (resp., $\mathcal{T}_{\text{uniq}}(\mathcal{S})$) is NP-complete (resp., NP-hard).*

Proof: See the appendix. ∎

I believe that the problem of deciding if a formula φ in $\mathcal{L}_{\text{uniq}}(\mathcal{S})$ or $\mathcal{L}^+(\mathcal{S})$ is satisfiable in $\mathcal{T}_{\text{uniq}}(\mathcal{S})$ and $\mathcal{T}(\mathcal{S})$ is NP-complete, as is the case of deciding if $\varphi \in \mathcal{L}_{\text{GP}}(\mathcal{S})$ is satisfiable in $\mathcal{T}_{\text{uniq}}(\mathcal{S})$. However, I have not been able to show this. What about the satisfiability problem for formulas in $\mathcal{L}_{\text{GP}}$ in $\mathcal{T}(\mathcal{S})$? This may well be in constant time! Indeed, if $\mathcal{S}$ is an infinite signature (that is, if $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R})$ and $|\mathcal{V}| = \infty$), then it is provably in constant time. The point is that a formula in $\mathcal{L}_{\text{GP}}(\mathcal{S})$ is trivially satisfiable in a structure $T \in \mathcal{L}_{\text{GP}}(\mathcal{S})$ where for all settings $\vec{X} \leftarrow \vec{x}$, the equations in $T_{\vec{X} \leftarrow \vec{x}}$ have no solutions, and there always is such model structure if $\mathcal{S}$ has infinitely many variables. If $\mathcal{S}$ has only finitely many variables, we do not have such trivial models, but it may still be possible to show that a "trivial enough" model exists that satisfies the formula. This just emphasizes that $\mathcal{L}_{\text{GP}}(\mathcal{S})$ is simply too weak a language to reason about models in $\mathcal{T}(\mathcal{S})$.

5. Conclusion

I have provided complete axiomatizations and decision procedures for propositional languages for reasoning about causality. I have tried to stress the important role of the choice of language (and the signature) in both the axiomatizations and, more generally, in the reasoning process.

Both the models and the languages considered here are somewhat limited. For example, a more general approach to modeling causality would allow there to be more than one value of X once we have set all the other variables. This would be appropriate if we model things at a somewhat coarser level of granularity, where the values of all the variables other than X do not suffice to completely determine the value of X. I believe the results of this paper can be extended in a straightforward way to deal with this generalization, although I have not checked the details. For general causal reasoning, I believe we need a richer language, which includes some first-order features. I hope to return to the issue of finding appropriate richer languages for causal reasoning in future work.

Acknowledgments

I'd like to thank Judea Pearl for his comments on a previous version of this paper, as well as his generous help in providing pointers to the literature. This work was supported in part by NSF under grant IRI-96-25901 and by the Air Force Office of Scientific Research under grant F49620-96-1-0323. A preliminary version of this paper appears in *Proc. Fourteenth Conference on Uncertainty in AI*, pp. 202–210, 1998.

Appendix A. Proofs

Theorem 3.2: *AX_{uniq} (resp., AX_{rec}) is a sound and complete axiomatization for $\mathcal{L}_{\mathrm{uniq}}(\mathcal{S})$ with respect to $\mathcal{T}_{\mathrm{uniq}}(\mathcal{S})$ (resp., $\mathcal{T}_{\mathrm{rec}}(\mathcal{S})$).*

Proof: Soundness is proved by Galles and Pearl. To make the paper self-contained, I reprove the only non-obvious case—the validity of C5 in $\mathcal{T}_{\mathrm{uniq}}$.

Let $T \in \mathcal{T}_{\mathrm{uniq}}$ and suppose that $T \models Y_{\vec{x}w}(\vec{u}) = y \wedge W_{\vec{x}y}(\vec{u}) = w$. We want to show that $T \models Y_{\vec{x}}(\vec{u}) = y$. Since we are in $\mathcal{T}_{\mathrm{uniq}}$, there is a unique vector $\vec{v}_1$ that satisfies the equations in $T_{\vec{x}w}(\vec{u})$ and a unique vector $\vec{v}_2$ that satisfies the equations in $T_{\vec{x}y}(\vec{u})$. I claim that $\vec{v}_1 = \vec{v}_2$. By assumption, the $\vec{X}$, Y, and W components of these vectors are the same ($\vec{x}$, y, and w, respectively). Now consider the $T_{\vec{x}yw}(\vec{u})$. I claim that $\vec{v}_1$ and $\vec{v}_2$ are both solutions to the equations in that causal theory. Note that for any variable Z other than those in $\vec{X} \cup \{W, Y\}$, the equation $F_Z^{\vec{x}w,\vec{u}}$ for Z in $T_{\vec{x}w}(\vec{u})$ is the same as the equations $F_Z^{\vec{x}y,\vec{u}}$ and $F_Z^{\vec{x}yw,\vec{u}}$ for Z in $T_{\vec{x}y}(\vec{u})$ and $T_{\vec{x}yw}(\vec{u})$, respectively, except that in the first case, w has been plugged in as the value of W, in the second case y has been plugged in as the value of Y, and in the third case, both w and y have been plugged in. However, since w and y are the values of W and Y, respectively, in both $\vec{v}_1$ and $\vec{v}_2$, and since these vectors satisfy both equation $F_Z^{\vec{x}w}$ and $F_Z^{\vec{x}y}$, they must also satisfy $F_Z^{\vec{x}wy}$. Since the equations for $T_{\vec{x}yw}(\vec{u})$ have a unique solution, we have that $\vec{v}_1 = \vec{v}_2$, as desired.

Next, I claim that $\vec{v}_1$ satisfies the equations in $T_{\vec{x}}(\vec{u})$. Again, as above, it is clear that it satisfies the equation for $Z \notin \vec{X} \cup \{W, Y\}$. A similar argument shows that it satisfies the equation for Y in $T_{\vec{x}}(\vec{u})$, since $\vec{v}_1$ satisfies the equation for Y in $T_{\vec{x}w}(\vec{u})$. Finally, a similar argument shows that it satisfies the equation for W in $T_{\vec{x}}(\vec{u})$, since $\vec{v}_2 = \vec{v}_1$ satisfies the equation for W in $T_{\vec{x}y}(\vec{u})$. Since the Y component of $\vec{v}_1$ is y, it follows that $Y_{\vec{x}}(\vec{u}) = y$.

So much for soundness. For completeness, as usual, it suffices to prove that if a formula in $\mathcal{L}_{\text{uniq}}$ is consistent with AX_{uniq} (resp., AX_{rec}), then it is satisfied in a causal model in $\mathcal{T}_{\text{uniq}}$ (resp., $\mathcal{T}_{\text{rec}}$). (Here's the argument: We want to show that every valid formula is provable. Suppose that we have shown that every consistent formula is satisfiable and that φ is valid. If φ is not provable, then $\neg\varphi$ is consistent. By assumption, this means that $\neg\varphi$ is satisfiable, contradicting the assumption that φ is valid.)

I now give the argument in the case of AX_{uniq}.

Suppose that a formula $\varphi \in \mathcal{L}_{\text{uniq}}(\mathcal{S})$, with $\mathcal{S} = (\mathcal{U}, \mathcal{V}, V)$, is consistent with AX_{uniq}. Consider a maximal consistent set C of formulas that includes φ. (A maximal consistent set is a set of formulas whose conjunction is consistent such that any larger set of formulas would be inconsistent.) It follows easily from standard propositional reasoning (i.e., using C0 and MP only) that such a maximal consistent set exists. Moreover, from C1 and C2, it follows that for each random variable $X \in \mathcal{V}$ and vector $\vec{y}$ of values, there exists exactly one element $x \in \mathcal{R}(X)$ such that $X_{\vec{y}} = x \in C$. I now construct a causal model $T = (\mathcal{S}, F) \in \mathcal{T}_{\text{uniq}}(\mathcal{S})$ that satisfies every formula in C (and, in particular, satisfies φ).

A term $X_{\vec{Y}\leftarrow\vec{y}}(\vec{u})$ is *complete (for X)* if $\vec{Y}$ consists of all the variables in $\mathcal{V} - \{X\}$. Thus, $X_{\vec{Y}\leftarrow\vec{y}}(\vec{u})$ is a complete term if every random variable other than X is determined. We use the complete terms to define the structural equations. For each variable in $X \in \mathcal{V}$, define $F_X(\vec{u}, \vec{y}) = x$ if $X_{\vec{y}}(\vec{u}) = x$, where $X_{\vec{y}}(\vec{u})$ is a complete term. This gives us a causal model T. Now we have to show that this model is in $\mathcal{T}_{\text{uniq}}$ and that all the formulas in C are satisfied by T.

I show that $X_{\vec{Y}\leftarrow\vec{y}}(\vec{u}) = x$ is in C iff $T \models X_{\vec{Y}\leftarrow\vec{y}}(\vec{u}) = x$ by induction on $|\mathcal{V}| - |\vec{Y}|$. The case where $|\mathcal{V}| - |\vec{Y}| = 0$ follows immediately from C4, since then X is in $\vec{Y}$. If $|\mathcal{V}| - |\vec{Y}| \neq 0$, we can assume without loss of generality that X is not in $\vec{Y}$, for otherwise the result again follows from C4. Given this assumption, if $|\mathcal{V}| - |\vec{Y}| = 1$, the result follows by definition of the equations F_X.

For the general case, suppose that $|\mathcal{V}| - |\vec{Y}| = k > 1$. We want to show that there is a unique solution to the equations in $T_{\vec{Y}\leftarrow\vec{y}}(\vec{u})$ and that, in this solution, X has value x. To see that there is a solution, we define a vector $\vec{v}$ and show that it is in fact a solution. If $W \in \vec{Y}$ and $W \leftarrow w$ is the assignment to W in $\vec{Y} \leftarrow \vec{y}$, then we set the W component of $\vec{v}$ to w. If W is not in $\vec{Y}$, then set the W component of $\vec{v}$ to the unique value w^* such that $W_{\vec{Y}\leftarrow\vec{y}}(\vec{u}) = w^*$ is in C. (By C1 and C2 there is such a unique value w.) I claim that $\vec{v}$ is a solution to the equations in $T_{\vec{Y}\leftarrow\vec{y}}(\vec{u})$.

To see this, let W be a variable in $\mathcal{V}$ not in $\vec{Y}$. Let $\vec{Y}' = \vec{Y}W$. By C3 and C4, for every variable $Z \in \mathcal{V} - \vec{Y}'$, we have $Z_{\vec{y}w^*}(\vec{u}) = z^*$. Since $|\mathcal{V}| - |\vec{Y}'| = k - 1$, by the inductive hypothesis, $\vec{v}$ is in fact the unique solution for $T_{\vec{y}w^*}(\vec{u})$. For every variable Z in $\mathcal{V} - \vec{Y}'$, the equation $F_Z^{\vec{y}w^*,\vec{u}}$ for Z in $T_{\vec{y}w^*}(\vec{u})$ is the same as the equation $F_Z^{\vec{y},\vec{u}}$ for Z in $T_{\vec{y}}(\vec{u})$, except that W is set to w^*. Thus, every equation in $T_{\vec{y}}(\vec{u})$ except possibly the equation $F_W^{\vec{y},\vec{u}}$ is satisfied by $\vec{v}$. To see that $F_W^{\vec{y},\vec{u}}$ is also satisfied by $\vec{v}$, simply repeat this argument above starting with another variable W' in $\mathcal{V}-\vec{Y}$. (Such a variable must exist since $|\mathcal{V}|-|\vec{Y}|$ was assumed to be at least 2.)

It remains to show that $\vec{v}$ is the *unique* solution to the equations in $T_{\vec{y}}(\vec{u})$. Suppose there were another solution, say $\vec{v}'$, to the equations. Suppose that for each variable W in $\mathcal{V}-\vec{Y}$, the W component of $\vec{v}'$ is w^{**}. For some variable Z, we must have $z^{**}\neq z^*$. Since $Z_{\vec{y}}(\vec{u})=z^*$, by assumption, it follows from C1 that $Z_{\vec{y}}(\vec{u})\neq z^{**}$ is in C (since C is a maximal consistent set). It is also easy to see that for each W in $\mathcal{V}-\vec{Y}$, the vector $\vec{v}'$ is also a solution to the equations in $T_{\vec{y}w^{**}}(\vec{u})$. Let W be a variable other than Z in $\mathcal{V}-\vec{Y}$. By the induction hypothesis, it follows that $W_{\vec{y}z^{**}}(\vec{u})=w^{**}$ and $Z_{\vec{y}w^{**}}(\vec{u})=z^{**}$ are both in C. By C5 (reversibility), $Z_{\vec{y}}(\vec{u})=z^{**}$ is in C. But this contradicts the consistency of C.

This completes the proof in the case of $\mathcal{T}_{\rm uniq}(\mathcal{S})$. Essentially the same proof works for $\mathcal{T}_{\rm rec}$. We just need to observe that C6 guarantees that the theory we construct can be taken to be recursive. To see this, given a formula φ consistent with $\mathcal{T}_{\rm rec}$, consider a maximal set C of formulas consistent with $\mathcal{T}_{\rm rec}$ that contains φ. Let T_C be the causal model determined by C, as above. The set C also determines a relation $\prec$ on the exogenous variables: define $Y\prec Z$ if $Y\rightsquigarrow Z\in C$. It easily follows from C6 that the transitive closure $\prec^*$ of $\prec$ is a partial order: if $X\prec^* Y$ and $Y\prec^* X$, then $X=Y$. Any total order on the variables consistent $\prec^*$ gives an ordering for which T_C is recursive. ∎

Theorem 3.3: *AX^+ (resp., $AX^+_{\rm uniq}$, $AX^+_{\rm rec}$) is a sound and complete axiomatization for $\mathcal{L}^+(\mathcal{S})$ with respect to $\mathcal{T}(\mathcal{S})$ (resp., $\mathcal{T}_{\rm uniq}(\mathcal{S})$, $\mathcal{T}_{\rm rec}(\mathcal{S})$).*

Proof: Soundness proceeds much as that of Theorem 3.2; I leave details to the reader. For completeness, we again proceed much as in the proof of Theorem 3.2. Because the proofs are so similar in spirit, I just sketch the proof for AX^+; the modifications for $\mathrm{AX}^+_{\rm uniq}$ and $\mathrm{AX}^+_{\rm rec}$ are left to the reader.

Again, given a formula φ consist with AX^+, we consider a maximal consistent set of formulas containing φ that is consistent with AX^+, and use it to construct a causal model T. Note that D9 suffices for this, because in defining $F_X(\vec{u},\vec{y})$, we needed to know only the unique x such that $[\vec{Y}\leftarrow\vec{y}](X(\vec{u})=x)$ for $\vec{Y}=\mathcal{V}-X$, and D9 (together with D1) assures us that there is a unique such x. Again, we want to show that all the formulas in C are satisfied by T.

To do this, it clearly suffices to show that for every formula ψ of the form $\langle\vec{Y}\leftarrow\vec{y}\varphi$, we have ψ in C iff $T\models\psi$. We can reduce to considering even simpler formulas, namely, ones where φ has the form $\vec{X}(\vec{u})=\vec{x}$, by applying some of the axioms. To see this, first observe that standard arguments of modal logic (using D0, D7, D8, and MP) show that $\langle\vec{Y}\leftarrow\vec{y}\rangle(\varphi_1\vee\varphi_2)$ is provably equivalent to $\langle\vec{Y}\leftarrow\vec{y}\rangle\varphi_1\vee\langle\vec{Y}\leftarrow\vec{y}\rangle\varphi_2$. That means we can assume without loss of generality that φ is a conjunction of formulas of the form $X(\vec{u})\leftarrow x$ and their negations. From D2 it follows that $\langle\vec{Y}\leftarrow\vec{y}\rangle(\varphi\wedge X(\vec{u})\neq x)$ is equivalent to $\langle\vec{Y}\leftarrow\vec{y}\rangle(\varphi\wedge(\vee_{x'\in\mathcal{R}(X)-\{x\}}X(\vec{u})=x')$. Thus, we can assume without loss of generality that φ has no negations. By applying D11, we can assume without loss of generality that the same setting $\vec{u}$ of the exogenous variables is used in all the conjuncts. Thus, it suffices to show that $\langle\vec{Y}\leftarrow\vec{y}\rangle(\vec{X}(\vec{u})=\vec{x})\in C$ iff $T\models\langle\vec{Y}\leftarrow\vec{y}\rangle(\vec{X}(\vec{u})=\vec{x})$ for $\vec{X}=\mathcal{V}-\vec{Y}$.

To do this, we proceed by induction on $|\mathcal{V}|-|\vec{Y}|$ again. The base case is dealt with using D4, as before. So assume that $k\geq 1$ and $|\mathcal{V}|-|\vec{Y}|=k+1$. Suppose that $\langle\vec{Y}\leftarrow\vec{y}\rangle(\vec{X}(\vec{u})=\vec{x})\in C$. Let $X_1,X_2\in\vec{X}$. Suppose that $X_1\leftarrow x_1$ and $X_2\leftarrow x_2$ are the assignments to

X_1 and X_2 in $\vec{X} \leftarrow \vec{x}$. Let $\vec{X}' \leftarrow \vec{x}'$ and $\vec{X}'' \leftarrow \vec{x}''$ be the result of removing $X_1 \leftarrow x_1$ and $X_2 \leftarrow x_2$, respectively, from $\vec{X} \leftarrow \vec{x}$. By D3, both $\langle \vec{Y} \leftarrow \vec{y}; X_1 \leftarrow x_1 \rangle (\vec{X}''(\vec{u}) = \vec{x}'')$ and $\langle \vec{Y} \leftarrow \vec{y}; X_2 \leftarrow x_2 \rangle (\vec{X}'(\vec{u}) = \vec{x}')$ are in C. By the induction hypothesis, both of these formulas are true in T. By the soundness of D5, it follows that $T \models \langle \vec{Y} \leftarrow \vec{y} \rangle (\vec{X}(\vec{u}) = \vec{x}')$, as desired.

Conversely, suppose that $T \models \langle \vec{Y} \leftarrow \vec{y} \rangle (\vec{X}(\vec{u}) = \vec{x}')$. Then, since D3 is sound, we have that $T \models \langle \vec{Y} \leftarrow \vec{y}; X_1 \leftarrow x_1 \rangle (\vec{X}''(\vec{u}) = \vec{x}'')$ and $T \models \langle \vec{Y} \leftarrow \vec{y}; X_2 \leftarrow x_2 \rangle (\vec{X}'(\vec{u}) = \vec{x}')$. By the induction hypothesis, we have that both $\langle \vec{Y} \leftarrow \vec{y}; X_1 \leftarrow x_1 \rangle (\vec{X}''(\vec{u}) = \vec{x}'')$ and $\langle \vec{Y} \leftarrow \vec{y}; X_2 \leftarrow x_2 \rangle (\vec{X}'(\vec{u}) = \vec{x}')$ are in C. We now apply D5 to complete the proof. ∎

Theorem 4.2: *A formula $\varphi \in \mathcal{L}^+(\mathcal{S})$ is satisfiable in $\mathcal{T}_{\mathrm{rec}}(\mathcal{S})$ (resp., $\mathcal{T}_{\mathrm{uniq}}(\mathcal{S})$) iff it is satisfiable in $\mathcal{T}_{\mathrm{rec}}(\mathcal{S}_\varphi)$ (resp., $\mathcal{T}_{\mathrm{uniq}}(\mathcal{S}_\varphi)$).*

Proof: Clearly, if a formula is satisfiable in $\mathcal{T}_{\mathrm{rec}}(\mathcal{S}_\varphi)$ (resp., $\mathcal{T}_{\mathrm{uniq}}(\mathcal{S}_\varphi)$), then it is satisfiable in $\mathcal{T}_{\mathrm{rec}}(\mathcal{S})$ (resp., $\mathcal{T}_{\mathrm{uniq}}(\mathcal{S})$). We can easily convert a causal model $T = (\mathcal{S}_\varphi, \mathcal{F}) \in \mathcal{T}_{\mathrm{rec}}(\mathcal{S}_\varphi)$ satisfying φ to a causal model $T' = (\mathcal{S}, \mathcal{F}') \in \mathcal{T}_{\mathrm{rec}}(\mathcal{S})$ satisfying φ by simply defining F'_X to be a constant, independent of its arguments, for $X \in \mathcal{V} - \mathcal{V}_\varphi$; if $X \in \mathcal{V}_\varphi$, define $F'_X(\vec{u}, \vec{x}, \vec{y}) = F_X(\vec{u}, \vec{x})$, where $\vec{u} \in \mathcal{R}(U^*)$, $\vec{x} \in \times_{Y \in \mathcal{V}_\varphi - \{X\}} \mathcal{R}(Y)$ and $\vec{y} \in \times_{Y \in \mathcal{V} - \mathcal{V}_\varphi} \mathcal{R}(Y)$; if $\vec{u} \notin \mathcal{R}(U^*)$, define $F'_X(\vec{u}, \vec{x}, \vec{y})$ to be an arbitrary constant. An identical transformation works for $T \in \mathcal{T}_{\mathrm{uniq}}(\mathcal{S}_\varphi)$.

For the converse, suppose that φ is satisfiable in a causal model $T = (\mathcal{S}, \mathcal{F}) \in \mathcal{T}_{\mathrm{rec}}(\mathcal{S})$. Thus, there is an ordering $\prec$ on the variables in $\mathcal{V}$ such that if $X \prec Y$, then F_X is independent of the value of Y. This means we can view F_X as a function of the exogenous variables in $\mathcal{U}$ and the variables $Y \in \mathcal{V}$ such that $Y \prec X$. Let $\mathrm{Pre}(X) = \{Y \in \mathcal{V} : Y \prec X\}$. For convenience, I allow F_X to take as arguments the values of only the variables in $\mathcal{U} \cup \mathrm{Pre}(X)$, rather than requiring its arguments to include the values of all the variables in $\mathcal{U} \cup \mathcal{V} - \{X\}$. Now define functions $F'_X : (\times_{U \in \mathcal{U}} \mathcal{R}(U)) \times (\times_{Y \in \mathcal{V}_\varphi - \{X\}} \mathcal{R}(Y)) \to \mathcal{R}(X)$ for all $X \in \mathcal{V}$ by induction on $\prec$ (that is, start with the $\prec$-minimal element, whose value is independent of that of all the other variables, and work up the $\prec$ chains). Suppose $X \in \mathcal{V}_\varphi$ and $\vec{x}$ is a vector of values for the variables in $\mathcal{V}_\varphi - \{X\}$. If X is $\prec$-minimal, then define $F'_X(\vec{u}, \vec{x}) = F_X(\vec{u})$. In general, define $F'_X(\vec{u}, \vec{x}) = F_X(\vec{u}, \vec{z})$, where $\vec{z}$ is a vector of values for the variables in $\mathrm{Pre}(X)$ defined as follows. If $Y \in \mathrm{Pre}(X) \cap \mathcal{V}_\varphi$, then the value of the Y component in $\vec{z}$ is the value of the Y component in $\vec{y}$; if $Y \in \mathrm{Pre}(X) - \mathcal{V}_\varphi$, then the value of the Y component in $\vec{z}$ is $F'_Y(\vec{u}, \vec{x})$. (By the induction hypothesis, $F'_Y(\vec{u}, \vec{x})$ has already been defined.) Now define a causal model $T' = (\mathcal{S}_\varphi, \mathcal{F}')$. It is easy to check that $T' \in \mathcal{T}_{\mathrm{rec}}(\mathcal{S}_\varphi)$ (the ordering of the variables is just $\prec$ restricted to $\mathcal{V}_\varphi$). Moreover, the construction guarantees that if $\vec{X} \subseteq \mathcal{V}_\varphi$, then the solutions to the equations $T'_{\vec{X} \leftarrow \vec{x}}(\vec{u})$ and $T_{\vec{X} \leftarrow \vec{x}}(\vec{u})$ are the same, when restricted to the variables in $\mathcal{V}_\varphi$. It follows that T' satisfies φ.

The argument in the case that $T \in \mathcal{T}_{\mathrm{uniq}}(\mathcal{S})$ is similar in spirit. For $X \in \mathcal{V}_\varphi$, $\vec{u} \in (\times_{U \in \mathcal{U}} \mathcal{R}(U))$, and $\vec{x} \in (\times_{Y \in \mathcal{V}_\varphi - \{X\}} \mathcal{R}(Y))$, define $F'_X(\vec{u}, \vec{x})$ to be the value of X in the unique solution to the equations in $T_{\mathcal{V}_\varphi - \{X\} \leftarrow \vec{x}}(\vec{u})$.[7] It is again straightforward to check that now $T' = (\mathcal{S}_\varphi, \mathcal{F}') \in \mathcal{T}_{\mathrm{uniq}}(S_\varphi)$ and satisfies φ. ∎

7. This definition is easily seen to agree with the earlier definition of F'_X if $T \in \mathcal{T}_{\mathrm{rec}}$.

Theorem 4.3: *A formula $\varphi \in \mathcal{L}^+(\mathcal{S})$ is satisfiable in $\mathcal{T}(\mathcal{S})$ iff φ is satisfiable in $\mathcal{T}(\mathcal{S}_\varphi^+)$.*

Proof: If $||\mathcal{S}|| \leq ||\mathcal{S}_\varphi||^2 + ||\mathcal{S}_\varphi||$ then the proof is immediate, so suppose that $||\mathcal{S}|| > ||\mathcal{S}_\varphi||^2 + ||\mathcal{S}_\varphi||$ and φ is satisfied in a causal model $T = (\mathcal{S}, \mathcal{F}) \in \mathcal{T}(\mathcal{S})$. Before going on with the proof, it is useful to define some notation. Let $\mathcal{V} = \{X_1, \ldots, X_m\}$, where $\mathcal{V}_\varphi = \{X_1, \ldots, X_k\}$ and $\mathcal{V} - \mathcal{V}_\varphi = \{X_{k+1}, \ldots, X_m\}$. Given a vector $\vec{x} \in \mathcal{R}(X^*) = \times_{X \in \mathcal{V}_\varphi} \mathcal{R}(X)$ and $X_i \in \mathcal{V}_\varphi$, let $\vec{x}_{-i}$ denote the vector excluding the value for X_i. For each $X_i \in \mathcal{V}_\varphi$, choose two values x_{i0} and x_{i1} in $\mathcal{R}(X_i)$. Define $T' = (\mathcal{S}_\varphi, \mathcal{F}')$ by defining $F'_X(\vec{u}, \vec{x}_{-i}, \vec{y}_{-i}, y_i) = x$, where

- $x = y_i$ if $\vec{x}_{-i} = \vec{y}_{-i}$ and $X = y_i$ in some solution to the equations in $T_{\mathcal{V}_\varphi - \{X_i\} \leftarrow \vec{x}_{-i}}(\vec{u})$;
- $x = x_{i0}$ if $y_i \neq x_{i0}$ and either $\vec{x}_{-i} \neq \vec{y}_{-i}$ or there is no solution to the equations in $T_{\mathcal{V}_\varphi - \{X\} \leftarrow \vec{x}_{-i}}(\vec{u})$ in which $X = y_i$;
- $x = x_{i1}$ otherwise.

Finally, define $F_{X^*}(\vec{u}, \vec{x}) = \vec{x}$.

I now show that the construction again guarantees that if $\vec{X} \subseteq \mathcal{V}_\varphi$, then the solutions to the equations $T'_{\vec{X} \leftarrow \vec{x}}(\vec{u})$ and $T_{\vec{X} \leftarrow \vec{x}}(\vec{u})$ are the same, when restricted to the variables in $\mathcal{V}_\varphi$. First suppose that $(\vec{y}, \vec{z})$ is a solution to the equations in $T_{\vec{X} \leftarrow \vec{x}}(\vec{u})$, where $\vec{y} \in \mathcal{R}(X^*)$ and $\vec{z} \in \times_{Y \in \mathcal{V} - \mathcal{V}_\varphi} \mathcal{R}(Y)$. It must be the case that $\vec{x}$ and $\vec{y}$ agree on the variables in $\vec{X}$, so $(\vec{y}, \vec{z})$ is also a solution of the equations in $T_{\mathcal{V}_\varphi - \{X_i\} \leftarrow \vec{y}_{-i}}(\vec{u})$ if $X_i \in \mathcal{V}_\varphi - \vec{X}$. Thus, $F'_{X_i}(\vec{u}, \vec{y}_{-i}, \vec{y}) = y_i$. It follows that $(\vec{y}, \vec{y})$ is a solution to the equations in $T'_{\vec{X} \leftarrow \vec{x}}(\vec{u})$.

Conversely, suppose that $(\vec{y}, \vec{y}')$ is a solution to the equations in $T'_{\vec{X} \leftarrow \vec{x}}(\vec{u})$. Then the definition of F_{X^*} guarantees that $\vec{y} = \vec{y}'$. Moreover, since $\vec{x}$ and $\vec{y}$ agree on the variables in $\vec{X}$, $(\vec{y}, \vec{y})$ must also be a solution to the equations in $T'_{\mathcal{V}_\varphi - \{X_1\} \leftarrow \vec{y}_{-1}}(\vec{u})$. Thus, $F'_{X_1}(\vec{u}, \vec{y}_{-1}, \vec{y}) = y_1$, which means that there must be some vector $\vec{z}$ of values for the variables in $\mathcal{V} - \mathcal{V}_\varphi$ such that $(\vec{y}, \vec{z})$ is a solution to the equations in $T_{\mathcal{V}_\varphi - \{X_1\} \leftarrow \vec{y}_{-1}}(\vec{u})$. But then it is easy to check that $(\vec{y}, \vec{z})$ must in fact be a solution to the equations in $T_{\mathcal{V}_\varphi - \{X_i\} \leftarrow \vec{y}_{-i}}(\vec{u})$ for all $i = 1, \ldots, k$. It follows that $(\vec{y}, \vec{z})$ is a solution to the equations in $T_{\vec{X} \leftarrow \vec{x}}(\vec{u})$, as desired. This suffices to prove this direction of the theorem.

Now suppose that φ is satisfied in a causal model $T = (\mathcal{S}_\varphi^+, \mathcal{F}) \in \mathcal{T}(\mathcal{S}_\varphi^+)$. Since $||\mathcal{S}|| > ||\mathcal{S}_\varphi||^2 + ||\mathcal{S}_\varphi||$, there must be an injective function $f : \mathcal{R}(X^*) \rightarrow \times_{Y \in \mathcal{V} - \mathcal{V}_\varphi} \mathcal{R}(Y)$ and two distinct vectors $\vec{y}_0 = (y_{01}, \ldots, y_{0k})$, $\vec{y}_1 = (y_{11}, \ldots, y_{1k})$ that are not in the range of f. Choose two distinct vectors $\vec{x}_0 = (x_{10}, \ldots, x_{k0})$, $\vec{x}_1 = (x_{11}, \ldots, x_{k1}) \in \mathcal{R}(X^*)$. Define $T' = (\mathcal{S}, \mathcal{F}') \in \mathcal{T}(\mathcal{S})$ as follows. If $X_i \in \mathcal{V}_\varphi$, $\vec{x}_{-i} \in \times_{Y \in \mathcal{V}_\varphi - \{X_i\}} \mathcal{R}(Y)$, $\vec{z} \in \mathcal{R}(X^*)$, and $\vec{y} \times_{Y \in \mathcal{V} - \mathcal{V}_\varphi} \mathcal{R}(Y)$, let

$$F'_{X_i}(\vec{x}_{-i}, \vec{y}) = \begin{cases} F_{X_i}(\vec{x}_{-i}, \vec{z}) & \text{if } f(\vec{z}) = \vec{y}, \\ x_{0i} & \text{if } \vec{y} \text{ is not in the range of } f,\ \vec{y} \neq \vec{y}_1, \\ x_{1i} & \text{otherwise.} \end{cases}$$

If $X_j \in \mathcal{V} - \mathcal{V}_\varphi$, $\vec{x} \in \mathcal{R}(X^*)$ and $\vec{y}_{-j} \in \times_{Y \in \mathcal{V} - \mathcal{V}_\varphi - \{X_j\}} \mathcal{R}(Y)$, then let

$$F'_{X_j}(\vec{x}, \vec{y}_{-j}) = \begin{cases} y & \text{if } f(F_{X^*}(\vec{x})) = (\vec{y}_{-j}, y), \\ y_{0j} & \text{if } f(F_{X^*}(\vec{x})) \neq (\vec{y}_{-j}, y') \text{ for all } y' \in \mathcal{R}(X_j),\ \vec{x} \neq \vec{x_0}, \\ y_{1j} & \text{otherwise.} \end{cases}$$

Again, I show that the construction guarantees that if $\vec{X} \subseteq \mathcal{V}_\varphi$, then the solutions to the equations $T'_{\vec{X}\leftarrow\vec{x}}(\vec{u})$ and $T_{\vec{X}\leftarrow\vec{x}}(\vec{u})$ are the same, when restricted to the variables in $\mathcal{V}_\varphi$. First suppose that $(\vec{y},\vec{z})$ is a solution to the equations in $T_{\vec{X}\leftarrow\vec{x}}(\vec{u})$, where $\vec{y},\vec{z} \in \mathcal{R}(X^*)$. It is easy to check that $(\vec{y}, f(\vec{z}))$ is a solution to the equations in $T'_{\vec{X}\leftarrow\vec{x}}(\vec{u})$. Conversely, suppose that $(\vec{y},\vec{z})$ is a solution to the equations in $T'_{\vec{X}\leftarrow\vec{x}}(\vec{u})$, where $\vec{y} \in \mathcal{R}(X^*)$ and $\vec{z} \in \times_{Y\in\mathcal{V}-\mathcal{V}_\varphi}\mathcal{R}(Y)$. I claim that we must have $\vec{z} = f(F_{X^*}(\vec{y}))$. If, in fact, this is the case, then it is easy to check that $(\vec{y}, F_{X^*}(\vec{y})$ is a solution to the equations in $T_{\vec{X}\leftarrow\vec{x}}(\vec{u})$. On the other hand, if $\vec{z} \neq f(F_{X^*}(\vec{y}))$, then the definition of F'_{X_j} for $X_j \in \mathcal{V} - \mathcal{V}_\varphi$ guarantees that $\vec{z} = \vec{y}_0$ unless $\vec{y} = \vec{x}_0$; if $\vec{y} = \vec{x}_0$, then $\vec{z} = \vec{y}_1$. But the definition of F_{X_i} for $X_i \in \mathcal{V}_\varphi$ guarantees that if $\vec{z} = \vec{y}_0$, then $\vec{y} = \vec{x}_0$: otherwise, $\vec{y} = \vec{x}_1$. Thus, $(\vec{y},\vec{z})$ is a solution iff $\vec{z} = f(F_{X^*}(\vec{y}))$. This suffices to prove the result. ∎

Theorem 4.4: *Given as input a pair $(\varphi, \mathcal{S})$, where $\varphi \in \mathcal{L}^+(\mathcal{S})$ (resp., $\mathcal{L}_{\rm uniq}(\mathcal{S})$) and $\mathcal{S}$ is a finite signature, the problem of deciding if φ is satisfiable with respect to $\mathcal{T}_{\rm rec}(\mathcal{S})$ (resp., $\mathcal{T}_{\rm uniq}(\mathcal{S})$, $\mathcal{T}(\mathcal{S})$) is NP-complete (resp., NP-hard) in $|\varphi|$; if $\varphi \in \mathcal{L}_{\rm GP}(\mathcal{S})$, then the problem of deciding if φ is satisfiable in $\mathcal{T}_{\rm rec}(\mathcal{S})$ (resp., $\mathcal{T}_{\rm uniq}(\mathcal{S})$) is NP-complete (resp., NP-hard).*

Proof: The NP-lower bound is easy for $\mathcal{L}^+(\mathcal{S})$ and $\mathcal{L}_{\rm uniq}(\mathcal{S})$, since there is an obvious way to encode the satisfiability problem for propositional logic into the satisfiability problem for $\mathcal{L}^+$ and $\mathcal{L}_{\rm uniq}$. Given a propositional formula φ with primitive propositions $p_1, \ldots, p_k$, let $\mathcal{S} = (\emptyset, \{X_1, \ldots, X_k\}, \mathcal{R})$, where $\mathcal{R}(X_i) = \{0, 1\}$ for $i = 1, \ldots, k$. Replace each occurrence of the primitive proposition p_i in φ with the formula $X_i = 1$. This gives us a formula φ' in $\mathcal{L}_{\rm uniq}(\mathcal{S})$. It is easy to see that if φ' is satisfiable in a causal model $T \in \mathcal{T}(\mathcal{S})$ (and, *a fortiori* if φ' is satisfiable in a causal model T in either $\mathcal{T}_{\rm rec}(\mathcal{S})$ or $\mathcal{T}_{\rm uniq}(\mathcal{S})$) then the solution to the equations in T defines a satisfying assignment for φ. Conversely, if φ is satisfiable, say by some truth assignment v, then we can trivially construct a causal model $T \in \mathcal{T}_{\rm rec}(\mathcal{S})$ such that $F_{X_i} = v(p_i)$. (For simplicity, I assume that valuations assign values 0 and 1 rather than **false** and **true**.)

This trivial construction of φ' will not work for $\mathcal{L}_{\rm GP}(\mathcal{S})$, since we do not have disjunctions or negations available. The lack of negations does not cause a problem. We can assume without loss of generality that the negations occur only in front of primitive propositions, and we can capture $\neg p_i$ by the formula $X_i = 0$. The idea for dealing with disjunctions is that a formula such as $p_1 \vee \neg p_2 \vee p_3$ is translated to $[X_1 \leftarrow 0; X_2 \leftarrow 1; X_3 \leftarrow 1](Y = 0)$, where Y is a fresh variable. Essentially, we are viewing $p_1 \vee \neg p_2 \vee p_3$ as $(\neg p_1 \wedge p_2 \wedge \neg p_3) \Rightarrow$ *false*, which is why we write, for example, $X_1 \leftarrow 0$ even though p_1 appears positively in the disjunction.

To make matters simpler, assume that φ is a formula in 3-CNF. This suffices for NP-hardness, since the satisfiability problem for 3-CNF formulas is also NP-hard (Garey & Johnson, 1979). Suppose φ is of the form $c_1 \wedge \ldots \wedge c_m$, where each c_l is a *clause* consisting of a disjunction of three primitive propositions and their negations. Suppose that the primitive propositions that appear in φ are $p_1, \ldots, p_k$. Let $\mathcal{S} = (\emptyset, \{X_1, \ldots, X_k, Y_1, \ldots, Y_m\}, \mathcal{R})$, where $\mathcal{R}(X_i) = \mathcal{R}(Y_j) = \{0, 1\}$ for all i, j. Suppose that c_j, the jth clause of φ, is of the form $q_{j1} \vee q_{j2} \vee q_{j3}$, where q_{ji} is either p_{j_i} or $\neg p_{j_i}$ for some j_i. Let c_j^t be the $\mathcal{L}_{\rm GP}$ formula

$$[X_{j_1} \leftarrow x_{j1}; X_{j_2} \leftarrow x_{j2}; X_{j_3} \leftarrow x_{j3}](Y_j = 0),$$

where x_{jh} is 0 if q_{jh} is p_{j_h} and x_{jh} is 1 if q_{jh} is $\neg p_{j_h}$ for $h = 1, 2, 3$. Let φ' be

$$[true](Y_1 = 1 \wedge \ldots \wedge Y_m = 1) \wedge c_1^t \wedge \ldots \wedge c_m^t.$$

I claim that φ is a satisfiable propositional formula iff the $\mathcal{L}_{\rm GP}$ formula φ' is satisfiable in $\mathcal{T}_{\rm rec}(\mathcal{S})$ (resp. $\mathcal{T}_{\rm uniq}(\mathcal{S})$). First suppose that φ' is satisfiable, say in some model $T \in \mathcal{T}_{\rm uniq}(\mathcal{S})$. (If this direction holds for $T \in \mathcal{T}_{\rm uniq}(\mathcal{S})$, it clearly holds *a fortiori* for $T \in \mathcal{T}_{\rm rec}(\mathcal{S})$.) Let $\vec{z}$ be the unique solution to the equations in T. By construction, the Y_j component of $\vec{z}$ is 1 for $j = 1, \ldots, m$. Let x_i^* be the value of the X_i component in $\vec{z}$. Consider the valuation v such that $v(p_i) = x_i^*$. I claim that $v(\varphi) = 1$. To see this, suppose that clause c_j is $q_{j1} \vee q_{j2} \vee q_{j3}$. If v makes q_{j1}, q_{j2}, and q_{j3} false, then we must have $x_{jh} = x_{j_h}^*$ for $h = 1, 2, 3$. Since $T \models [X_{j_1} \leftarrow x_{j1}; X_{j_2} \leftarrow x_{j2}; X_{j_3} = x_{j3})](Y_j = 0)$ and the value of the X_{j_h} component of $\vec{z}$ is x_{jh} for $h = 1, 2, 3$, it follows that $\vec{z}$ is a solution to the equations in $T_{X_{j_1} \leftarrow x_{j1}; X_{j_2} \leftarrow x_{j2}; X_{j_3} \leftarrow x_{j3}}$. But this contradicts the fact that $T \models [X_{j_1} \leftarrow x_{j1}; X_{j_2} \leftarrow x_{j2}; X_{j_3} \leftarrow x_{j3}](Y_j = 0)$ (since the Y_j component of $\vec{z}$ is 1). It follows that $v(c_j) = v(q_{j1} \vee q_{j2} \vee q_{j3}) = 1$. Since this is true for all clauses c_j, we must have that $v(\varphi) = 1$.

For the converse, suppose that φ is satisfiable, say by valuation v. I show that φ' is satisfiable in $T \in \mathcal{T}_{\rm rec}(\mathcal{S})$. Order the variables so that $X_{j_1}, X_{j_2}, X_{j_3} \prec Y_j$. (There are many orderings of the variables that satisfy these constraints; any one will do.) Define $F_{X_i} = v(p_i)$ (so that F_{X_i} is a constant, independent of its arguments); define $F_{Y_j}(x_{j_1}, x_{j_2}, x_{j_3}) = 1$ if $(x_{j_1}, x_{j_2}, x_{j_3}) = (v(p_{j_1}), v(p_{j_2}), v(p_{j_3}))$ and 0 otherwise. It is easy to check that $T \models \varphi'$, as desired.

For the NP upper bound in the case of $\mathcal{T}_{\rm rec}(\mathcal{S})$, it clearly suffices to deal with $\varphi \in \mathcal{L}^+$. Suppose we are given $(\varphi, \mathcal{S})$ with $\varphi \in \mathcal{L}^+$. We want to check if φ is satisfiable in $\mathcal{T}_{\rm rec}(\mathcal{S})$. The basic idea in to guess a causal model T and verify that it indeed satisfies φ. There is a problem with this though. To completely describe a model T, we need to describe the functions F_X. However, there may be many variables X in $\mathcal{S}$ and they can have many possible inputs. Just describing these functions may take time much longer than polynomial in φ. Part of the solution to this problem is provided by Theorem 4.2, which tells us that it suffices to check whether φ is satisfiable in $\mathcal{T}_{\rm rec}(\mathcal{S}_\varphi)$. In light of this, for the remainder of this part of the proof, I assume without loss of generality that $\mathcal{S} = \mathcal{S}_\varphi$. This limits the number of variables that we must consider to $O(|\varphi|)$. But even this does not solve our problem completely. Since we are not given any bounds on $|\mathcal{R}(Y)|$ for variables Y in $\mathcal{S}_\varphi$, even describing the functions F_Y for the variables Y that appear in φ on all their possible input vectors could take time much more than polynomial in φ. The solution is to give only a short partial description of a model T and show that this suffices.

Consider all pairs $(\vec{Y} \leftarrow \vec{y}, \vec{u})$ such that there is a subformula of φ of the form $[\vec{Y} \leftarrow \vec{y}]\psi$ and $\vec{u}$ appears in ψ. Let R be the set of all such pairs. Note that $|R| < |\varphi|^2$. We say that two causal models T and T' in $\mathcal{T}_{\rm rec}(\mathcal{S})$ *agree on* R if, for all pairs $(\vec{Y} \leftarrow \vec{y}, \vec{u}) \in R$, the (unique) solutions to the equations in $T_{\vec{Y} \leftarrow \vec{y}}(\vec{u})$ and $T'_{\vec{Y} \leftarrow \vec{y}}(\vec{u})$ are the same. It is easy to see that if T and T' agree on R, then either both T and T' satisfy φ or neither do. That is, all we need to know about a causal model is how it deals with the *relevant equations*—those corresponding to pairs in R.

For each pair $(\vec{Y} \leftarrow \vec{y}, \vec{u}) \in R$, guess a vector $\vec{v}(\vec{Y} \leftarrow \vec{y}, \vec{u})$ of values for the endogenous variables; intuitively, these are the unique solutions to the relevant equations in a model satisfying T. Given these guesses, it is easy to check if φ is satisfied in a model where these guesses do indeed represent the solutions to the relevant equations. It remains to show that there exists a causal model in $\mathcal{T}_{\mathrm{rec}}(\mathcal{S})$ where the relevant equations have these solutions.

To do this, first guess an ordering $\prec$ on the variables. We can then verify, for each fixed $\vec{u}$ that appears in φ, whether the solution vectors $\vec{v}(\vec{Y} \leftarrow \vec{y}, \vec{u})$ guessed for the relevant equations are *compatible with* $\prec$, in the sense that it is not the case that there are two solutions $(\vec{u}, \vec{x})$ and $(\vec{u}, \vec{x}')$ such that some variable X takes on different values in $\vec{x}$ and $\vec{x}'$, but all variables Y such that $Y \prec X$ take on the same values in $\vec{x}$ and $\vec{x}'$. It is easy to see that if the solutions are compatible with $\prec$, we can define the functions F_X for $X \in \mathcal{V}$ such that all the equations hold and F_X is independent of the values of Y if $X \prec Y$ for all $X, Y \in \mathcal{V}$. (Note we never actually have to write out the functions F_X, which may take too long; we just have to know they exist.) To summarize, as long as we can guess some solutions to the relevant equations such that a causal model that has these solutions satisfies φ, and an ordering $\prec$ such that these solutions are compatible with $\prec$, then φ is satisfiable in $\mathcal{T}_{\mathrm{rec}}(\mathcal{S})$. Conversely, if φ is satisfiable in $T \in \mathcal{T}_{\mathrm{rec}}(\mathcal{S})$, then there clearly are solutions to the relevant equations that satisfy φ and an ordering $\prec$ such that these solutions are compatible with $\prec$. (We just take the solutions and the ordering $\prec$ from T.) This shows that the satisfiability problem for $\mathcal{T}_{\mathrm{rec}}$ is in NP, as desired. ∎

References

Chajewska, U., & Halpern, J. Y. (1997). Defining explanation in probabilistic systems. In *Proc. Thirteenth Conference on Uncertainty in Artificial Intelligence (UAI '97)*, pp. 62–71.

Druzdzel, M. J., & Simon, H. A. (1993). Causality in bayesian belief networks. In *Uncertainty in Artificial Intelligence 9*, pp. 3–11.

Galles, D., & Pearl, J. (1997). Axioms of causal relevance. *Artificial Intelligence*, *97*(1–2), 9–43.

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

Garey, M., & Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-completeness.* W. Freeman and Co., San Francisco, Calif.

Goldberger, A. S. (1972). Structural equation methods in the social sciences. *Econometrica*, *40*(6), 979–1001.

Harel, D. (1979). *First-Order Dynamic Logic.* Lecture Notes in Computer Science, Vol. 68. Springer-Verlag, Berlin/New York.

Heckerman, D., & Shachter, R. (1995). Decision-theoretic foundations for causal reasoning. *Journal of Artificial Intelligence Research*, *3*, 405–430.

Henrion, M., & Druzdzel, M. J. (1990). Qualitative propagation and scenario-based approaches to explanation of probabilistic reasoning. In *Uncertainty in Artificial Intelligence 6*, pp. 17–32.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, *82*(4), 669–710.

Pearl, J. (1999). *Causality*. Cambridge University Press, New York. Forthcoming.

Pearl, J., & Verma, T. (1991). A theory of inferred causation. In *Principles of Knowledge Representation and Reasoning: Proc. Second International Conference (KR '91)*, pp. 441–452.

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, Prediction, and Search*. Springer-Verlag, New York.

Strotz, R. H., & Wold, H. O. A. (1960). Recursive vs. nonrecursive systems: an attempt at synthesis. *Econometrica*, *28*(2), 417–427.