In: Proc. Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-01), 2001.

Complexity Results for Structure-Based Causality

Thomas Eiter and **Thomas Lukasiewicz**
Institut und Ludwig Wittgenstein Labor für Informationssysteme, TU Wien
Favoritenstraße 9–11, A-1040 Wien, Austria
{eiter, lukasiewicz}@kr.tuwien.ac.at

Abstract

We analyze the computational complexity of causal relationships in Pearl's structural models, where we focus on causality between variables, event causality, and probabilistic causality. In particular, we analyze the complexity of the sophisticated notions of weak and actual causality by Halpern and Pearl. In the course of this, we also prove an open conjecture by Halpern and Pearl, and establish other semantic results. To our knowledge, no complexity aspects of causal relationships have been considered so far, and our results shed light on this issue.

1 Introduction

Representing and reasoning with causal knowledge has received much attention in the recent decade. The existing approaches to causality in the AI literature can be roughly divided into those that have been developed as modal nonmonotonic logics (especially in the context of logic programming) and those that evolved from the area of Bayesian networks.

A representative of the former is Geffner's modal nonmonotonic logic for handling causal knowledge [1990; 1992], which has been inspired by default reasoning from conditional knowledge bases. Other more specialized formalisms play an important role in dealing with causal knowledge about actions and change; see especially the work by Turner [1999] and the references therein for an overview.

A representative of the latter is Pearl's approach to modeling causality by structural equations [Balke and Pearl, 1994; Galles and Pearl, 1997; Pearl, 1999; 2000], which is central to a number of recent research efforts. In particular, the evaluation of deterministic and probabilistic counterfactuals has been explored, which is at the core of problems in fault diagnosis, planning, decision making, and determination of liability [Balke and Pearl, 1994].

In a recent paper, Halpern [2000] gave an axiomatization of reasoning about causal formulas in the structural-model approach, and explored its computational aspects.

It has been shown that the structural-model approach allows a precise modeling of many important causal relationships, which can especially be used in natural language processing [Galles and Pearl, 1997]. In particular, it allows an elegant definition of the important notions of actual causation and explanation [Halpern and Pearl, 2000; 2001].

We give a simple example due to Halpern and Pearl [2000], which illustrates the structural-model approach.

Example 1.1 Suppose that two arsonists lit matches in different parts of a dry forest, and both cause trees to start burning. Assume now that either match by itself suffices to burn down the whole forest. In the structural-model framework, such a scenario may be modeled as follows. We assume two binary background variables U_1 and U_2, which determine the motivation and the state of mind of the two arsonists, where U_i has the value 1 iff the arsonist i intends to start a fire. Moreover, we have three binary variables A_1, A_2, and F, which describe the observable situation, where A_i has the value 1 iff the arsonist i drops the match, and F has the value 1 iff the whole forest burns down. The causal dependencies between these variables are expressed through functions, which say that the value of A_i is given by the value of U_i, and that F has the value 1 iff either A_1 or A_2 has the value 1. These dependencies can be graphically represented as in Fig. 1.

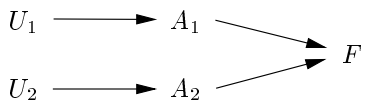

Figure 1: Causal Graph

While the semantic aspects of causal relationships in the structural-model approach have been explored in depth (see especially the work by Pearl [2000]), studies about their computational properties are missing so far. In this paper, we try to fill this gap by giving a precise account of the complexity of deciding causal relationships in structural models.

Note that Halpern's work [2000] is orthogonal to ours, as it focuses on the computational aspects of deciding whether a given causal formula has a causal model, while our work in this paper deals with the complexity of deciding whether a given causal relationship holds in a given causal model.

The main contributions can be summarized as follows:

- We analyze the complexity of deciding causal relationships between variables in structural causal models. We consider the notions of causal irrelevance, cause, cause in a context, direct cause, and indirect cause. It turns out that testing

these notions has a complexity among NP, co-NP, and D^{P}. Hardness holds even in restricted cases.

- We analyze the complexity of deciding causal relationships between events. We consider the notions of necessary and possible causality, and the sophisticated notions of weak and actual causality by Halpern and Pearl [2000]. In particular, checking the latter is shown to be Σ_2^P-complete in general, and NP-complete in the case of binary variables.
- We prove some semantic results related to the notions of actual and weak causality. More precisely, we prove an open conjecture by Halpern and Pearl [2000], and we give a new characterization of weak causality for the binary case.
- As a representative for probabilistic causal relationships, we finally analyze the complexity of deciding the notion of probabilistic causal irrelevance, which is shown to be complete for the class $\mathbf{C}_{=}$, and thus harder than co-NP. Note that few $\mathbf{C}_{=}$-complete problems, and none in AI, are known.

Our results draw a clear picture of the complexity of structural causality, and give useful insight for exploiting it, e.g., in counterfactual reasoning (see Section 6).

For space reasons, we give only proof sketches of some results. Proofs of all results are given in the extended paper [Eiter and Lukasiewicz, 2001].

2 Preliminaries

We assume a finite set of *random variables*. Each variable X_i may take on *values* from a finite *domain* $D(X_i)$. A *value* for a set of variables $X = \{X_1, \ldots, X_n\}$ is a mapping $x\colon X \to D(X_1) \cup \cdots \cup D(X_n)$ such that $x(X_i) \in D(X_i)$ (where the empty mapping $\emptyset$ is the unique value for $X = \emptyset$). The *domain* of X, denoted $D(X)$, is the set of all values for X. For $Y \subseteq X$ and $x \in D(X)$, denote by $x|Y$ the restriction of x to Y. For sets of variables X, Y and values $x \in D(X)$ and $y \in D(Y)$, we use xy to denote the union of x and y. As usual, we often identify singletons $\{X_i\}$ with X_i, and their values x with $x(X_i)$.

2.1 Causal and Probabilistic Causal Models

A *causal model* M is a triple (U, V, F), where U is a finite set of *exogenous* variables, V is a finite set of *endogenous* variables such that $U \cap V = \emptyset$, and $F = \{F_X \mid X \in V\}$ is a set of functions $F_X\colon D(PA_X) \to D(X)$ that assign a value to X for each value of the *parents* $PA_X \subseteq U \cup V \setminus \{X\}$ of X.

We focus here on the principal class of *recursive* causal models $M = (U, V, F)$ (as argued in [Halpern and Pearl, 2000], we do not lose much generality by concentrating on recursive causal models) in which a total ordering $\prec$ on V exists such that $Y \in PA_X$ implies $Y \prec X$, for all $X, Y \in V$. In such models, every assignment to the exogenous variables $U = u$ determines a unique value y for every set of endogenous variables $Y \subseteq V$, denoted $Y_M(u)$ (or simply $Y(u)$). In the sequel, M is reserved for denoting a recursive causal model.

For causal models $M = (U, V, F)$, $X \subseteq V$, and $x \in D(X)$, the causal model $M_{X=x} = (U, V, F_{X=x})$, where $F_{X=x} = \{F_Y \mid Y \in V \setminus X\} \cup \{F_{X'} = x(X') \mid X' \in X\}$, is a *submodel* of M. Intuitively, X is set to x by an "external action". We abbreviate $M_{X=x}$ and $F_{X=x}$ by M_x and F_x, respectively. For $Y \subseteq V$, we abbreviate $Y_{M_x}(u)$ by $Y_x(u)$.

A *probabilistic causal model* (M, P) consists of a causal model $M = (U, V, F)$ and a probability function P on $D(U)$.

Example 2.1 In our introductory example, the causal model $M = (U, V, F)$ is given by $U = \{U_1, U_2\}$, $V = \{A_1, A_2, F\}$, and $F = \{F_{A_1}, F_{A_2}, F_F\}$, where $F_{A_1} = U_1$, $F_{A_2} = U_2$, and $F_F = 1$ iff $A_1 = 1$ or $A_2 = 1$ (Fig. 1 shows the parent relationship between the variables). In a probabilistic causal model (M, P), we may use the uniform distribution P over $D(U)$.

2.2 Model Representation for Computation

We assume the following representation of causal models $M = (U, V, F)$ and probabilistic causal models (M, P) (see the full paper for a discussion of these assumptions):

- The domain $D(X)$ of each variable $X \in U \cup V$ is explicit, i.e., $D(X) = \{v_1, \ldots, v_k\}$ is enumerated.
- Each function $F_X\colon D(PA_X) \to D(X)$, $X \in V$, is computable in polynomial time.
- P is given by a pair (f, b), where $f\colon D(U) \to \{0,1,2,\ldots\}$ is a polynomial-time computable function and $b > 0$ is an integer, such that $P(u) = f(u) \,/\, b$ for every $u \in D(U)$.

The following proposition is immediate.

Proposition 2.2 *For every $X, Y \subseteq V$ and $x \in D(X)$, the values $Y(u)$ and $Y_x(u)$, given $u \in D(U)$, are computable in polynomial time.*

We say M (resp., (M, P)) is *binary*, if $|D(X)| = 2$ for all $X \in V$. Furthermore, M (resp., (M, P)) is *bounded*, if $|PA_X| \leq k$ holds for each $X \in V$, i.e., X has at most k parents, where k is an arbitrary but fixed constant.

Note that in a bounded model, as for polynomiality, each F_X is w.l.o.g. given by a table that lists the output values $F_X(v)$ for all $v \in D(PA_X)$ (which is similar to a conditional probability table in Bayesian networks).

2.3 Complexity Classes

Complexity classes that we encounter are shown in Fig. 2, where arrows denote containment. P, NP, co-NP, Σ_2^P, and Π_2^P are from the Polynomial Hierarchy (PH); $\mathrm{D}^{\mathrm{P}} = \{L \cap L' \mid L \in \mathrm{NP}, L' \in \text{co-NP}\}$ is the "conjunction" of NP and co-NP. The class $\mathbf{C}$ is from the Counting Hierarchy (CH) of complexity classes [Torán, 1991]. Informally, $\mathbf{C}$ contains all problems which can be expressed as deciding whether a given instance I has at least $f(|I|)$ many polynomial size "proofs" $J_1, \ldots, J_{f(|I|)}$ that I is a Yes-instance, for some function f, where computing $f(|I|)$ and checking each proof J_i is polynomial. The class $\mathbf{C}$ coincides with the famous class PP (probabilistic P) [Papadimitriou, 1994], which contains the problems decidable by a polynomial-time Turing machine which accepts an input iff the majority of its runs halt in an accepting state.

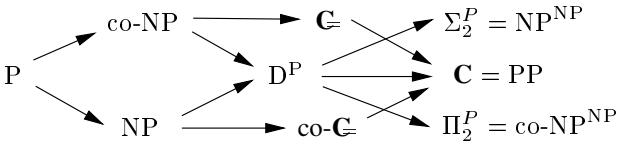

Figure 2: Containment between Complexity Classes

The class $\mathbf{C}^{=}$ is a variant of $\mathbf{C}$, where "exactly $f(|I|)$" replaces "at least $f(|I|)$". While this difference seems marginal, $\mathbf{C}$ and $\mathbf{C}^{=}$ have quite different properties [Torán, 1991]. Intuitively, $\mathbf{C}^{=}$ is an extension of co-NP, and has many properties of this class. $\mathbf{C}$ and $\mathbf{C}^{=}$ are contained in PSPACE, and it is believed that they are not contained in PH.

3 Causality between Variables

We first focus on causal relationships between variables due to Galles and Pearl [1997]; see also [Pearl, 1999; 2000].

3.1 Definitions

In the sequel, we assume a causal model $M = (U, V, F)$ and sets of variables $X, Y, Z \subseteq V$ with $X, Y \neq \emptyset$. We say

- *X is causally irrelevant to Y given Z* iff for every $W \subseteq V \setminus X \cup Y \cup Z$, $u \in D(U)$, $x, x' \in D(X)$, $z \in D(Z)$, and $w \in D(W)$, we have $Y_{xzw}(u) = Y_{x'zw}(u)$.
- *X is a cause of Y* iff values $x, x' \in D(X)$ and $u \in D(U)$ exist such that $Y_x(u) \neq Y_{x'}(u)$.
- *X is a cause of Y in the context $Z = z$* iff values $x, x' \in D(X)$ and $u \in D(U)$ exist with $Y_{xz}(u) \neq Y_{x'z}(u)$.
- *X is a direct cause* of Y iff values $x, x' \in D(X)$, $u \in D(U)$, and $z \in D(V \setminus X \cup Y)$ exist with $Y_{xz}(u) \neq Y_{x'z}(u)$.
- *X is an indirect cause of Y* iff X is a cause of Y and X is not a direct cause of Y.

We give some examples of such causal relationships. For a more detailed discussion of these concepts see [Galles and Pearl, 1997] and [Pearl, 1999; 2000].

Example 3.1 In our running example, A_1 is not causally irrelevant to F. For instance, if we set $A_2 \in V \setminus \{A_1, F\}$ to 0 and A_1 to 0 and 1, then F has the values 0 and 1, respectively. Informally, the actions of arsonist 1 are not causally irrelevant to the state of the forest. In fact, A_1 is a cause of F, but not a cause of F in the context $A_2 = 1$. Informally, the actions of arsonist 1 are in general a cause of the state of the forest, but not when arsonist 2 starts a fire. Finally, it is easy to verify that A_1 is in fact a direct cause of F.

3.2 Results

Our complexity results for checking the above notions of causality are summarized in Table 1. It is important to point out that for all these causal relationships, hardness holds even if M is binary and bounded, and X is a singleton.

Table 1: Complexity of Causality between Variables

X is causally irrelevant to Y given Z	co-NP-complete
X is a cause of Y	NP-complete
X is a cause of Y in the context $Z = z$	NP-complete
X is a direct cause of Y	NP-complete
X is an indirect cause of Y	D^{P}-complete

Our first result shows that deciding causal irrelevance is co-NP-complete. We sketch the main ideas of its proof.

Theorem 3.2 *Given $M = (U, V, F)$ and $X, Y, Z \subseteq V$ with $X, Y \neq \emptyset$, deciding whether X is causally irrelevant to Y given Z is* co-NP*-complete. Hardness holds even if (1) M is binary and bounded, (2) Z is empty, (3) X is a singleton, and (4a) Y is a singleton or (4b) $V = X \cup Y \cup Z$.*

Proof (sketch). The problem is in co-NP, as $W \subseteq V \setminus X \cup Y \cup Z$ and values $u \in D(U)$, $x, x' \in D(X)$, $z \in D(Z)$, and $w \in D(W)$ such that $Y_{xzw}(u) \neq Y_{x'zw}(u)$ can be guessed and verified in polynomial time (see Proposition 2.2).

We show co-NP-hardness by a polynomial transformation from the co-NP-complete problem of deciding whether a given propositional formula in 3DNF $\phi = \phi_1 \vee \cdots \vee \phi_k$ on the atoms $A_1, \ldots, A_n$, where w.l.o.g. $k \geq 2$ and $n \geq 1$, is a tautology.

We now construct $M = (U, V, F)$ and $X, Y, Z \subseteq V$ such that X is causally irrelevant to Y given Z iff ϕ is a tautology. We define $X = \{A\}$, $Y = \{B\}$, and $Z = \emptyset$. Moreover, $U = \{A_1, \ldots, A_n\}$ and $V = X \cup Y \cup \{D_1, \ldots, D_{k-1}\}$. Let all variables have the domain $\{0, 1\}$, where 0 and 1 are identified with the truth values **false** and **true**, respectively. The values 0 and 1 of X are denoted by x_0 and x_1, respectively. The functions F_S for $S \in V$ are as follows:

- $F_A = 1$ and $F_{D_1} = A \vee \phi_1$,
- $F_{D_i} = D_{i-1} \vee \phi_i$ for all $i \in \{2, \ldots, k-1\}$,
- $F_B = D_{k-1} \vee \phi_k$.

It can now be shown that X is causally irrelevant to Y given Z iff ϕ is a tautology. Informally, under any $u \in D(U)$, if X is set to x_1, then Y becomes 1. Whereas, if X is set to x_0, then Y is the truth value of ϕ under u. That is, setting X to x_0 and x_1 yields the same value of Y under any $u \in D(U)$ iff ϕ is a tautology. Notice that assignments to some nonempty $W \subseteq D_1, \ldots, D_{k-1}$ always yield the same value of Y.

Note that (1)–(3) and (4a) are satisfied. The proof of hardness in the case where (1)–(3) and (4b) holds is similar. □

The following result shows that deciding causes and causes in a context is NP-complete. Here, NP-hardness can be shown by a reduction from the NP-complete problem of deciding whether a propositional formula is not a tautology.

Theorem 3.3 *Given $M = (U, V, F)$ and $X, Y \subseteq V$ (resp., $X, Y, Z \subseteq V$ and $z \in D(Z)$) with $X, Y \neq \emptyset$, deciding whether X is a cause of Y (resp., X is a cause of Y in the context $Z = z$) is* NP*-complete. Hardness holds even if (1) M is binary and bounded, and (2) X and Y are singletons.*

4 Event Causality

We now focus on causality between events. In particular, we consider causal relationships between events due to Galles and Pearl [1997]; see also [Pearl, 1999; 2000], and the notions of weak and actual causality by Halpern and Pearl [2000], which are inspired by Pearl's causal beams [2000].

4.1 Definitions

A *primitive event* is an expression of the form $Y = y$, where Y is a variable and y is a value for Y. The set of *events* is the

closure of the set of primitive events under the Boolean operations $\neg$ and $\wedge$ (that is, every primitive event is an event, and if ϕ and ψ are events, then also $\neg\phi$ and $\phi \wedge \psi$).

The *truth* of an event ϕ in a causal model $M = (U, V, F)$ under $u \in D(U)$, denoted $(M, u) \models \phi$, is inductively given by:

- $(M, u) \models Y = y$ iff $Y_M(u) = y$,
- $(M, u) \models \neg\phi$ iff $(M, u) \models \phi$ does not hold,
- $(M, u) \models \phi \wedge \psi$ iff $(M, u) \models \phi$ and $(M, u) \models \psi$.

We write $\phi(u)$ to abbreviate $(M, u) \models \phi$. For $X \subseteq V$ and $x \in D(X)$, we write $\phi_x(u)$ to abbreviate $(M_x, u) \models \phi$. For $X = \{X_1, \ldots, X_k\} \subseteq V$ with $k \geq 1$ and $x_i \in D(X_i)$, we use $X = x_1 \cdots x_k$ to abbreviate $X_1 = x_1 \wedge \ldots \wedge X_k = x_k$.

The following proposition is immediate.

Proposition 4.1 *Let $X \subseteq V$ and $x \in D(X)$. Given $u \in D(U)$ and an event ϕ, deciding whether $\phi(u)$ and $\phi_x(u)$ (given x) hold can be done in polynomial time.*

We are now ready to define causal relationships between events. We first define the notions of necessary and possible causality (which are slightly more general than in [Galles and Pearl, 1997]). Let $M = (U, V, F)$ be a causal model. Let $X \subseteq V$ and $x \in D(X)$. Let ϕ be an event. We say

- $X = x$ *always causes* ϕ iff (i) $\phi_x(u)$ for all $u \in D(U)$, and (ii) some $x' \in D(X)$, $u' \in D(U)$ exist with $x' \neq x$ and $\neg\phi_{x'}(u')$.
- $X = x$ *may have caused* ϕ iff (i) $X = x$ and ϕ are observed (which implies that $X(u) = x$ and $\phi(u)$ for some $u \in D(U)$) and (ii) some $x' \in D(X)$ and $u \in D(U)$ exist such that $x' \neq x$, $X(u) = x$, $\phi(u)$, and $\neg\phi_{x'}(u)$.

We next define weak and actual causality. We say $X = x$ is a *weak cause* of ϕ under u iff the following conditions hold:

AC1. $X(u) = x$ and $\phi(u)$.

AC2. Some set of variables $W \subseteq V \setminus X$ and some values $\overline{x} \in D(X)$ and $w \in D(W)$ exist such that:

(a) $\neg\phi_{\overline{x}w}(u)$,

(b) $\phi_{x\,w\,\hat{z}}(u)$ for all $\hat{Z} \subseteq V \setminus (X \cup W)$ and $\hat{z} = \hat{Z}(u)$.

We say $X = x$ is an *actual cause* of ϕ under u iff additionally the following condition is satisfied:

AC3. X is minimal. That is, no proper subset of X satisfies both AC1 and AC2.

We give some illustrative examples.

Example 4.2 In our running example, each event among $A_1 = 1$, $A_2 = 1$, and $A_1 = 1 \wedge A_2 = 1$ always causes $F = 1$. For instance, let us show that $A_1 = 1$ always causes $F = 1$: (i) if A_1 is set to 1, then F has the value 1 under every $u \in D(U)$, and (ii) if U_2 is set to 0, and A_1 to 0, then F has the value 0. Informally, at least one arsonist starting a fire always has the effect that the whole forest burns down.

Consider now the background context $u = (1, 1)$ in which both arsonists intend to start a fire. Then, each among $A_1 = 1$, $A_2 = 1$, and $A_1 = 1 \wedge A_2 = 1$ is a weak cause of $F = 1$. For instance, let us show that $A_1 = 1$ is a weak cause of $F = 1$: (AC1) both A_1 and F have the value 1 under u, (AC2 (a)) if both A_1 and A_2 are set to 0, then F has the value 0, and (AC2 (b)) if A_1 is set to 1 and A_2 to 0, then F has the value 1. In fact, $A_1 = 1$ and $A_2 = 1$ are actual causes of $F = 1$, while $A_1 = 1 \wedge A_2 = 1$ is not an actual cause of $F = 1$.

4.2 Results

Our complexity results for the above causal relationships between events are summarized in Table 2. We distinguish between the general and the binary case, in which we assume a syntactic restriction to binary causal models. We remark that for all these causal relationships between events, hardness holds even if M is bounded and X is a singleton.

Table 2: Complexity of Event Causality

Problem	general case	binary case
$X = x$ always causes ϕ	D^{P}-complete	D^{P}-complete
$X = x$ may cause ϕ	NP-complete	NP-complete
$X = x$ is a weak cause of ϕ	Σ_2^P-complete	NP-complete
$X = x$ is an actual cause of ϕ	Σ_2^P-complete	NP-complete

The following theorem shows that deciding weak causality is Σ_2^P-complete in the general case. We sketch the main ideas behind the technically quite involved proof of this result.

Theorem 4.3 *Given $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, $u \in D(U)$, and an event ϕ, deciding whether $X = x$ is a weak cause of ϕ under u is Σ_2^P-complete. Hardness holds even if X is a singleton, and either M is bounded or ϕ is primitive.*

Proof (sketch). As for membership in Σ_2^P, recall that $X = x$ is a weak cause of ϕ under u iff AC1 and AC2 hold. By Proposition 4.1, verifying $X(u) = x$ and $\phi(u)$ is polynomial. Moreover, some W, $\overline{x}$, and w as in AC2 can be guessed and verified with an NP-oracle (needed for (b)) in polynomial time. In summary, checking AC1 and AC2 is in Σ_2^P.

Σ_2^P-hardness is shown by a reduction from evaluating a quantified Boolean formula $\Phi = \exists A \forall B\, \gamma$, where γ is a propositional formula on the variables $A = \{A_1, \ldots, A_m\}$ and $B = \{B_1, \ldots, B_n\}$. We now define M, etc. as stated such that $X = x$ is a weak cause of ϕ under u iff Φ is true (deciding this is Σ_2^P-complete [Papadimitriou, 1994]).

We define $X = \{G\}$ and $Y = \{H\}$. The exogenous and endogenous variables are $U = \{E\}$ and $V = X \cup Y \cup \{C\} \cup A \cup B$, respectively. Define $D(S) = \{0, 1, 2\}$ for all $S \in B$, and $D(S) = \{0, 1\}$ for all $S \in U \cup V \setminus B$. We define

$$\phi' = (\gamma' \wedge \bigwedge_{S \in B} S \neq 2) \vee (C = 0) \vee (G = 1 \wedge C = 1 \wedge \bigvee_{S \in B} S \neq 2),$$

where γ' is obtained from γ by replacing each $S \in A \cup B$ by "$S = 1$". The functions F_S with $S \in V$ are defined by:

- $F_S = 0$ for all $S \in X \cup \{C\} \cup A$,
- $F_S = G + C$ for all $S \in B$, and $F_H = 1$ iff ϕ' is true.

Let $u \in D(U)$ and $x = 0$. Let ϕ be $Y = 1$. It can now be shown that $X = x$ is a weak cause of ϕ under u iff Φ is true. More precisely, AC1 is trivially satisfied, and AC2 holds iff Φ is true. Roughly speaking, the existential and universal quantification over A and B in Φ is reflected by the variables in W and $V \setminus (X \cup W)$, respectively. We then especially have to ensure that (i) $B \cap W = \emptyset$ and (ii) no truth assignment to the variables in B is ignored. In detail, to make ϕ false in AC2 (a), C must be set to 1 and all $S \in B$ must have

the value 2. Whereas, to make ϕ true in AC2 (b), all $S \in B$ must have a value from $\{0,1\}$. This already ensures (i). Since G is set to 0 in AC2 (b), and $F_S = G + C$ for all $S \in B$, each variable in B has the value 1 in AC2 (b). As every $S \in B$ has the value 0 in M, this then ensures (ii).

Note that X is a singleton and that ϕ is primitive. To show Σ_2^P-hardness for a singleton X and bounded M, define $M' = (U, V \setminus \{H\}, F \setminus \{F_H\})$. Then, $X = x$ is a weak cause of ϕ' under u iff Φ evaluates to true. □

The just sketched proof of Σ_2^P-hardness makes use of *non-binary* causal models. Thus, we may ask whether deciding weak causality in the binary case has a lower complexity.

In fact, the following semantic result shows that in the binary case, AC2 can be expressed in a different way.

Theorem 4.4 *Let $M = (U, V, F)$ be binary. Let $X \subseteq V$ and $x \in D(X)$. Let ϕ be an event. Then, $X = x$ is a weak cause of ϕ under $u \in D(U)$ iff AC1 holds, and (AC2') some $W \subseteq V \setminus X$, $\overline{x} \in D(X)$, and $w \in D(W)$ exist such that (a) $\neg\phi_{\overline{x}w}(u)$, (b) $\phi_{xw}(u)$, and (c) $Z_{xw}(u) = Z(u)$ for $Z = V \setminus (X \cup W)$.*

Proof (sketch). Notice that AC2' (a) is identical to AC2 (a). Moreover, AC2' (b) can be replaced by AC2 (b), as AC2' (c) implies that setting $\hat{Z} \subseteq V \setminus (X \cup W)$ to $\hat{z} = \hat{Z}(u)$ is immaterial in $\phi_{xw\hat{z}}(u)$ of AC2 (b). Thus, it is now sufficient to show that for binary M, we can add AC2' (c) to AC2 (a) and (b).

Roughly speaking, we can additionally satisfy AC2' (c) by iteratively moving variables from $V \setminus (X \cup W)$ to the W-part in AC2 (a) and (b). More precisely, any singleton $S \subseteq V \setminus (X \cup W)$ with $S_{xw}(u) \neq S(u)$ can be moved to the W-part assigning them $S_{\overline{x}w}(u)$. This is always feasible for $S_{\overline{x}w}(u) = S_{xw}(u)$. If M is binary, this is also feasible for $S_{\overline{x}w}(u) \neq S_{xw}(u)$, which then implies $S_{\overline{x}w}(u) = S(u)$. This construction can now be iterated until AC2' (c) holds. □

Based on this result, it can now be shown that deciding weak causality in the binary case is NP-complete. This is more formally expressed by the following theorem.

Theorem 4.5 *Given a binary $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, $u \in D(U)$, and an event ϕ, deciding whether $X = x$ is a weak cause of ϕ under u is* NP*-complete. Hardness holds even if M is bounded, X is a singleton, and ϕ is primitive.*

We next focus on the problem of deciding actual causality. We first prove a conjecture by Halpern and Pearl [2000], which says that only primitive events can be actual causes. Note that the proof also goes through for their setting of possibly infinite domains and sets of endogenous variables.

Theorem 4.6 *Let $M = (U, V, F)$ be a causal model. Let $X \subseteq V$ and $x \in D(X)$. Let ϕ be an event. If $X = x$ is an* actual cause *of ϕ under u, then X is a singleton.*

Proof (sketch). We give a proof by contradiction. Let $X = x$ be an *actual cause* of ϕ under u. That is, AC1–AC3 hold. In particular, AC2 (a) and (b) hold for some $W \subseteq V \setminus X$ and some $\overline{x} \in D(X)$ and $w \in D(W)$. Suppose that X is not a singleton. We consider two cases:

Case 1: There exists some nonempty $X'' \subseteq X$, $X'' \neq X$, such that $\phi_{x'\overline{x}''w\hat{z}}(u)$ for all $\hat{Z} \subseteq V \setminus (X \cup W)$, where $X' = X \setminus X''$, $x' = x|X'$, $\overline{x}'' = \overline{x}|X''$, and $\hat{z} = \hat{Z}(u)$. Informally, we can then move all variables in X'' to the W-part of AC2 assigning them $\overline{x}''$. That is, set in AC2 $\overline{x}' = \overline{x}|X'$ and $w' = \overline{x}''w$ where $W' = X'' \cup W$ is the respective W-part. Then, $X' = x'$ is another weak cause of ϕ under u, which is smaller than $X = x$. This is a contradiction.

Case 2: For every nonempty $X'' \subseteq X$, $X'' \neq X$, there exists some $\hat{Z} \subseteq V \setminus (X \cup W)$ such that $\neg\phi_{x'\overline{x}''w\hat{z}}(u)$, where $X' = X \setminus X''$, $x' = x|X'$, $\overline{x}'' = \overline{x}|X''$, and $\hat{z} = \hat{Z}(u)$. Informally, we can then take any such X'' and move the variables in X' and $\hat{Z}$ to the W-part of AC2 assigning them values $x' = x|X'$ and $\hat{z} = \hat{Z}(u)$. That is, set in AC2 $x'' = x|X''$ and $w' = x'w\hat{z}$ where $W' = X' \cup W \cup \hat{Z}$ is the respective W-part. Then, $X'' = x''$ is a weak cause of ϕ under u (note that each instance of AC2(b) for $X'' = x''$ amounts to an instance of AC2(b) for $X = x$). This is a contradiction. □

Hence, $X = x$ is an actual cause of ϕ under u iff (i) $X = x$ is a weak cause of ϕ under u, and (ii) $X = x$ is primitive. As an immediate corollary of Theorems 4.3 and 4.6, it thus follows that deciding whether $X = x$ is an actual cause of ϕ under u is Σ_2^P-complete in the general case.

Corollary 4.7 *Given $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, $u \in D(U)$, and an event ϕ, deciding whether $X = x$ is an actual cause of ϕ under u is Σ_2^P-complete. Hardness holds even if X is a singleton, and M is bounded or ϕ is primitive.*

Furthermore, as an immediate corollary of Theorems 4.5 and 4.6, we obtain that deciding whether $X = x$ is an actual cause of ϕ under u is NP-complete in the binary case.

Corollary 4.8 *Given a binary $M = (U, V, F)$, $X \subseteq V$, $x \in D(X)$, $u \in D(U)$, and an event ϕ, deciding whether $X = x$ is an actual cause of ϕ under u is* NP*-complete. Hardness holds even if M is bounded, X is a singleton, and ϕ is primitive.*

5 Probabilistic Causality

As a representative of probabilistic causal relationships, we finally consider the notion of probabilistic causal irrelevance by Galles and Pearl [1997]; see also [Pearl, 1999; 2000].

A *counterfactual formula* is an expression $Y_x = y$, where $X, Y \subseteq V$, $x \in D(X)$, and $y \in D(Y)$. Given a probabilistic causal model (M, P), where $M = (U, V, F)$, the *probability* of $Y_x = y$, denoted $P(Y_x = y)$, is the sum of all $P(u)$ such that $Y_x(u) = y$. For $X, Y, Z \subseteq V$ with $X, Y \neq \emptyset$, we say *X is probabilistically causally irrelevant to Y given Z*, denoted $(X \not\to Y \mid Z)_P$, iff for all $x, x' \in D(X)$, $y \in D(Y)$, and $z \in D(Z)$, it holds that $P(Y_{xz} = y) = P(Y_{x'z} = y)$. Intuitively, once the value of Z is fixed at z, the value of X does not affect the probability of Y.

The following theorem shows that deciding probabilistic causal irrelevance is complete for $\mathbf{C}_{=}$.

Theorem 5.1 *Given a probabilistic causal model (M, P), where $M = (U, V, F)$, and $X, Y, Z \subseteq V$, deciding whether $(X \not\to Y \mid Z)_P$ is complete for $\mathbf{C}_{=}$.*

Theorem 5.1 is nontrivial and needs some explanations. Firstly, the result means that, in a sense, testing probabilistic causal irrelevance is harder than co-NP, and thus not polynomially reducible to SAT-testing. Moreover, it cannot be reduced to any problem solver for problems in the polynomial

hierarchy. On the other hand, the problem is "easier" than PP-complete problems, which could perhaps help in finding polynomial time (randomized) approximation algorithms.

The easier part of this result is the lower bound. Hardness for $\mathbf{C}_{=}$ can be proved by a reduction from the following $\mathbf{C}_{=}$-complete problem HALFSAT: Given a propositional formula in 3DNF γ on the variables $x_1, \ldots, x_n$, decide whether exactly half of the assignments to $x_1, \ldots, x_n$ satisfy γ. The construction is based on ideas in the proof of Theorem 3.2, but more involved. In fact, it establishes hardness for the case where M is binary and bounded, Z is empty, X, Y are singletons, and P is the uniform distribution. Thus, the problem shows its full complexity already in a minimalistic setting.

The more difficult part is membership in $\mathbf{C}_{=}$. Recall that $(X \not\to Y \mid Z)_P$ iff $P(Y_{xz} = y) = P(Y_{x'z} = y)$ for all values x', x, y, and z. Checking that $P(Y_{xz} = y) = P(Y_{x'z} = y)$, for given x, x', y, and z, can be reduced to the following problem EQUALRUN, which is in $\mathbf{C}_{=}$: Given two NP Turing machines T_1 and T_2 and an input string w, decide whether T_1 and T_2 have the same number of accepting runs on w.

Indeed, let P be given by (f, b) as described in Section 2.2. Let the NP Turing machines T_1 and T_2 have input x, x', y, z and nondeterministically generate all $u \in D(U)$. Then, for each u, let T_1 and T_2 generate $f(u)$ paths. On each of these paths, T_1 (resp., T_2) computes deterministically $Y_{xz}(u)$ (resp., $Y_{x'z}(u)$), and accepts if this value coincides with y, otherwise it rejects. Then, $P(Y_{xz} = y) = P(Y_{x'z} = y)$ iff T_1 and T_2 have the same numbers of accepting paths.

Obviously, T_1 and T_2 can be constructed in polynomial time from M and x, x', y, z. However, we actually need to test that $P(Y_{xz} = y) = P(Y_{x'z} = y)$ for *all* values x', x, y, and z. What we obtain (by slight adaptations) is that the problem is in the class $\forall\mathbf{C}_{=}$, which is a generalization of $\mathbf{C}_{=}$ similar to Π_2^P for co-NP: deciding whether I is a Yes-instance can be informally expressed as deciding whether for every polynomial size string J, it holds that I, J is a Yes-instance of the same problem in $\mathbf{C}_{=}$. Since $\forall\mathbf{C}_{=}$ is equal to $\mathbf{C}_{=}$ (cf. [Green, 1993]), this reduction actually proves membership in $\mathbf{C}_{=}$.

6 Conclusion and Outlook

We analyzed the complexity of causal relationships in Pearl's structural models. In particular, we considered causality between variables, event causality, and probabilistic causality. It turned out that all discussed notions of causality are intractable, where the sophisticated notions of weak and actual causality, and the notion of probabilistic causal irrelevance have the highest complexity (Σ_2^P and $\mathbf{C}_{=}$, respectively).

Our results give useful insight, and may be exploited e.g. in evaluating probabilistic counterfactuals as defined in [Balke and Pearl, 1994]. Note that the evaluation of probabilistic counterfactuals can be polynomially reduced to standard inference tasks in Bayesian networks, and thus has similar computational properties. It is easy to see that the complexity of computing conditional probabilities in Bayesian networks, which is complete for $\#\mathrm{P}$ [Roth, 1996], carries over to computing probabilities of counterfactual statements. Similar to independencies [Pearl, 2000], deterministic and probabilistic causal relationships might now be used to simplify the evaluation of probabilistic counterfactuals. By our results, this seems reasonable, as the complexity of testing simple causal relationships (at most D^{P}, $\mathbf{C}_{=}$) is much lower than the complexity of evaluating probabilistic counterfactuals ($\#\mathrm{P}$-hard).

Moreover, our results may help to analyze the computational aspects of explanations and partial explanations as introduced by Halpern and Pearl [2000], which are crucially based on the notions of weak and actual causality.

An interesting topic of future research is to explore whether there are restricted cases in which testing causal relationships in the structural-model approach is tractable. For example, probabilistic causal irrelevance in stable causal models [Galles and Pearl, 1997] can be tested in polynomial time as it coincides with path interception in their causal graphs.

Acknowledgments

This work has been partially supported by the Austrian Science Fund Project N Z29-INF and by a DFG grant. We are very grateful to Jacobo Torán for useful comments related to counting classes and for pointing out relevant literature. Many thanks also to the reviewers for their useful comments.

References

[Balke and Pearl, 1994] A. Balke and J. Pearl. Probabilistic evaluation of counterfactual queries. In *Proceedings AAAI-94*, pages 230–237, 1994.

[Eiter and Lukasiewicz, 2001] T. Eiter and T. Lukasiewicz. Complexity results for structure-based causality. Technical Report INFSYS RR-1843-01-01, Institut für Informationssysteme, TU Wien, 2001.

[Galles and Pearl, 1997] D. Galles and J. Pearl. Axioms of causal relevance. *Artif. Intell.*, 97:9–43, 1997.

[Geffner, 1990] H. Geffner. Causal theories for nonmonotonic reasoning. In *Proc. AAAI-90*, pages 524–530, 1990.

[Geffner, 1992] H. Geffner. *Default Reasoning: Causal and Conditional Theories*. MIT Press, 1992.

[Green, 1993] F. Green. On the power of deterministic reductions to $\mathrm{C}{=}\mathrm{P}$. *Math. Syst. Theory*, 26(2):215–233, 1993.

[Halpern and Pearl, 2000] J. Y. Halpern and J. Pearl. Causes and explanations: A structural-model approach. Technical Report R-266, UCLA Cognitive Systems Lab, 2000.

[Halpern and Pearl, 2001] J. Y. Halpern and J. Pearl. Causes and explanations: A structural-model approach — Part II: Explanation. In *Proceedings IJCAI-01*, 2001.

[Halpern, 2000] J. Y. Halpern. Axiomatizing causal reasoning. *J. Artif. Intell. Res.*, 12:317–337, 2000.

[Papadimitriou, 1994] C. H. Papadimitriou. *Computational Complexity*. Addison-Wesley, Reading, MA, 1994.

[Pearl, 1999] J. Pearl. Reasoning with cause and effect. In *Proceedings IJCAI-99*, pages 1437–1449, 1999.

[Pearl, 2000] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.

[Roth, 1996] D. Roth. On the hardness of approximate reasoning. *Artif. Intell.*, 82(1–2):273–302, 1996.

[Torán, 1991] J. Torán. Complexity classes defined by counting quantifiers. *J. ACM*, 38(3):753–774, 1991.

[Turner, 1999] H. Turner. A logic of universal causation. *Artif. Intell.*, 113:87–123, 1999.